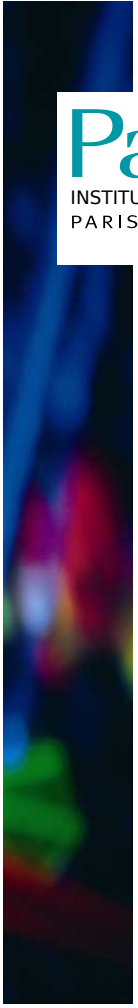
Contrôle multi-objectifs d'ordre réduit

Christian Fischer



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n°84 :
Sciences et technologies de l'information et de la communication

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École Nationale Supérieure des Mines de Paris

Spécialité « Contrôle, Optimisation et Prospective »

présentée et soutenue publiquement par

Christian FISCHER

le 27 Juillet 2011

Contrôle multi-objectifs d'Ordre réduit

Directeur de thèse : **Nadia MAÏZI**
Co-encadrement de la thèse : **Marc BORDIER**

Jury

M. Michel ZASADZINSKI, Professeur, Université Henri Poincaré, Cosnes-et-Romain	Président
M. Alban QUADRAT, Chargé de Recherche, INRIA Saclay	Rapporteur
Mme Nadia MAÏZI, Professeur, Mines ParisTech	Examinateur
M. Marc BORDIER, Chargé de Recherche, Mines ParisTech	Examinateur
Mme Martine OLIVI, Chargé de Recherche, INRIA Sophia Antipolis	Membre invité
M. Rémi DRAI, Docteur-Ingénieur, ESA, Noordwijk, Pays-Bas	Membre invité

MINES ParisTech
Centre de Mathématiques Appliquées
Rue Claude Daunesse B. P. 207, 06904 Sophia Antipolis Cedex, France

T H È S E

Contents

Remerciements

Tous mes remerciements vont aux membres de l'équipe du Centre de Mathématiques Appliquées pour leur accueil chaleureux et pour la patience avec laquelle ils m'ont répondu à toutes mes questions,

spécialement à Professeur Nadia Maïzi qui m'a donné l'opportunité d'effectuer ma thèse au sein de son laboratoire et qui m'a confié un travail exigeant mais particulièrement intéressant,

à mes tuteurs de thèse Monsieur Marc Bordier et Monsieur Jean-Paul Marmorat pour leur créativité exceptionnelle face aux problèmes scientifiques et pour la bonne atmosphère de travail qu'ils ont su créer,

à Madame Valérie Roy pour l'expérience édifiante comme tuteur dans le cadre des projets MIG de l'école que je ne veux pas m'en passer,

à Madame Josiane Bedos, à Madame Dominique Micollier, à Monsieur Jean-Charles Bonin et à Monsieur Gilles Guerassimoff pour leur soutien, leur aide directe et leurs conseils précieux,

et à mes camarades avec lesquels j'ai travaillé ensemble, qui n'ont jamais hésité de m'aider et qui, grâce à leur humeur, ont rendu mon séjour inoubliable.

Résumé

Le présent rapport décrit une méthode de résolution d'un problème de contrôle multi-objectifs d'un système linéaire multi-entrées multi-sorties. Étant donné un système, on cherche un régulateur tel que la dynamique du système rebouclé sur son régulateur vérifie certaines propriétés. Outre ces contraintes sur la dyamique, on exige que le régulateur optimal ait une complexité – mesurée par son degré de McMillan – donnée. Ce problème, réputé difficile, a fait l'objet, depuis le début des années 2000, de nombreuses recherches [3], [8] et il existe désormais des boites à outils Matlab telles que *Hinfstruct* et *Hifoo* qui, à ordre de régulateur fixé, en donnent une solution. Essentiellement, les algorithmes proposés reposent sur des techniques d'optimisation non différentiable du critère par rapport au régulateur lui même. Ce rapport présente une autre façon de résoudre ce problème. A la différence des techniques mentionnées ci-dessus, on construit un calcul differentiel fondé sur un paramétrage minimal des régulateurs stabilisants. La démarche proposée connecte des idées de trois domaines différents présentés au cours des trois chapitres principaux. Le chapitre 2 développe, à partir du concept de factorisation sur un anneau initié par le travail de [13], une méthode générique de factorisation à ordre donné d'une boucle fermée en faisant apparaitre un paramétrage nouveau du régulateur. Le chapitre 3 introduit la notion de système conservatif qui, selon [2], forment une variété différentiable. Cette propriété, appliquée au paramétrage développé à la fin du chapitre précédent, permet alors de bouger le correcteur d'une manière différentiable : pour un régulateur réel d'ordre n et de dimension $p \times m$, le nombre de paramètres réels nécessaires à sa représentation est $n(p+m)$ à comparer avec les $n^2 + n(m+p)$ utilisés par [3] et [8]. Le chapitre 4 introduit le problème d'optimisation multi-objectifs dans lequel les spécifications sont exprimées au moyen de quelques normes communes de système. En appliquant les résultats trouvés en [4] et en utilisant le calcul différentiel développé à la fin du chapitre précédent, il devient possible de calculer un sous gradient du critère qui, par une méthode de descente, permet d'améliorer le régulateur relativement aux spécifications données tout en gardant son ordre fixé.

Dans le chapitre 2, étant donné l'unité $\boldsymbol{P}$, on cherche un régulateur $\boldsymbol{K}$ tel que les éléments formant la matrice de transfert $\boldsymbol{H}$ de la boucle fermée construite à partir ces deux systèmes linéaires, soient des fonctions de transfert propres ayant toutes leurs pôles dans une zone Z prédéfinie par l'utilisateur. L'ensemble de ces fonctions de transfert a une structure d'anneau et les résultats de la théorie des anneaux présentés en A.1 peuvent être appliqués afin d'obtenir une description de tous les régulateurs qui satisfont cette exigence de conception. En particulier, on montre comment ils engendrent une factorisation de la boucle fermée sur l'anneau de référence et comment l'ensemble des régulateurs factorisants

peut être décrit au moyen du paramètre de Youla-Kucera $\boldsymbol{Q}$. Le passage à la représentation d'état, qui est numériquement plus pratique, introduit une nouvelle variable à savoir la dimension de l'espace d'état de la représentation. Bien qu'une factorisation soit toujours possible, celle-ci se paye, en général, par des réalisations de grande taille pour chacun des facteurs et par des régulateurs complexes d'ordre élevé. Une factorisation de la boucle fermée à ordre fixé est alors introduite comme un cas spécial de la factorisation générale obtenu en fixant les dimensions des espaces d'états de l'unité et du régulateur. L'avantage de fixer ces dimensions consiste en ce que les régulateurs à ordre réduit, avec une dimension d'espace d'état égale ou inférieure à celle de l'unité, peuvent être paramétrés directement, ce qui est intéressant si la situation de contrôle demande des régulateurs à complexité réduite. Enfin, on montre pourquoi les paramètrages à ordre fixé $\boldsymbol{Q}_L, \boldsymbol{Q}_R$ partagent avec le paramètre $\boldsymbol{Q}$ la propriété d'être chacun un élément de l'anneau pour que la boucle $\boldsymbol{H}$ le soit aussi, un fait qui permettra à $\boldsymbol{Q}_L, \boldsymbol{Q}_R$ d'être eux mêmes paramétrés au moyen des systèmes conservatifs au cours du chapitre 3.

Le chapitre 3 introduit la notion de système conservatif et fournit un paramétrage des paires stables observables en résolvant un problème d'interpolation du type Nudelman. Il a été montré en [2] que les systèmes conservatifs et les paires stables observables qui peuvent en être extraites forment une variété différentiable. Le régulateur $\boldsymbol{K}$ pouvant ne pas être stable, ce paramétrage ne peut pas être appliqué directement à $\boldsymbol{K}$. Cependant il a été montré à la fin du chapitre 2 que le paramètre gauche $\boldsymbol{Q}_L$ devait être stable pour que la boucle $\boldsymbol{H}$ soit stable et observable pour que le régulateur $\boldsymbol{K}$ soit observable. Le paramétrage de la paire stable observable contenue dans $\boldsymbol{Q}_L$ permet alors de faire varier de fa con infinitésimale $\boldsymbol{Q}_L$ tout entier et de construire un calcul différentiel sur $\boldsymbol{Q}_L$ ce qui sera important au cours l'étape d'optimisation du chapitre 4.

Le chapitre 4 définit un problème d'optimisation multi-objectifs. Pour une unité $\boldsymbol{P}$ donnée, un régulateur $\boldsymbol{K}$ doit être trouvé tel qu'il soit optimal relativement à une fonction de coût donnée et des contraintes données. Ces objectifs d'optimisation sont formulés en termes de normes L1, L2/H2, H∞ de système. L'avantage d'utiliser les normes L1, L2/H2, H∞ consiste en ce que leurs calculs peuvent eux-même être énoncés sous forme de problèmes d'optimisation semi-définis standards ce qui permet de calculer les sensibilités locales ainsi que le sous gradient du critère d'optimisation multi-objectifs. La formulation du problème d'optimisation en termes du paramétrage élaboré au cours des chapitres 2 et 3 permet alors de bouger de fa con infinitésimale le paramètre $\boldsymbol{Q}_L$ à la place de $\boldsymbol{K}$ dans la direction descendante indiquée par le sous gradient. Dès lors que la tâche d'optimisation est complète, le régulateur $\boldsymbol{K}$ peut être récupéré du paramètre $\boldsymbol{Q}_L$. Ceci permet une amélioration locale du régulateur initial $\boldsymbol{K}_0$ donné vers un régulateur supérieur $\boldsymbol{K}^*$ relativement à la fonction de coût et dans le cadre des contraintes.

Pour conclure cette introduction, disons qu'une méthode d'amélioration d'un régulateur relativement à des spécifications multi-objectifs formulée en termes des normes de système L1, L2/H2, H∞, est présentée. Trois éléments clés distinguent l'approche proposée. Premièrement, un paramétrage des régulateurs à ordre fixé au moyen d'un paramètrage des systèmes stables et observables, basé sur le travail en [13]. Deuxièmement, un paramétrage de toutes les paires stables observables qui permet un ajustement infinitésimal du régulateur basé sur le tra-

vail en [2]. Et troisièmement, l'utilisation d'une information de sensibilité pour calculer le sous gradient local du problème d'optimisation pour que le régulateur puisse être ajusté de la bonne fa con, basé sur le travail en [4]. Bien que beaucoup de mesures de performance pour quantifier les spécifications d'un problème d'optimisation multi-objectifs existent, les normes de système L1, L2/H2, H∞ sont possiblement les plus importantes. Elles ont été choisies dans ce rapport à cause de leur pertinence pratique et aussi pour illustrer le champ d'application de l'approche proposée. Pour un système donné, si les normes H2 et H∞ se traduisent par des inégalités matricielles linéaires, la norme L1 s'obtient comme inégalité matricielle bilinéaire. Malgré ces différences, la méthode que nous proposons traite ces contraintes d'une manière uniforme, comme une partie d'une contrainte semi-définie plus générale. La nécessité pour l'approche proposée d'accéder à l'information de sensibilité ne limite nullement les contraintes possibles à ces trois normes. De fait, la solution décrite dans cette thèse est basée sur le paramétrage gauche $\boldsymbol{Q}_L$ du régulateur auquel correspond un paramétrage des paires stables observables. Cependant l'approche duale est aussi possible. Elle utilise le paramètrage droit $\boldsymbol{Q}_R$ du régulateur et le paramétrage des paires stables contrôlables. Dans les applications pratiques on peut imaginer d'utiliser $\boldsymbol{Q}_L$ et le paramétrage des paires stables observables dans le cas où $\boldsymbol{K}$ a plus d'entrées que de sorties vu que les matrices de sortie $\boldsymbol{C}_K, \boldsymbol{C}_{QL}$ ont des dimensions inférieures à celles des matrices d'entrée $\boldsymbol{B}_K, \boldsymbol{B}_{QR}$ et que le paramétrage des paires stables observables de $(\boldsymbol{A}_{QL}, \boldsymbol{C}_{QL})$ devient alors plus simple que le paramétrage des paires stables contrôlables de $(\boldsymbol{A}_{QR}, \boldsymbol{B}_{QR})$ et d'utiliser $\boldsymbol{Q}_R$ et le paramétrage des paires stables contrôlables dans l'autre cas. Les équations de Riccati (2.21), (2.22) peuvent posséder des solutions $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ multiples pour une unité $\boldsymbol{P}$ et un régulateur $\boldsymbol{K}$ donnés. Les conséquences sur le comportement de l'algorithme de descente ne sont pas encore connues. Il peut être possible que ces comportements soient différents pour des $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ différents de sorte que, démarrant d'un point selle ou d'un maximum de la fonction de coût par exemple, le résultat de l'optimisation locale soit totalement différent.

Chapter 1

Introduction

This report describes an attempt to solve a multiobjective optimization problem in the domain of multi input multi output linear time invariant control. This problem consists of a specified plant whose output signals are fed back to its input signals by an unspecified controller so that multiple specified objectives are met. As an additional constraint, the resulting optimal controller is required to have a specified complexity, measured by its order which is, simply put, the dimension of its associated state space.

1.1 Motivation

The idea for this approach started while trying to improve a working controller for a satellite with respect to a specified cost function and constrained by multiple specified design requirements. The environment in which the controller has to run on board of a satellite strongly affects the choice of suitable controllers. Objects in space operate in a very low density environment, almost vacuum. On the one hand, the low material density surrounding a satellite implies almost no fluid damping of oscillations of the main body and its most often long elastic appendages, like solar panels. Modeling a satellite therefore leads to a complex plant of high order with almost undamped elastic modes that can not be neglected. On the other hand, the almost lack of a material screen around the satellite has it exposed to strong electromagnetic radiation. The onboard computers therefore need to be radiation hardened which leads to a lower computation performance as a tradeoff. This explains why in this environment the controller is required to be of a specified order, almost always lower than the order of the plant, which is the satellite model. While common sense suggests that it must be easier to find an optimal controller of high order since there are more parameters and thus more freedom to play with, the converse argument indicates that it also might be more difficult to find a reduced order controller that matches the specifications. This challenge is the motivation for this work.

1.2 Overview

There exist methods *Hinfstruct* and *Hifoo* which have been cast in Matlab toolboxes of the same name, that allow the controller order to be fixed, as presented

in [3] and [8]. These algorithms apply nonsmooth optimization directly to the controller. Another attempt to solve this problem is described in this report. In contrast to the aforementioned techniques, a differentiable parametrization of the controller is developed so that smooth optimization can be applied to the controller parameter and thus indirectly to the controller. It connects ideas from three different areas that are presented in the three main chapters.

Chapter 2 introduces the concept of feedback loop factorizations on algebraic rings pioneered by the work in [13]. However, this approach, while providing a parametrization of all controllers, in general leads to high order controllers that more or less cancel out the plant dynamics and replace it with any desired dynamics. The controller is thus in general working against the nature of the plant. To avoid this, a new approach is presented that uses fixed orders from ground up, starting with fixed order loop factorizations and ending with a fixed order controller parametrization. The parametrization of fixed order controllers inherits an important property from the aforementioned parametrization of all controllers in that the parameter needs to be stable for the feedback loop to be stable. In addition to that the fixed order parameter needs to be observable for the controller to be observable. Hence the fixed order controller parameter is necessarily stable and observable, a fact that is important for the next chapter.

Chapter 3 introduces the notion of square lossless systems and shows that any stable observable pair of a state space realization can be completed to form a lossless system. A parametrization of stable observable pairs is then obtained by solving an interpolation problem for lossless systems. It has been shown in [2] that lossless systems and the stable observable pairs that can be extracted from them form a differentiable manifold. Thus, parametrizing the stable observable pair of the fixed order controller parameter found in the previous chapter allows to move the fixed order controller parameter smoothly, a fact that is important for the next chapter.

Chapter 4 introduces the multiobjective optimization setup with objectives formulated in terms of some common system norms. This problem turns out to be nonconvex and thus difficult to solve. However, fixing the controller parameter leads to a convex solvable problem. Solving the multiobjective optimization problem for a current controller parameter value then provides a local sensitivity information in form of the optimal Lagrange multiplier value. By applying the results in [4] this sensitivity information can be used to calculate the subgradient of the cost function for the current controller parameter value. Due to the smooth parametrization of the controller parameter found in the previous chapter, the information provided by the subgradient enables a local improvement of the controller parameter value with respect to the cost function. Then the multiobjective optimization problem is again solved for the updated controller parameter value and the process repeats until satisfactory results are achieved or no further progress is made.

To ease the reading, an annex is provided with some of the most frequently used formulas and theorems in algebra, analysis and linear time invariant system theory.

1.3 Notation

Within the scope of this report bold lowercase letters represent vectors, bold uppercase letters represent matrices and non bold letters represent scalars. The constants $\mathbf{0}$ and $\mathbf{1}$ represent the additive and multiplicative matrix identities respectively. The matrices $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ represent the transposed and transconjugated matrix $\boldsymbol{A}$ respectively. The functions $\operatorname{tr}\boldsymbol{A}$ and $\det\boldsymbol{A}$ denote the trace and determinant of matrix $\boldsymbol{A}$ respectively. The functions $\operatorname{rnk}\boldsymbol{A}$ and $\operatorname{col}\boldsymbol{A}$ return the rank and number of columns of matrix $\boldsymbol{A}$ respectively. In an abuse of language, a matrix $\boldsymbol{A}$ is called *stable* if and only if $\operatorname{Re}\lambda < 0$ for continuous systems or $|\lambda| < 1$ for discrete systems for all eigenvalues λ of matrix $\boldsymbol{A}$. The sets $\mathbb{C}, \mathbb{R}, \mathbb{Z}, \mathbb{N}$ denote the sets of complex, real, entire and natural numbers respectively, the set $\mathbb{N}_n$ denotes the natural numbers up to and including the natural number n. A zero subscript denotes an initial value, a star superscript denotes an optimal value.

Chapter 2

Controller Parametrizations

In this chapter, a feedback loop $\boldsymbol{H}$ is constructed from two linear systems, a given plant $\boldsymbol{P}$ and a controller $\boldsymbol{K}$ which is to be designed. The controller has to assure that the transfer matrix of the feedback loop becomes a matrix of proper transfer functions whose poles are confined to a user defined region Z. These transfer functions form the algebraic structure of a ring. Results of the ring theory provided in A.1 are then applied to obtain a description of all controllers that meet this design requirement. It is shown how these controllers admit a general factorization of the closed feedback loop on the given ring. A parametrization of all these factorizing controllers with a parameter $\boldsymbol{Q}$ is then given in transfer function representation. Passing to the numerically more practical state space representation of linear systems introduces a new variable which is the state space dimension or order of a representation. Albeit a factorization is always possible, it is shown that this in general has to be paid by high state space dimensions of the factors which leads to complex controllers of high state dimension. A special fixed order factorization of the feedback loop is then presented in state space representation, which is the special case of the general factorization in that the state space dimensions of plant and controller are fixed. Fixing the dimension has the advantage that reduced order controllers, with state dimension equal or less than that of the plant, can be parametrized directly, which is of interest if controllers of low complexity are to be found. The fixed order parameters $\boldsymbol{Q}_L, \boldsymbol{Q}_R$ share the property of the parameter $\boldsymbol{Q}$ in that the parameter needs to be a ring member for the loop $\boldsymbol{H}$ to be a ring member, a fact that allows the parameter itself to be parametrized in chapter 3.

2.1 Ring of Transfer Functions

This section serves as a bridge between algebraic ring structures and the transfer functions that arise in the study of linear time invariant systems.

2.1.1 Transfer Functions

A linear time invariant system may be expressed either in time domain, where it is described by a system of linear differential equations over time $t \in \mathbb{R}$ or $k \in \mathbb{Z}$, or in frequency domain where it is described by a system of linear

Figure 2.1: A linear time invariant system $\boldsymbol{P}$ with input $\boldsymbol{u}$ and output $\boldsymbol{y}$.

algebraic equations over frequency $s \in \mathbb{C}$ or $z \in \mathbb{C}$, for continuous or discrete systems respectively. The link between the two descriptions is the Laplace transformation.

Figure 2.1 shows a block diagram of a linear time invariant system. The input $\boldsymbol{u}$ and output $\boldsymbol{y}$ are in general vectors containing multiple scalar inputs and outputs respectively. Due to the linearity, the system description in frequency domain can be brought in the following form:

$$\boldsymbol{y}(x) = \boldsymbol{P}(x)\boldsymbol{u}(x) \tag{2.1}$$

with $x \in \{s, z\}$ for continuous and discrete systems respectively. For reason of clarity the arguments s, z are omitted in the following. The matrix function $\boldsymbol{P}$ between input and output is called the *transfer matrix*. Its entries, the scalar *transfer functions*, are rational functions of frequency $x \in \{s, z\}$

$$y_j = [\boldsymbol{P}]_{ij} u_i = k \frac{\prod_{k=0}^{m}(x - a_k)}{\prod_{l=0}^{n}(x - b_l)} u_i$$

with zeros $a_k \in \mathbb{C}$, poles $b_l \in \mathbb{C}$ and static gain $k \in \mathbb{C}$. A physically realizable transfer function is characterized by $m \leq n$ and is called *proper*, so the numerator polynomial is never of higher degree than the denominator polynomial, see also section 2.3.1. Transfer functions are either *stable* or unstable. A continuous transfer function is stable if and only if all its poles b_l have negative real part $\operatorname{Re} b_l < 0, \ \forall b_l$. A discrete transfer function is stable if and only if all its poles b_l lie inside the unit circle $|b_l| < 1, \ \forall b_l$.

The transfer function may have common zeros and poles. In this case they can be mutually cancelled and the degree of the numerator and denominator polynomials can be decreased until they have no common zeros and poles. Then the transfer function is said to be *minimal* since the degree of the numerator and denominator polynomials can not be decreased further.

2.1.2 Ring Structure

Verification of (A.1) to (A.9) as well as (A.13) and (A.14) shows that proper transfer functions with all poles contained in a zone Z form a commutative integral ring R_Z. It has been shown in [6] that these transfer functions more precisely form a *principal* commutative integral ring, so that every ideal is principal. The multiplicative inverse of proper transfer functions with $m = n$ and all poles and zeros contained in Z are again proper transfer functions with all poles in Z, so they are the units in this ring.

Definition Let R_Z be the ring of transfer functions with all poles contained in a selected zone $Z \subset \mathbb{C}$ and let $F_Z = \frac{R_Z}{R_Z \setminus \{0\}}$ be the field of fractions on R_Z.

Proposition All physically realizable and thus proper transfer functions are a member of F_Z.

Proof Let an arbitrary physically realizable transfer function be given by

$$P = k\frac{\prod_{k=0}^{m}(x - c_k)}{\prod_{l=0}^{n}(x - d_l)}$$

with $x \in \{s, z\}$ for continuous and discrete systems respectively. This transfer function is physically realizable if it is proper so that $m \leq n$. Then let $c_1, \ldots, c_n \in Z$ and define

$$A = k\frac{\prod_{k=0}^{m}(x - a_k)}{\prod_{l=0}^{n}(x - c_l)},\ B = \frac{\prod_{l=0}^{n}(x - b_l)}{\prod_{l=0}^{n}(x - c_l)}$$

then obviously $P = \frac{A}{B}$ with $A, B \in R_Z$ and therefore $P \in F_Z$.

Remark 1 This means that the ring theory of chapter A.1 can now be applied to proper transfer functions with all poles in Z. The extension to matrices of section A.1.5 allows the treatment of entire proper transfer matrices with all poles in Z. The extension to fractions of section A.1.6 allows the treatment of general transfer matrices with poles anywhere. That way, the transfer matrix of any linear time invariant system can be described by means of proper transfer functions with poles in Z.

Remark 2 If the zone Z coincides with the left half plane for continuous systems or the inner unit disk for discrete systems, then proper transfer functions with all poles in this zone Z are stable. Therefore stability, which is a main objective in linear control, arises naturally in this representation but such a ring is only a special case as the zone Z may also be defined so that a maximal settling time or a minimal amount of damping is respected.

2.2 Controller Parametrization

This section introduces the concept of loop factorization which provides a way of parametrizing the set of all linear controllers in transfer function representation that keep the loop formed with a linear plant in the selected underlying ring R_Z. Details can be found in [13].

2.2.1 The Control Loop

Figure 2.2 shows the control loop underlying this problem. The plant $\boldsymbol{P} \in F_Z^{m,n}$ is controlled by the controller $\boldsymbol{K} \in F_Z^{n,m}$. The input signals $\boldsymbol{v}$ and $\boldsymbol{w}$ might be the reference and measurement noise, respectively and the output signals $\boldsymbol{y}$ and $\boldsymbol{u}$ are the plant and controller outputs, respectively.

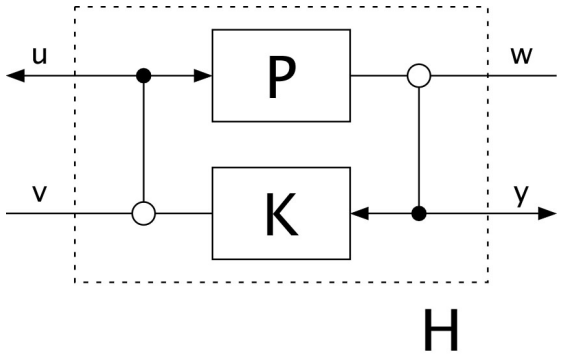


Figure 2.2: The plant $\boldsymbol{P}$ and the controller $\boldsymbol{K}$ form the control loop.

Proposition The loop shown in figure 2.2 has a transfer matrix $\boldsymbol{H}$ given by

$$\boldsymbol{H} = \begin{bmatrix} -\boldsymbol{K} & \mathbf{1} \\ \mathbf{1} & -\boldsymbol{P} \end{bmatrix}^{-1} \tag{2.2}$$

or equivalently by

$$\begin{aligned} \boldsymbol{H} = \begin{bmatrix} \boldsymbol{P}(\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1} & \mathbf{1}+\boldsymbol{P}(\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1}\boldsymbol{K} \\ (\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1} & (\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1}\boldsymbol{K} \end{bmatrix} = \\ = \begin{bmatrix} \boldsymbol{P} \\ \mathbf{1} \end{bmatrix} (\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1} \begin{bmatrix} \mathbf{1} & \boldsymbol{K} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{aligned} \tag{2.3}$$

or equivalently by

$$\begin{aligned} \boldsymbol{H} = \begin{bmatrix} (\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1}\boldsymbol{P} & (\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1} \\ \mathbf{1}+\boldsymbol{K}(\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1}\boldsymbol{P} & \boldsymbol{K}(\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1} \end{bmatrix} = \\ = \begin{bmatrix} \mathbf{1} \\ \boldsymbol{K} \end{bmatrix} (\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1} \begin{bmatrix} \boldsymbol{P} & \mathbf{1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \end{aligned} \tag{2.4}$$

Proof From figure 2.2 it can be seen that

$$\left.\begin{aligned} \boldsymbol{y} &= \boldsymbol{w} + \boldsymbol{P}\boldsymbol{u} \\ \boldsymbol{u} &= \boldsymbol{v} + \boldsymbol{K}\boldsymbol{y} \end{aligned}\right\} \Rightarrow \begin{bmatrix} \mathbf{1} & -\boldsymbol{P} \\ -\boldsymbol{K} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix}$$

The transfer matrix $\boldsymbol{H}$ may then be defined by

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix}$$

so that (2.2) follows. The equivalent expressions (2.3), (2.4) then follow by application of (A.59) and use of the relations (A.56).

Remark The loop $\boldsymbol{H}$ is symmetric with respect to $\boldsymbol{P}, \boldsymbol{K}$ as can be seen by (2.2). In the following, the plant $\boldsymbol{P}$ is assumed to be given and the controller $\boldsymbol{K}$ is the variable to assure that the requirements posed on $\boldsymbol{H}$ are met.

2.2.2 Connectibility

Proposition A controller $\boldsymbol{K}$ can be connected to a plant $\boldsymbol{P}$ to form a feedback loop $\boldsymbol{H}$ as given in (2.2) if and only if

$$\exists(\mathbf{1}-\boldsymbol{P}\boldsymbol{K})^{-1} \Leftrightarrow \exists(\mathbf{1}-\boldsymbol{K}\boldsymbol{P})^{-1} \Leftrightarrow \exists \boldsymbol{H} = \begin{bmatrix} -\boldsymbol{K} & \mathbf{1} \\ \mathbf{1} & -\boldsymbol{P} \end{bmatrix}^{-1} \tag{2.5}$$

In that case the loop $\boldsymbol{H}$ is said to be *well connected*.

Proof This follows from (2.3) or (2.4).

Remark From now on, all feedback loops $\boldsymbol{H}$ are supposed to be well connected.

2.2.3 Doubly Coprime Factorization

Definition For a given plant $\boldsymbol{P}$, all controllers $\boldsymbol{K}$ that lead to $\boldsymbol{H} \in R_Z^{m+n,m+n}$, with R_Z being the selected ring of proper transfer functions with all poles contained in a zone Z, form the set of *factorizing controllers*

$$\boldsymbol{K} \in K(\boldsymbol{P}) \Leftrightarrow \boldsymbol{H} \in R_Z^{m+n,m+n} \tag{2.6}$$

Proposition A controller $\boldsymbol{K}$ is a factorizing controller if and only if there exist $\boldsymbol{M}_i \in R_Z^{m,n}$, $\boldsymbol{X}_i \in R_Z^{n,m}$, $\boldsymbol{N}_L, \boldsymbol{Y}_R \in R_Z^{m,m}$, $\boldsymbol{N}_R, \boldsymbol{Y}_L \in R_Z^{n,n}$ with $i \in \{R, L\}$ that satisfy the doubly coprime factorization

$$\begin{aligned} \mathbf{1} &= \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \\ \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} \begin{bmatrix} -\boldsymbol{X}_R & \boldsymbol{N}_R \\ \boldsymbol{Y}_R & -\boldsymbol{M}_R \end{bmatrix} = \begin{bmatrix} -\boldsymbol{X}_L & \boldsymbol{Y}_L \\ \boldsymbol{N}_L & -\boldsymbol{M}_L \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_R & \boldsymbol{Y}_R \\ \boldsymbol{N}_R & \boldsymbol{X}_R \end{bmatrix} \\ \text{with } &\begin{cases} \boldsymbol{P} = \boldsymbol{M}_R \boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1} \boldsymbol{M}_L \\ \boldsymbol{K} = \boldsymbol{X}_R \boldsymbol{Y}_R^{-1} = \boldsymbol{Y}_L^{-1} \boldsymbol{X}_L \end{cases} \end{aligned} \tag{2.7}$$

Furthermore all valid matrices $\boldsymbol{X}_i, \boldsymbol{Y}_i$ and so $\boldsymbol{K}$ are a function of a free parameter matrix $\boldsymbol{Q} \in R_Z^{n,m}$ and can be given by

$$\begin{aligned} \begin{bmatrix} \boldsymbol{X}_R \\ \boldsymbol{Y}_R \end{bmatrix} &= \begin{bmatrix} \boldsymbol{X}_{R0} & \boldsymbol{N}_R \\ \boldsymbol{Y}_{R0} & \boldsymbol{M}_R \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \boldsymbol{Q} \end{bmatrix} \qquad \text{with } \boldsymbol{Y}_R \text{ invertible} \\ \begin{bmatrix} \boldsymbol{X}_L & \boldsymbol{Y}_L \end{bmatrix} &= \begin{bmatrix} \mathbf{1} & \boldsymbol{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{L0} & \boldsymbol{Y}_{L0} \\ \boldsymbol{N}_L & \boldsymbol{M}_L \end{bmatrix} \quad \text{with } \boldsymbol{Y}_L \text{ invertible} \end{aligned} \tag{2.8}$$

with $\boldsymbol{X}_{i0}, \boldsymbol{Y}_{i0}$ being a set of valid initial matrices that satisfy (2.7) and allow the construction of all other solutions $\boldsymbol{X}_i, \boldsymbol{Y}_i$ using (2.8).

Proof Let plant $\boldsymbol{P} \in F_Z^{m,n}$ and controller $\boldsymbol{K} \in F_Z^{n,m}$ be factorized in coprime transfer matrices that are members of the underlying ring R_Z

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{M}_R \boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1} \boldsymbol{M}_L \\ \boldsymbol{K} &= \boldsymbol{X}'_R \boldsymbol{Y}'^{-1}_R = \boldsymbol{Y}'^{-1}_L \boldsymbol{X}'_L \end{aligned}$$

using (A.50) with $\boldsymbol{M}_i \in R_Z^{m,n}; \boldsymbol{X}'_i \in R_Z^{n,m}; \boldsymbol{N}_L, \boldsymbol{Y}'_R \in R_Z^{m,m}; \boldsymbol{N}_R, \boldsymbol{Y}'_L \in R_Z^{n,n}$. With this factorization, $\boldsymbol{H}$ can be expressed as

$$\begin{aligned}\boldsymbol{H} &= \begin{bmatrix}\boldsymbol{M}_R\boldsymbol{N}_R^{-1}\\ \mathbf{1}\end{bmatrix}(\mathbf{1} - \boldsymbol{Y}'_L{}^{-1}\boldsymbol{X}'_L\boldsymbol{M}_R\boldsymbol{N}_R^{-1})^{-1}\begin{bmatrix}\mathbf{1} & \boldsymbol{Y}'_L{}^{-1}\boldsymbol{X}'_L\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{1}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} = \\ &= \begin{bmatrix}\boldsymbol{M}_R\\ \boldsymbol{N}_R\end{bmatrix}(\boldsymbol{Y}'_L\boldsymbol{N}_R - \boldsymbol{X}'_L\boldsymbol{M}_R)^{-1}\begin{bmatrix}\boldsymbol{Y}'_L & \boldsymbol{X}'_L\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{1}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad (2.9)\end{aligned}$$

according to (2.3), and as

$$\begin{aligned}\boldsymbol{H} &= \begin{bmatrix}\mathbf{1}\\ \boldsymbol{X}'_R\boldsymbol{Y}'_R{}^{-1}\end{bmatrix}(\mathbf{1} - \boldsymbol{N}_L^{-1}\boldsymbol{M}_L\boldsymbol{X}'_R\boldsymbol{Y}'_R{}^{-1})^{-1}\begin{bmatrix}\boldsymbol{N}_L^{-1}\boldsymbol{M}_L & \mathbf{1}\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{1} & \mathbf{0}\end{bmatrix} = \\ &= \begin{bmatrix}\boldsymbol{Y}'_R\\ \boldsymbol{X}'_R\end{bmatrix}(\boldsymbol{N}_L\boldsymbol{Y}'_R - \boldsymbol{M}_L\boldsymbol{X}'_R)^{-1}\begin{bmatrix}\boldsymbol{M}_L & \boldsymbol{N}_L\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{1} & \mathbf{0}\end{bmatrix} \quad (2.10)\end{aligned}$$

according to (2.4). The transfer matrix $\boldsymbol{H}$ has its poles in the predefined zone Z and is realizable if and only if $\boldsymbol{H} \in R_Z^{m+n,m+n}$ that is if the elements of $\boldsymbol{H}$ are in R_Z and thus are all proper transfer functions with poles in Z. It can be seen from the two expressions for $\boldsymbol{H}$ above that this is true if and only if the two inverses $(\boldsymbol{Y}'_L\boldsymbol{N}_R - \boldsymbol{X}'_L\boldsymbol{M}_R)^{-1}$ and $(\boldsymbol{N}_L\boldsymbol{Y}'_R - \boldsymbol{M}_L\boldsymbol{X}'_R)^{-1}$ are matrices over R_Z which is equivalent of stating that $\boldsymbol{Y}'_L\boldsymbol{N}_R - \boldsymbol{X}'_L\boldsymbol{M}_R$ and $\boldsymbol{N}_L\boldsymbol{Y}'_R - \boldsymbol{M}_L\boldsymbol{X}'_R$ must be units over R_Z so that

$$\begin{aligned}\boldsymbol{Y}'_L\boldsymbol{N}_R - \boldsymbol{X}'_L\boldsymbol{M}_R &= \boldsymbol{U}_L\\ \boldsymbol{N}_L\boldsymbol{Y}'_R - \boldsymbol{M}_L\boldsymbol{X}'_R &= \boldsymbol{U}_R\end{aligned} \quad (2.11)$$

with $\boldsymbol{U}_R \in R_Z^{m,m}, \boldsymbol{U}_L \in R_Z^{n,n}$ being unit matrices. The set of matrices

$$\begin{aligned}\begin{bmatrix}\boldsymbol{X}_R\\ \boldsymbol{Y}_R\end{bmatrix} &= \begin{bmatrix}\boldsymbol{X}'_R\\ \boldsymbol{Y}'_R\end{bmatrix}\boldsymbol{U}_R^{-1} && \text{right coprime}\\ \begin{bmatrix}\boldsymbol{X}_L & \boldsymbol{Y}_L\end{bmatrix} &= \boldsymbol{U}_L^{-1}\begin{bmatrix}\boldsymbol{X}'_L & \boldsymbol{Y}'_L\end{bmatrix} && \text{left coprime}\end{aligned} \quad (2.12)$$

also form a coprime factorization of the controller since $\boldsymbol{K} = \boldsymbol{X}_R\boldsymbol{Y}_R^{-1} = \boldsymbol{Y}_L^{-1}\boldsymbol{X}_L$. The constraints for loop ring membership (2.11) can now be written as

$$\begin{aligned}\boldsymbol{Y}_L\boldsymbol{N}_R - \boldsymbol{X}_L\boldsymbol{M}_R &= \mathbf{1}\\ \boldsymbol{N}_L\boldsymbol{Y}_R - \boldsymbol{M}_L\boldsymbol{X}_R &= \mathbf{1}\end{aligned}$$

This requirement together with the coprime factorizations $\boldsymbol{P} = \boldsymbol{M}_R\boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1}\boldsymbol{M}_L$ and $\boldsymbol{K} = \boldsymbol{X}_R\boldsymbol{Y}_R^{-1} = \boldsymbol{Y}_L^{-1}\boldsymbol{X}_L$ which lead to $\boldsymbol{M}_L\boldsymbol{N}_R - \boldsymbol{N}_L\boldsymbol{M}_R = \mathbf{0}$ and $\boldsymbol{X}_L\boldsymbol{Y}_R - \boldsymbol{Y}_L\boldsymbol{X}_R = \mathbf{0}$ can be represented by the matrix equation of (2.7). According to (A.51) the set of matrices $\boldsymbol{X}_i, \boldsymbol{Y}_i$ forms a coprime completion of the plant's coprime factorization. With the additional requirement of $\boldsymbol{Y}_i$ to be invertible, the set of matrices $\boldsymbol{X}_i, \boldsymbol{X}_i$ also form a coprime factorization of the controller. Hence, this represents a doubly coprime factorization of $\boldsymbol{P}, \boldsymbol{K}$. According to (A.52), all coprime factors $\boldsymbol{X}_i, \boldsymbol{Y}_i$ of the controller $\boldsymbol{K}$ that solve this doubly coprime factorization for $\boldsymbol{P}$ can be parametrized by a parameter $\boldsymbol{Q} \in R_Z^{n,m}$ and can be brought into the block matrix form of (2.8).

2.2.4 Loop Factorization

Proposition All controllers $\boldsymbol{K}$ from the set $K(\boldsymbol{P})$ of solutions of the doubly coprime factorization (2.7) factorize the loop $\boldsymbol{H}$ in the form of

$$\begin{aligned} \boldsymbol{H} &= \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \\ \boldsymbol{H} &= \begin{bmatrix} \boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{0} \end{bmatrix} \end{aligned} \tag{2.13}$$

with $\boldsymbol{M}_i \in R_Z^{m,n}$, $\boldsymbol{X}_i \in R_Z^{n,m}$, $\boldsymbol{N}_L, \boldsymbol{Y}_R \in R_Z^{m,m}$, $\boldsymbol{N}_R, \boldsymbol{Y}_L \in R_Z^{n,n}$ and $i \in \{R, L\}$.

Proof This immediately follows from the above constraints for ring membership (2.11) of the loop given by (2.9), (2.10) using the factorization (2.12).

Remark An alternative form of $\boldsymbol{H}$ in the form of a serial connection can be given by

$$\begin{aligned} \boldsymbol{H} &= \begin{bmatrix} \boldsymbol{Y}_R & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} -\boldsymbol{X}_R & \boldsymbol{N}_R \\ \boldsymbol{Y}_R & -\boldsymbol{M}_R \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{Y}_R & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \\ \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} \\ \boldsymbol{H} &= \begin{bmatrix} -\boldsymbol{X}_L & \boldsymbol{Y}_L \\ \boldsymbol{N}_L & -\boldsymbol{M}_L \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_L \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_R & \boldsymbol{Y}_R \\ \boldsymbol{N}_R & \boldsymbol{X}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_L \end{bmatrix} \end{aligned} \tag{2.14}$$

which can be directly derived from (2.2). First using the coprime factorizations of $\boldsymbol{P}$ and $\boldsymbol{K}$ gives

$$\begin{aligned} \boldsymbol{H} &= \begin{bmatrix} -\boldsymbol{K} & \boldsymbol{1} \\ \boldsymbol{1} & -\boldsymbol{P} \end{bmatrix}^{-1} = \\ &= \begin{bmatrix} \boldsymbol{Y}_R & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} -\boldsymbol{X}_R & \boldsymbol{N}_R \\ \boldsymbol{Y}_R & -\boldsymbol{M}_R \end{bmatrix}^{-1} = \begin{bmatrix} -\boldsymbol{X}_L & \boldsymbol{Y}_L \\ \boldsymbol{N}_L & -\boldsymbol{M}_L \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_L \end{bmatrix} \end{aligned}$$

Then using the fact that the two block matrices in (2.7) are mutual inverses as constraint for ring membership this immediately leads to (2.14).

2.2.5 Loop Decomposition

Proposition All controllers $\boldsymbol{K}$ from the set $K(\boldsymbol{P})$ provide a decomposition of the loop $\boldsymbol{H}$ that is linear in the parameter $\boldsymbol{Q}$

$$\boldsymbol{H}(\boldsymbol{Q}) = \boldsymbol{H}(\boldsymbol{0}) + \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix} \tag{2.15}$$

with $\boldsymbol{H}(\boldsymbol{0})$ being the loop of plant $\boldsymbol{P}$ fed back by the initial controller $\boldsymbol{K}(\boldsymbol{0})$. The decomposition is illustrated by figure 2.3.

Proof The controller parametrization works in the way that if one valid controller $\boldsymbol{K}(\boldsymbol{0})$ is known, then all other valid controllers $\boldsymbol{K}(\boldsymbol{Q})$ that keep the loop formed with a given plant $\boldsymbol{P}$ in the selected ring R_Z can be found by varying the free ring transfer matrix $\boldsymbol{Q}$ according to (2.8). This variation will change

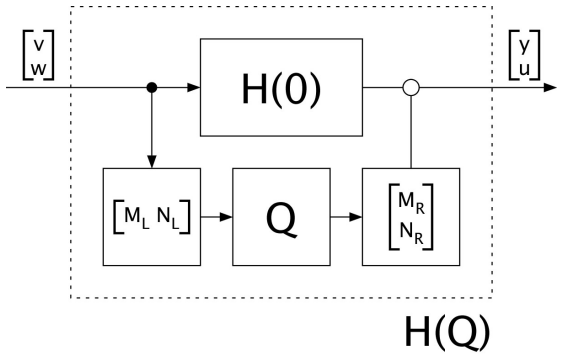


Figure 2.3: Decomposition of the feedback loop by extraction of the free parameter $\boldsymbol{Q}$.

the closed loop transfer matrix $\boldsymbol{H}$ which therefore is also a function of $\boldsymbol{Q}$. This function can be derived by substituting (2.8) in the loop factorization (2.13). The first expression for $\boldsymbol{H}$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

together with

$$\begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{L0} & \boldsymbol{X}_{L0} \\ \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix}$$

from (2.8) leads to

$$\boldsymbol{H}(\boldsymbol{Q}) = \underbrace{\begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{L0} & \boldsymbol{X}_{L0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{H}(\boldsymbol{0})} + \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix}$$

which shows (2.15). The second expression for $\boldsymbol{H}$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{0} \end{bmatrix}$$

together with

$$\begin{bmatrix} \boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_{R0} & \boldsymbol{M}_R \\ \boldsymbol{X}_{R0} & \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{Q} \end{bmatrix}$$

from (2.8) leads to

$$\boldsymbol{H}(\boldsymbol{Q}) = \underbrace{\begin{bmatrix} \boldsymbol{Y}_{R0} \\ \boldsymbol{X}_{R0} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{H}(\boldsymbol{0})} + \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \boldsymbol{Q} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix}$$

which also shows (2.15).

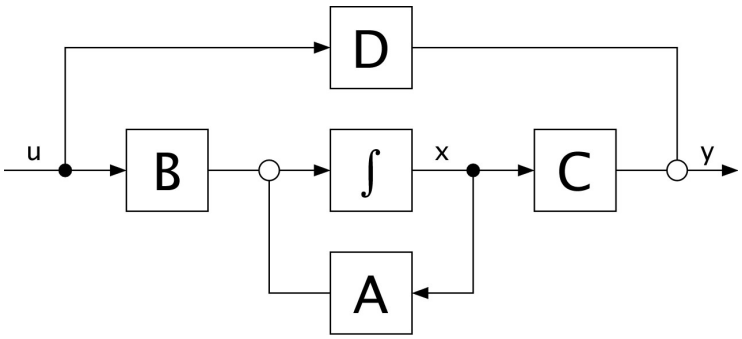


Figure 2.4: Realization of a linear time invariant system with input $\boldsymbol{u}$, output $\boldsymbol{y}$ and system state $\boldsymbol{x}$.

2.3 Passage to State Space

This section shows that a linear time invariant systems in transfer function representation may also be described in state space representation. This has an advantage in computer aided design of controllers since constant matrices instead of polynomials are treated.

2.3.1 Realization of Transfer Functions

Proper transfer functions are physically realizable with the three common elements adder, gain and integrator, whereas improper transfer functions would require a differentiator that is not realizable. Any realizable system such as the plant in figure 2.1 may be written in state space representation which is therefore called a *realization*. In time domain it is given by

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \\ \boldsymbol{y}(t) &= \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) \end{aligned} \tag{2.16}$$

for continuous systems and

$$\begin{aligned} \boldsymbol{x}[k+1] &= \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{B}\boldsymbol{u}[k] \\ \boldsymbol{y}[k] &= \boldsymbol{C}\boldsymbol{x}[k] + \boldsymbol{D}\boldsymbol{u}[k] \end{aligned} \tag{2.17}$$

for discrete systems, with system state vectors $\boldsymbol{x}(t), \boldsymbol{x}[k]$ and constant matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$. For reason of clarity the argument t is omitted in the following. This representation is readily realizable with adders, gains and integrators or delays as shown in figure 2.4. The link between this state space representation in time domain and the transfer matrix $\boldsymbol{P}$ of the plant in frequency domain is the Laplace transformation. By transforming (2.16), (2.17) and eliminating the state $\boldsymbol{x}$ it can be compared to (2.1) which then leads to

$$\boldsymbol{P}(x) = \boldsymbol{C}(x\mathbf{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D} =: \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right]$$

with $x \in \{s, z\}$ for continuous and discrete systems respectively. Since $(x\mathbf{1} - \boldsymbol{A})^{-1} = \mathrm{adj}(x\mathbf{1} - \boldsymbol{A})(\det(x\mathbf{1} - \boldsymbol{A}))^{-1}$ the eigenvalues of $\boldsymbol{A}$ are the poles of $\boldsymbol{P}(x)$. The static gain is $\boldsymbol{P}(\infty) = \boldsymbol{D}$.

The equivalent of a transfer function having common zeros and poles is a state space representation having uncontrollable or unobservable states. Like transfer functions with no common zeros and poles, a completely controllable and observable system is called *minimal*. However, whole transfer matrices describing systems with multiple inputs and outputs may have no common zeros and poles despite being non minimal. This is one of the reasons for the description of systems in state space representation. The other reason is that constant matrices instead of polynomials are treated which is a huge advantage in computer aided controller design.

2.3.2 Realization of the Control Loop

Now let the state space representations of plant $\boldsymbol{P}$, controller $\boldsymbol{K}$ and loop $\boldsymbol{H}$ as in figure 2.2 be given by

$$\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \hline \boldsymbol{C}_P & \boldsymbol{D}_P \end{array}\right],\ \boldsymbol{K} = \left[\begin{array}{c|c} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \hline \boldsymbol{C}_K & \boldsymbol{D}_K \end{array}\right],\ \boldsymbol{H} = \left[\begin{array}{c|c} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \hline \boldsymbol{C}_H & \boldsymbol{D}_H \end{array}\right] \tag{2.18}$$

then with (C.48), the state space matrices of $\boldsymbol{H}$ amount to

$$\begin{aligned} \boldsymbol{A}_H &= \begin{bmatrix} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K \\ \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K \end{bmatrix} \\ \boldsymbol{B}_H &= \begin{bmatrix} \boldsymbol{B}_P\boldsymbol{E}_{KP} & \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} \\ \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP} & \boldsymbol{B}_K\boldsymbol{E}_{PK} \end{bmatrix} \\ \boldsymbol{C}_H &= \begin{bmatrix} \boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K \\ \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{E}_{KP}\boldsymbol{C}_K \end{bmatrix} \\ \boldsymbol{D}_H &= \begin{bmatrix} \boldsymbol{D}_P\boldsymbol{E}_{KP} & \boldsymbol{E}_{PK} \\ \boldsymbol{E}_{KP} & \boldsymbol{D}_K\boldsymbol{E}_{PK} \end{bmatrix} \end{aligned} \tag{2.19}$$

The inverses $\boldsymbol{E}_{ij} = (\boldsymbol{1} - \boldsymbol{D}_i\boldsymbol{D}_j)^{-1}$ with $i, j \in \{P, K\}$ must exist for $\boldsymbol{P}$ and $\boldsymbol{K}$ to be placed in a feedback loop due to (2.5) and $\boldsymbol{D}_P = \boldsymbol{P}(\infty), \boldsymbol{D}_K = \boldsymbol{K}(\infty)$. A useful relation is $\boldsymbol{E}_{ij}\boldsymbol{D}_i\boldsymbol{D}_j = \boldsymbol{D}_i\boldsymbol{E}_{ji}\boldsymbol{D}_j = \boldsymbol{D}_i\boldsymbol{D}_j\boldsymbol{E}_{ij} = \boldsymbol{E}_{ij} - \boldsymbol{1}$ which will be applied throughout this text.

2.4 Fixed Order Controller Parametrization

This section shows why the controller parametrization in transfer function representation is not well adapted to problems formulated in state space representations and presents a new fixed order parametrization derived from the set of all linear controllers in state space representation that factorize the loop formed with a linear plant on a ring R_Z.

2.4.1 Need for a new Parametrization

Transferring the controller parametrization to state space description reveals a practical problem. The loop decomposition (2.15) shows that by starting with an initial controller $\boldsymbol{K}(\boldsymbol{0})$ that leads to an initial loop $\boldsymbol{H}(\boldsymbol{0})$, all possible loops are found by varying a parameter $\boldsymbol{Q}$. This means, in order to obtain a new dynamic in $\boldsymbol{H}(\boldsymbol{Q})$, the initial dynamic of $\boldsymbol{H}(\boldsymbol{0})$ has to be cancelled out by $\boldsymbol{Q}$.

Cancellations of poles and zeros in transfer function representation correspond to reductions of the state dimension in state space representation. If a system's state dimension can be reduced, then it is not minimal. Therefore $\boldsymbol{H}(\boldsymbol{Q})$ is in general not minimal. To give an idea of the state dimensions encountered, look at the way a new controller $\boldsymbol{K}(\boldsymbol{Q})$ is constructed from an initial $\boldsymbol{K}(\boldsymbol{0})$ as given in (2.7) and (2.8).

The new controller's factors $\boldsymbol{X}_i, \boldsymbol{Y}_i$ with $i \in \{L, R\}$ are obtained by (2.8) which is a serial connection between the plant's factors $\boldsymbol{M}_i, \boldsymbol{N}_i$ of state dimension p_P and the parameter $\boldsymbol{Q}$ of state dimension p_Q and then a parallel connection of the result with the initial controller's factors $\boldsymbol{X}_{i0}, \boldsymbol{Y}_{i0}$ of state dimension p_{K0}. According to the formulas for serial (C.46) and parallel (C.44) connection, each such connection has the state of the connecting systems combined so that $p_{Xi,Yi} = p_P + p_{K0} + p_Q$. The formulas for right (C.59) and left (C.61) factorization show that the state dimension of $\boldsymbol{K}(\boldsymbol{Q})$ is the same as its factors so $p_K = p_P + p_{K0} + p_Q$. With the formula for the feedback loop (C.48), the state dimension of $\boldsymbol{H}(\boldsymbol{Q})$ is $p_H = 2p_P + p_{K0} + p_Q$. Using this parametrization in state space description therefore lead to very large matrices that have to be treated by the multiobjective optimization stage later on, which is not desirable. Thus, in order to obtain reduced order controllers directly, the state space expansion to large non minimal systems has to be avoided if possible.

The use of the parameter $\boldsymbol{Q}$ to parametrize the solutions of the doubly coprime factorization (2.7) is therefore adapted to the transfer function representation. To find the new parametrization adapted to the state space representation, the doubly coprime factorization has to be solved directly in state space form. Then it is shown in the following that among all possible solutions there exist fixed order solutions so that the loop $\boldsymbol{H}$ has a factorization (2.13) with fixed order factors. Reduced order controllers can then be found by fixing the controller state dimension to a value lower or equal to the plant state dimension so that $p_K \leq p_P$. Since reduced controllers are a subset of all controllers, this comes at the cost of not parametrizing all possible controllers.

2.4.2 Fixed Order Loop Factorization

Definition For a given plant $\boldsymbol{P}$ with realization $(\boldsymbol{A}_P, \boldsymbol{B}_P, \boldsymbol{C}_P, \boldsymbol{D}_P)$, all controllers $\boldsymbol{K} \in K(\boldsymbol{P})$ with a state space representation of fixed dimension p_K form the set of *p_K-factorizing controllers*

$$\boldsymbol{K} \in K_{p_K}(\boldsymbol{A}_P, \boldsymbol{B}_P, \boldsymbol{C}_P, \boldsymbol{D}_P) \Leftrightarrow \boldsymbol{K} \in K(\boldsymbol{P}) \text{ with state dimension } p_K \quad (2.20)$$

Proposition For a given plant $\boldsymbol{P}$ with realization $(\boldsymbol{A}_P, \boldsymbol{B}_P, \boldsymbol{C}_P, \boldsymbol{D}_P)$, a controller $\boldsymbol{K} \in K(\boldsymbol{P})$ is a p_K-factorizing controller if and only if at least one of the following Riccati equations has a solution

$$\boldsymbol{0} = \begin{bmatrix} -\boldsymbol{R}_{KP} & \boldsymbol{1} \end{bmatrix} \boldsymbol{A}_H \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{R}_{KP} \end{bmatrix} \quad (2.21)$$

$$\boldsymbol{0} = \begin{bmatrix} \boldsymbol{1} & -\boldsymbol{R}_{PK} \end{bmatrix} \boldsymbol{A}_H \begin{bmatrix} \boldsymbol{R}_{PK} \\ \boldsymbol{1} \end{bmatrix} \quad (2.22)$$

If $\boldsymbol{R}_{KP}$ exists, the control loop $\boldsymbol{H}$ has a fixed order loop factorization in the form of

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \text{ with} \tag{2.23}$$

$$\left[\begin{array}{c} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P & \boldsymbol{B}_P\boldsymbol{U}_P \\ \hline \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P & \boldsymbol{D}_P\boldsymbol{U}_P \\ \boldsymbol{F}_P & \boldsymbol{U}_P \end{array}\right]$$

$$\left[\begin{array}{cc} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{array}\right] = \left[\begin{array}{c|cc} \boldsymbol{A}_K + \boldsymbol{L}_K\boldsymbol{C}_K & \boldsymbol{L}_K & \boldsymbol{B}_K + \boldsymbol{L}_K\boldsymbol{D}_K \\ \hline \boldsymbol{V}_K\boldsymbol{C}_K & \boldsymbol{V}_K & \boldsymbol{V}_K\boldsymbol{D}_K \end{array}\right]$$

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{L}_K \\ \boldsymbol{F}_P & \boldsymbol{U}_P\boldsymbol{V}_K \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} -\boldsymbol{R}_{KP} & \boldsymbol{1} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \boldsymbol{C}_H & \boldsymbol{D}_H \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{R}_{KP} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{0} \end{bmatrix} \end{bmatrix}$$

If $\boldsymbol{R}_{PK}$ exists, the control loop $\boldsymbol{H}$ has a fixed order loop factorization in the form of

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_L & \boldsymbol{N}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{0} \end{bmatrix} \text{ with} \tag{2.24}$$

$$\left[\begin{array}{cc} \boldsymbol{M}_L & \boldsymbol{N}_L \end{array}\right] = \left[\begin{array}{c|cc} \boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P & \boldsymbol{B}_P + \boldsymbol{L}_P\boldsymbol{D}_P & \boldsymbol{L}_P \\ \hline \boldsymbol{V}_P\boldsymbol{C}_P & \boldsymbol{V}_P\boldsymbol{D}_P & \boldsymbol{V}_P \end{array}\right]$$

$$\left[\begin{array}{c} \boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{F}_K & \boldsymbol{B}_K\boldsymbol{U}_K \\ \hline \boldsymbol{F}_K & \boldsymbol{U}_K \\ \boldsymbol{C}_K + \boldsymbol{D}_K\boldsymbol{F}_K & \boldsymbol{D}_K\boldsymbol{U}_K \end{array}\right]$$

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{L}_P \\ \boldsymbol{F}_K & \boldsymbol{U}_K\boldsymbol{V}_P \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{1} & -\boldsymbol{R}_{PK} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \boldsymbol{C}_H & \boldsymbol{D}_H \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{R}_{PK} \\ \boldsymbol{1} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{1} \end{bmatrix} \end{bmatrix}$$

The matrices $\boldsymbol{A}_H, \boldsymbol{B}_H, \boldsymbol{C}_H, \boldsymbol{D}_H$ are the state space matrices of the loop $\boldsymbol{H}$ as given in (2.19).

Proof Condition (2.21) can be verified by using (C.59) and (C.61) to obtain the right and left factors of plant $\boldsymbol{P}$ and controller $\boldsymbol{K}$ of state dimension p_P and p_K respectively, see [7] for details. All other factors of that dimension are related to these factors by change of basis, the freedom which will be accounted for by using the basis transformation $\boldsymbol{T}$ further below. Using these factors, the loop factorization can be given by

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{M}_R \\ \boldsymbol{N}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_L & \boldsymbol{X}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right] =$$

$$= \left[\begin{array}{cc|cc} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P & \boldsymbol{B}_P\boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{C}_K & \boldsymbol{B}_P\boldsymbol{U}_P\boldsymbol{V}_K & \boldsymbol{B}_P\boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{D}_K \\ \boldsymbol{0} & \boldsymbol{A}_K + \boldsymbol{L}_K\boldsymbol{C}_K & \boldsymbol{L}_K & \boldsymbol{B}_K + \boldsymbol{L}_K\boldsymbol{D}_K \\ \hline \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P & \boldsymbol{D}_P\boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{C}_K & \boldsymbol{D}_P\boldsymbol{U}_P\boldsymbol{V}_K & \boldsymbol{D}_P\boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{D}_K + \boldsymbol{1} \\ \boldsymbol{F}_P & \boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{C}_K & \boldsymbol{U}_P\boldsymbol{V}_K & \boldsymbol{U}_P\boldsymbol{V}_K\boldsymbol{D}_K \end{array}\right]$$

using (C.46). The general non factorized feedback loop is given by (2.19) so that

$$\boldsymbol{H} = \left[\begin{array}{c|c} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \hline \boldsymbol{C}_H & \boldsymbol{D}_H \end{array}\right]$$

These two state space representations of $\boldsymbol{H}$ have the same state dimension and must therefore be equivalent in the sense of (C.35) for the controller to be a member of $K(\boldsymbol{P})$. Then according to (C.36) there must exist an invertible basis transformation matrix $\boldsymbol{T}$ so that $\boldsymbol{TA} = \boldsymbol{A}_H\boldsymbol{T}, \boldsymbol{TB} = \boldsymbol{B}_H, \boldsymbol{C} = \boldsymbol{C}_H\boldsymbol{T}, \boldsymbol{D} = \boldsymbol{D}_H$ which can now be calculated using the partitioning

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X} \\ \boldsymbol{Y} & \boldsymbol{Z} \end{bmatrix}$$

The condition $\boldsymbol{D} = \boldsymbol{D}_H$ provides

$$\boldsymbol{U}_P\boldsymbol{V}_K = \boldsymbol{E}_{KP} \tag{2.25}$$

The condition $\boldsymbol{C} = \boldsymbol{C}_H\boldsymbol{T}$ provides

$$\begin{aligned} \boldsymbol{C}_P\boldsymbol{W} &= \boldsymbol{C}_P \\ \boldsymbol{C}_P\boldsymbol{X} &= \boldsymbol{0} \\ \boldsymbol{C}_K\boldsymbol{Z} &= \boldsymbol{C}_K \\ \boldsymbol{F}_P &= \boldsymbol{E}_{KP}(\boldsymbol{D}_K\boldsymbol{C}_P + \boldsymbol{C}_K\boldsymbol{Y}) \end{aligned} \tag{2.26}$$

The condition $\boldsymbol{TB} = \boldsymbol{B}_H$ provides

$$\begin{aligned} \boldsymbol{X}\boldsymbol{B}_K &= \boldsymbol{0} \\ \boldsymbol{Z}\boldsymbol{B}_K &= \boldsymbol{B}_K \\ \boldsymbol{X}\boldsymbol{L}_K &= (\boldsymbol{1} - \boldsymbol{W})\boldsymbol{B}_P\boldsymbol{E}_{KP} \\ \boldsymbol{Z}\boldsymbol{L}_K &= (\boldsymbol{B}_K\boldsymbol{D}_P - \boldsymbol{Y}\boldsymbol{B}_P)\boldsymbol{E}_{KP} \end{aligned} \tag{2.27}$$

The condition $\boldsymbol{TA} = \boldsymbol{A}_H\boldsymbol{T}$ provides

$$\begin{aligned} \boldsymbol{X}\boldsymbol{A}_K &= \boldsymbol{A}_P\boldsymbol{X} \\ \boldsymbol{Z}\boldsymbol{A}_K &= \boldsymbol{A}_K\boldsymbol{Z} \\ \boldsymbol{W}\boldsymbol{A}_P &= \boldsymbol{A}_P\boldsymbol{W} + (\boldsymbol{1} - \boldsymbol{W})\boldsymbol{B}_P\boldsymbol{F}_P \end{aligned} \tag{2.28}$$

and

$$\begin{aligned} \boldsymbol{Y}(\boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P) + \boldsymbol{Y}\boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K\boldsymbol{Y} = \\ = \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P + (\boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K)\boldsymbol{Y} \end{aligned} \tag{2.29}$$

using (2.26) and (2.27). If there exists a solution $\boldsymbol{R}_{KP} = \boldsymbol{Y}$ to this Riccati equation, then a solution to (2.25), (2.26), (2.27), (2.28) also exists in the form of

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{1} \\ \boldsymbol{X} &= \boldsymbol{0} \\ \boldsymbol{Y} &= \boldsymbol{R}_{KP} \\ \boldsymbol{Z} &= \boldsymbol{1} \\ \boldsymbol{U}_P\boldsymbol{V}_K &= \boldsymbol{E}_{KP} \\ \boldsymbol{F}_P &= \boldsymbol{E}_{KP}(\boldsymbol{D}_K\boldsymbol{C}_P + \boldsymbol{C}_K\boldsymbol{R}_{KP}) \\ \boldsymbol{L}_K &= (\boldsymbol{B}_K\boldsymbol{D}_P - \boldsymbol{R}_{KP}\boldsymbol{B}_P)\boldsymbol{E}_{KP} \end{aligned}$$

Hence $\boldsymbol{T}$ is invertible so that the state space representations $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D})$ and $(\boldsymbol{A}_H,\boldsymbol{B}_H,\boldsymbol{C}_H,\boldsymbol{D}_H)$ are equivalent and thus $\boldsymbol{H}$ can be factorized in the form of (2.23) by a controller of state dimension p_K. This verifies that (2.21) is the condition for $\boldsymbol{K}$ to be a member of $K_{p_K}(\boldsymbol{A}_P,\boldsymbol{B}_P,\boldsymbol{C}_P,\boldsymbol{D}_P) \subset K(\boldsymbol{P})$. The proof of condition (2.22) is similar.

Remark 1 The Riccati equations (2.21) and (2.22) can also be written in the Radon form

$$\begin{aligned}\boldsymbol{A}_H \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{R}_{KP}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{R}_{KP}\end{bmatrix}(\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{F}_P)\\ \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{R}_{PK}\end{bmatrix}\boldsymbol{A}_H &= (\boldsymbol{A}_P+\boldsymbol{L}_P\boldsymbol{C}_P)\begin{bmatrix}\boldsymbol{1} & -\boldsymbol{R}_{PK}\end{bmatrix}\end{aligned}$$

which follows by calculating

$$\begin{aligned}\boldsymbol{A}_H \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{R}_{KP}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P+\boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K\boldsymbol{R}_{KP}\\ \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P+(\boldsymbol{A}_K+\boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K)\boldsymbol{R}_{KP}\end{bmatrix} =\\ &= \begin{bmatrix}\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{E}_{KP}(\boldsymbol{D}_K\boldsymbol{C}_P+\boldsymbol{C}_K\boldsymbol{R}_{KP})\\ \boldsymbol{R}_{KP}(\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P)+\boldsymbol{R}_{KP}\boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K\boldsymbol{R}_{KP}\end{bmatrix} =\\ &= \begin{bmatrix}\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{F}_P\\ \boldsymbol{R}_{KP}(\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{F}_P)\end{bmatrix} = \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{R}_{KP}\end{bmatrix}(\boldsymbol{A}_P+\boldsymbol{B}_P\boldsymbol{F}_P)\end{aligned}$$

using $\boldsymbol{F}_P = \boldsymbol{E}_{KP}(\boldsymbol{D}_K\boldsymbol{C}_P+\boldsymbol{C}_K\boldsymbol{R}_{KP})$ from (2.23) and the Riccati equation (2.21). The proof for the second line is similar, this time using $\boldsymbol{L}_P = (\boldsymbol{B}_P\boldsymbol{D}_K - \boldsymbol{R}_{PK}\boldsymbol{B}_K)\boldsymbol{E}_{PK}$ from (2.24) and the Riccati equation (2.22).

Remark 2 The existence of solutions of (2.21) and (2.22) does not depend on the selected state space basis of plant or controller which can be seen by taking all equivalent state space representations of $\boldsymbol{P},\boldsymbol{K}$ in the form of

$$\boldsymbol{P} = \left[\begin{array}{c|c}\boldsymbol{T}_P\boldsymbol{A}_P\boldsymbol{T}_P^{-1} & \boldsymbol{T}_P\boldsymbol{B}_P\\ \hline \boldsymbol{C}_P\boldsymbol{T}_P^{-1} & \boldsymbol{D}_P\end{array}\right],\ \boldsymbol{K} = \left[\begin{array}{c|c}\boldsymbol{T}_K\boldsymbol{A}_K\boldsymbol{T}_K^{-1} & \boldsymbol{T}_K\boldsymbol{B}_K\\ \hline \boldsymbol{C}_K\boldsymbol{T}_K^{-1} & \boldsymbol{D}_K\end{array}\right]$$

with $\boldsymbol{T}_P,\boldsymbol{T}_K$ invertible. This leads to the loop dynamic matrix

$$\boldsymbol{A}_H(\boldsymbol{T}_P,\boldsymbol{T}_K) = \begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}\boldsymbol{A}_H\begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}^{-1}$$

with $\boldsymbol{A}_H$ given in (2.19). The Riccati equations (2.21), (2.22) can then be written as

$$\begin{aligned}\boldsymbol{0} &= \begin{bmatrix}-\boldsymbol{R}_{KP} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}^{-1}\boldsymbol{A}_H(\boldsymbol{T}_P,\boldsymbol{T}_K)\begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{R}_{KP}\end{bmatrix}\\ \boldsymbol{0} &= \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{R}_{PK}\end{bmatrix}\begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}^{-1}\boldsymbol{A}_H(\boldsymbol{T}_P,\boldsymbol{T}_K)\begin{bmatrix}\boldsymbol{T}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{T}_K\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_{PK}\\ \boldsymbol{1}\end{bmatrix}\end{aligned}$$

Since $\boldsymbol{T}_P,\boldsymbol{T}_K$ are invertible, this can be reshaped to

$$\begin{aligned}\boldsymbol{0} &= \begin{bmatrix}-\boldsymbol{T}_K\boldsymbol{R}_{KP}\boldsymbol{T}_P^{-1} & \boldsymbol{1}\end{bmatrix}\boldsymbol{A}_H(\boldsymbol{T}_P,\boldsymbol{T}_K)\begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{T}_K\boldsymbol{R}_{KP}\boldsymbol{T}_P^{-1}\end{bmatrix}\\ \boldsymbol{0} &= \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{T}_P\boldsymbol{R}_{PK}\boldsymbol{T}_K^{-1}\end{bmatrix}\boldsymbol{A}_H(\boldsymbol{T}_P,\boldsymbol{T}_K)\begin{bmatrix}\boldsymbol{T}_P\boldsymbol{R}_{PK}\boldsymbol{T}_K^{-1}\\ \boldsymbol{1}\end{bmatrix}\end{aligned}$$

Hence the solutions of the transformed Riccati equations are given by

$$\boldsymbol{R}_{KP}(\boldsymbol{T}_P, \boldsymbol{T}_K) = \boldsymbol{T}_K \boldsymbol{R}_{KP} \boldsymbol{T}_P^{-1}$$
$$\boldsymbol{R}_{PK}(\boldsymbol{T}_P, \boldsymbol{T}_K) = \boldsymbol{T}_P \boldsymbol{R}_{PK} \boldsymbol{T}_K^{-1}$$

so that these solutions exist if and only if $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ exist.

2.4.3 Existence of Solutions

Proposition A sufficient condition for the solvability of both Riccati equations (2.21) and (2.22) is given by

$$\boldsymbol{A}_H \text{ is diagonalizable} \Rightarrow \boldsymbol{R}_{PK} \text{ and } \boldsymbol{R}_{KP} \text{ exist} \tag{2.30}$$

with $\boldsymbol{A}_H$ being the dynamic matrix of the loop $\boldsymbol{H}$ as given in (2.19). In this case both Riccati equations (2.21) and (2.22) have solutions and $\boldsymbol{K}$ admits both loop factorizations (2.23) and (2.24). Some examples are given after the proof to illustrate the situation if $\boldsymbol{A}_H$ is not diagonalizable.

Proof Since $\boldsymbol{A}_H$ is diagonalizable we can write

$$\boldsymbol{A}_H \boldsymbol{V} = \boldsymbol{V} \boldsymbol{\Lambda} \tag{2.31}$$

with invertible $\boldsymbol{V}$ containing the eigenvectors of $\boldsymbol{A}_H$ and diagonal $\boldsymbol{\Lambda}$ containing the eigenvalues of $\boldsymbol{A}_H$. Since $\boldsymbol{V}$ is invertible, it can be arranged in a way

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{PP} & \boldsymbol{V}_{PK} \\ \boldsymbol{V}_{KP} & \boldsymbol{V}_{KK} \end{bmatrix}$$

with $\boldsymbol{V}_{PP}$ invertible. Since $\boldsymbol{\Lambda}$ is diagonal we can now write (2.31) in the form

$$\boldsymbol{A}_H \begin{bmatrix} \boldsymbol{V}_{PP} & \boldsymbol{V}_{PK} \\ \boldsymbol{V}_{KP} & \boldsymbol{V}_{KK} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_{PP} & \boldsymbol{V}_{PK} \\ \boldsymbol{V}_{KP} & \boldsymbol{V}_{KK} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_K \end{bmatrix}$$

Right projection with $\begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \end{bmatrix}^T$ leads to

$$\boldsymbol{A}_H \begin{bmatrix} \boldsymbol{V}_{PP} \\ \boldsymbol{V}_{KP} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_{PP} & \boldsymbol{V}_{PK} \\ \boldsymbol{V}_{KP} & \boldsymbol{V}_{KK} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_P \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_{PP} \\ \boldsymbol{V}_{KP} \end{bmatrix} \boldsymbol{\Lambda}_P$$

Right multiplication with $\boldsymbol{V}_{PP}^{-1}$ gives

$$\boldsymbol{A}_H \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1} \end{bmatrix} \boldsymbol{V}_{PP} \boldsymbol{\Lambda}_P \boldsymbol{V}_{PP}^{-1}$$

Left projection with $\begin{bmatrix} -\boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1} & \boldsymbol{1} \end{bmatrix}$ leads to the Riccati equation

$$\begin{bmatrix} -\boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1} & \boldsymbol{1} \end{bmatrix} \boldsymbol{A}_H \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1} \end{bmatrix} = \boldsymbol{0}$$

which is identical to (2.21) so that a solution $\boldsymbol{R}_{KP} = \boldsymbol{V}_{KP} \boldsymbol{V}_{PP}^{-1}$ has been found. The proof for $\boldsymbol{R}_{PK}$ is similar.

Example 1 Let a plant $\boldsymbol{P}$ and a controller $\boldsymbol{K}$ be given by

$$\boldsymbol{P} = \left[\begin{array}{c|cc} \lambda+1 & 1 & 0 \\ \hline 1 & 0 & 1 \end{array}\right],\ \boldsymbol{K} = \left[\begin{array}{c|c} \lambda-1 & 1 \\ \hline 1 & -1 \\ 1 & 0 \end{array}\right]$$

The matrices

$$\boldsymbol{E}_{PK} = 1,\ \boldsymbol{E}_{KP} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

are invertible so the loop $\boldsymbol{H}$ is well connected. Its dynamic matrix is then given by

$$\boldsymbol{A}_H = \begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$$

Of the Riccati equations (2.21) and (2.22) that simplify to

$$\begin{aligned} 0 &= 1 \\ 0 &= \boldsymbol{R}_{PK}^2 \end{aligned}$$

only the second one has a solution

$$\boldsymbol{R}_{PK} = 0$$

This is therefore an example of a controller that is a member of $K_1(\boldsymbol{P}) \subset K(\boldsymbol{P})$ albeit one that admits only one loop factorization.

Example 2 Let a plant $\boldsymbol{P}$ and a controller $\boldsymbol{K}$ be given by

$$\boldsymbol{P} = \left[\begin{array}{cc|cc} -1 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ \hline 1 & -6 & 0 & 3 \\ 0 & 1 & 0 & 0 \end{array}\right],\ \boldsymbol{K} = \left[\begin{array}{c|cc} -3 & 1 & 0 \\ \hline -3 & 1 & 0 \\ 1 & 0 & 2 \end{array}\right]$$

The matrices

$$\boldsymbol{E}_{PK} = \begin{bmatrix} 1 & 6 \\ 0 & 1 \end{bmatrix},\ \boldsymbol{E}_{KP} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

are invertible so the loop $\boldsymbol{H}$ is well connected. Its dynamic matrix is then given by

$$\boldsymbol{A}_H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

However the Riccati equations (2.21) and (2.22) that simplify to

$$\begin{aligned} \boldsymbol{0} &= \boldsymbol{R}_{KP} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \boldsymbol{R}_{KP} - \begin{bmatrix} 1 & 0 \end{bmatrix} \\ \boldsymbol{0} &= \boldsymbol{R}_{PK} \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{R}_{PK} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

have both no solution. This is therefore an example of a controller that is not a member of $K_1(\boldsymbol{P})$ since it admits no loop factorization.

Example 3 Let a plant $\boldsymbol{P}$ and a controller $\boldsymbol{K}$ be given by

$$\boldsymbol{P} = \left[\begin{array}{cc|cc} -1 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ \hline 1 & -6 & 0 & 3 \\ 0 & 1 & 0 & 0 \end{array}\right], \quad \boldsymbol{K} = \left[\begin{array}{cc|cc} -3 & 0 & 1 & 0 \\ 0 & \lambda & 0 & 1 \\ \hline -3 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{array}\right]$$

The plant and the matrices $\boldsymbol{E}_{PK}, \boldsymbol{E}_{KP}$ are identical to the ones in example 2 and the controller has the same transfer function than the one in example 2, but an unobservable state has been added to the controller. The loop dynamic matrix is now given by

$$\boldsymbol{A}_H = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \lambda \end{bmatrix}$$

Of the Riccati equations (2.21) and (2.22) that simplify to

$$\begin{aligned} \mathbf{0} &= \boldsymbol{R}_{KP} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \boldsymbol{R}_{KP} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \lambda \end{bmatrix} \boldsymbol{R}_{KP} \\ \mathbf{0} &= \boldsymbol{R}_{PK}^2 - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \boldsymbol{R}_{PK} \begin{bmatrix} 0 & 0 \\ 0 & \lambda \end{bmatrix} \end{aligned}$$

the second one has a solution if $\lambda \neq 0$

$$\boldsymbol{R}_{PK} = \begin{bmatrix} 0 & 0 \\ -\frac{1}{\lambda} & -\lambda \end{bmatrix}$$

This is therefore an example of a controller that is a member of $K_2(\boldsymbol{P}) \subset K(\boldsymbol{P})$. This stands in contrast to example 2 since the only difference is an unobservable state with a new eigenvalue $\lambda \neq 0$ that is added to the controller, while the transfer functions are all exactly as in example 2. Together with the result of example 2 this means that $\boldsymbol{K}$ is a member of $K_2(\boldsymbol{P}) \setminus K_1(\boldsymbol{P})$ and therefore $\boldsymbol{K}$ is a loop factorizing controller but the dimension of its state space representation given in example 2 was too low. Thus a controller in state space representation can become loop factorizing by increasing its state dimension.

2.4.4 Relation to Estimator Corrector Type Controllers

An *estimator corrector* type control is characterized by exhibiting the *separation principle* which states that the control task can be separated into a state estimator part and a state corrector part and both parts can be designed independently, a fact that is reflected by the eigenvalues of the loop dynamic matrix $\boldsymbol{A}_H$ being the union of the eigenvalues independently determined by the estimator and corrector stage. In the following it is shown that the fixed order factorized loop resembles an estimator corrector type controlled loop in that an estimator mismatch $\boldsymbol{\epsilon}$ is formed between the controller part of the state vector $\boldsymbol{x}_K$ and a projection of the plant part of the state vector $\boldsymbol{x}_P$ in the form of $\boldsymbol{\epsilon} = \dot{\boldsymbol{x}}_K - \boldsymbol{R}_{KP}\dot{\boldsymbol{x}}_P$ in the case of factorization (2.23) or with indices exchanged in the case of factorization (2.24). Despite the similarities, it does however not exhibit the separation principle, so that the notation *general estimator corrector* seems more appropriate.

Proposition For a given plant $\boldsymbol{P}$, a controller

$$\boldsymbol{K} \text{ is of the general estimator corrector type} \Leftrightarrow \boldsymbol{R}_{KP} \text{ or } \boldsymbol{R}_{PK} \text{ exist} \quad (2.32)$$

with $\boldsymbol{R}_{KP}$ and $\boldsymbol{R}_{PK}$ being the solutions of the Riccati equations (2.21) and (2.22).

Proof Assume the existence of $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$. Then the basis of the loop's state space representation $\boldsymbol{H}$ as given in (2.19) can be changed using matrices

$$\boldsymbol{T}_{KP} = \begin{bmatrix} \mathbf{1} & \boldsymbol{R}_{KP} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \; \boldsymbol{T}_{PK} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{R}_{PK} & \mathbf{1} \end{bmatrix}$$

so that

$$\boldsymbol{H} = \left[\begin{array}{c|c} \boldsymbol{T}_{KP}^{-1}\boldsymbol{A}_H\boldsymbol{T}_{KP} & \boldsymbol{T}_{KP}^{-1}\boldsymbol{B}_H \\ \hline \boldsymbol{C}_H\boldsymbol{T}_{KP} & \boldsymbol{D}_H \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{T}_{PK}^{-1}\boldsymbol{A}_H\boldsymbol{T}_{PK} & \boldsymbol{T}_{PK}^{-1}\boldsymbol{B}_H \\ \hline \boldsymbol{C}_H\boldsymbol{T}_{PK} & \boldsymbol{D}_H \end{array}\right]$$

Using (2.23) and (2.24) this leads to the following equivalent state space representations of the closed loop system

$$\begin{aligned} \begin{bmatrix} \dot{\boldsymbol{x}}_K - \boldsymbol{R}_{KP}\dot{\boldsymbol{x}}_P \\ \dot{\boldsymbol{x}}_P \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_K + \boldsymbol{L}_K\boldsymbol{C}_K & \mathbf{0} \\ \boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_K - \boldsymbol{R}_{KP}\boldsymbol{x}_P \\ \boldsymbol{x}_P \end{bmatrix} + \\ &+ \begin{bmatrix} \boldsymbol{B}_K + \boldsymbol{L}_K\boldsymbol{D}_K & \boldsymbol{L}_K \\ \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} & \boldsymbol{B}_P\boldsymbol{E}_{KP} \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{v} \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{F}_P \\ \boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_K - \boldsymbol{R}_{KP}\boldsymbol{x}_P \\ \boldsymbol{x}_P \end{bmatrix} + \\ &+ \begin{bmatrix} \boldsymbol{D}_K\boldsymbol{E}_{PK} & \boldsymbol{E}_{KP} \\ \boldsymbol{E}_{PK} & \boldsymbol{D}_P\boldsymbol{E}_{KP} \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{v} \end{bmatrix} \end{aligned} \quad (2.33)$$

for $\boldsymbol{R}_{KP}$ and

$$\begin{aligned} \begin{bmatrix} \dot{\boldsymbol{x}}_P - \boldsymbol{R}_{PK}\dot{\boldsymbol{x}}_K \\ \dot{\boldsymbol{x}}_K \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P & \mathbf{0} \\ \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{F}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_P - \boldsymbol{R}_{PK}\boldsymbol{x}_K \\ \boldsymbol{x}_K \end{bmatrix} + \\ &+ \begin{bmatrix} \boldsymbol{B}_P + \boldsymbol{L}_P\boldsymbol{D}_P & \boldsymbol{L}_P \\ \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP} & \boldsymbol{B}_K\boldsymbol{E}_{PK} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{F}_K \\ \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{C}_K + \boldsymbol{D}_K\boldsymbol{F}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_P - \boldsymbol{R}_{PK}\boldsymbol{x}_K \\ \boldsymbol{x}_K \end{bmatrix} + \\ &+ \begin{bmatrix} \boldsymbol{D}_P\boldsymbol{E}_{KP} & \boldsymbol{E}_{PK} \\ \boldsymbol{E}_{KP} & \boldsymbol{D}_K\boldsymbol{E}_{PK} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \end{aligned} \quad (2.34)$$

for $\boldsymbol{R}_{PK}$. It can be seen by the structure of the transformed closed loop system (2.33), that the fixed order factorization describes a form of general estimator corrector type control in which the plant's state $\boldsymbol{x}_P$ is projected onto the range of $\boldsymbol{P}_{KP}$ and then compared to the controller's state $\boldsymbol{x}_K$. The fact that the estimator mismatch $\boldsymbol{x}_K - \boldsymbol{R}_{KP}\boldsymbol{x}_P$ becomes an independent state of $\boldsymbol{H}$ due to the block triangular structure of the dynamic matrix $\boldsymbol{A}_H$ already shows the separation property of estimator corrector type control. However, unlike with actual estimator corrector type control, the matrices $\boldsymbol{F}_P, \boldsymbol{L}_K$ are in general not independent design variables but are rather both a function of $\boldsymbol{R}_{KP}$ according to

(2.23), so the separation principle that helps in the design of common estimator corrector type controllers is in general not valid. The same is true in the dual case (2.34) for $\boldsymbol{F}_K, \boldsymbol{L}_P$ that are a function of $\boldsymbol{R}_{PK}$ according to (2.24).

Proposition For a given plant $\boldsymbol{P}$, a controller

$$\boldsymbol{K} \text{ is of the estimator corrector type} \Leftrightarrow \boldsymbol{R}_{KP} \text{ and } \boldsymbol{R}_{PK} \text{ invertible} \tag{2.35}$$

with $\boldsymbol{R}_{KP}$ and $\boldsymbol{R}_{PK}$ being the solutions of the Riccati equations (2.21) and (2.22).

Proof Assume the existence of an invertible $\boldsymbol{R}_{PK}$. Then $\boldsymbol{R}_{PK}$ solves

$$\begin{bmatrix}\mathbf{1} & -\boldsymbol{R}_{PK}\end{bmatrix} \boldsymbol{A}_H \begin{bmatrix}\boldsymbol{R}_{PK} \\ \mathbf{1}\end{bmatrix} = \mathbf{0}$$

according to (2.24). Left multiplication by $-\boldsymbol{R}_{PK}^{-1}$ and right multiplication by $\boldsymbol{R}_{PK}^{-1}$ leads to

$$\begin{bmatrix}-\boldsymbol{R}_{PK}^{-1} & \mathbf{1}\end{bmatrix} \boldsymbol{A}_H \begin{bmatrix}\mathbf{1} \\ \boldsymbol{R}_{PK}^{-1}\end{bmatrix} = \mathbf{0}$$

Comparing this to (2.23) shows that there exists a solution $\boldsymbol{R}_{KP} = \boldsymbol{R}_{PK}^{-1}$. This also works in the reverse direction so that $\boldsymbol{R}_{PK}^{-1}$ exists if and only if $\boldsymbol{R}_{KP}^{-1}$ exists. The state space basis transformation $\boldsymbol{x}'_K = \boldsymbol{R}_{PK}\boldsymbol{x}_K$ with $\boldsymbol{R}_{KP} = \boldsymbol{R}_{PK}^{-1}$ and accordingly

$$\begin{bmatrix}\boldsymbol{A}'_K & \boldsymbol{B}'_K \\ \boldsymbol{C}'_K & \boldsymbol{D}_K\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{PK} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K\end{bmatrix} \begin{bmatrix}\boldsymbol{R}_{KP} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}\end{bmatrix}$$

together with $\boldsymbol{F}'_K = \boldsymbol{F}_K\boldsymbol{R}_{KP}$ and $\boldsymbol{L}'_K = \boldsymbol{R}_{PK}\boldsymbol{L}_K$ leads to

$$\begin{aligned}
\boldsymbol{F}_P &= \boldsymbol{E}_{KP}(\boldsymbol{D}_K\boldsymbol{C}_P + \boldsymbol{C}'_K) \\
\boldsymbol{L}_P &= (\boldsymbol{B}_P\boldsymbol{D}_K - \boldsymbol{B}'_K)\boldsymbol{E}_{PK} \\
\boldsymbol{F}'_K &= \boldsymbol{E}_{PK}(\boldsymbol{D}_P\boldsymbol{C}'_K + \boldsymbol{C}_P) \\
\boldsymbol{L}'_K &= (\boldsymbol{B}'_K\boldsymbol{D}_P - \boldsymbol{B}_P)\boldsymbol{E}_{KP}
\end{aligned}$$

using (2.23), (2.24). This further gives

$$\begin{aligned}
\boldsymbol{A}'_K + \boldsymbol{B}'_K\boldsymbol{F}'_K &= \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P \\
\boldsymbol{A}'_K + \boldsymbol{L}'_K\boldsymbol{C}'_K &= \boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P \\
\boldsymbol{B}'_K + \boldsymbol{L}'_K\boldsymbol{D}_K &= -\boldsymbol{L}_P \\
-\boldsymbol{L}'_K &= \boldsymbol{B}_P + \boldsymbol{L}_P\boldsymbol{D}_P \\
\boldsymbol{C}'_K + \boldsymbol{D}_K\boldsymbol{F}'_K &= \boldsymbol{F}_P \\
\boldsymbol{F}'_K &= \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P
\end{aligned}$$

Since $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ are assumed to exist, the state space representations of the closed loop system can be written in the forms (2.33), (2.34) which can now be

given by

$$\begin{aligned}\begin{bmatrix}\dot{\boldsymbol{x}}'_K - \dot{\boldsymbol{x}}_P \\ \dot{\boldsymbol{x}}_P\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P & \boldsymbol{0} \\ \boldsymbol{B}_P(\boldsymbol{F}_P - \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P) & \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P\end{bmatrix}\begin{bmatrix}\boldsymbol{x}'_K - \boldsymbol{x}_P \\ \boldsymbol{x}_P\end{bmatrix} + \\ &+ \begin{bmatrix}-\boldsymbol{L}_P & -(\boldsymbol{B}_P + \boldsymbol{L}_P\boldsymbol{D}_P) \\ \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} & \boldsymbol{B}_P\boldsymbol{E}_{KP}\end{bmatrix}\begin{bmatrix}\boldsymbol{w} \\ \boldsymbol{v}\end{bmatrix} \\ \begin{bmatrix}\boldsymbol{u} \\ \boldsymbol{y}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{F}_P - \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{F}_P \\ \boldsymbol{D}_P(\boldsymbol{F}_P - \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P) & \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P\end{bmatrix}\begin{bmatrix}\boldsymbol{x}'_K - \boldsymbol{x}_P \\ \boldsymbol{x}_P\end{bmatrix} + \\ &+ \begin{bmatrix}\boldsymbol{D}_K\boldsymbol{E}_{PK} & \boldsymbol{E}_{KP} \\ \boldsymbol{E}_{PK} & \boldsymbol{D}_P\boldsymbol{E}_{KP}\end{bmatrix}\begin{bmatrix}\boldsymbol{w} \\ \boldsymbol{v}\end{bmatrix}\end{aligned} \tag{2.36}$$

or alternatively by

$$\begin{aligned}\begin{bmatrix}\dot{\boldsymbol{x}}_P - \dot{\boldsymbol{x}}'_K \\ \dot{\boldsymbol{x}}'_K\end{bmatrix} &= \begin{bmatrix}\boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P & \boldsymbol{0} \\ (\boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} - \boldsymbol{L}_P)\boldsymbol{C}_P & \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_P - \boldsymbol{x}'_K \\ \boldsymbol{x}'_K\end{bmatrix} + \\ &+ \begin{bmatrix}\boldsymbol{B}_P + \boldsymbol{L}_P\boldsymbol{D}_P & \boldsymbol{L}_P \\ (\boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} - \boldsymbol{L}_P)\boldsymbol{D}_P & \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK} - \boldsymbol{L}_P\end{bmatrix}\begin{bmatrix}\boldsymbol{v} \\ \boldsymbol{w}\end{bmatrix} \\ \begin{bmatrix}\boldsymbol{y} \\ \boldsymbol{u}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{C}_P + \boldsymbol{D}_P\boldsymbol{F}_P \\ \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{F}_P\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_P - \boldsymbol{x}'_K \\ \boldsymbol{x}'_K\end{bmatrix} + \\ &+ \begin{bmatrix}\boldsymbol{D}_P\boldsymbol{E}_{KP} & \boldsymbol{E}_{PK} \\ \boldsymbol{E}_{KP} & \boldsymbol{D}_K\boldsymbol{E}_{PK}\end{bmatrix}\begin{bmatrix}\boldsymbol{v} \\ \boldsymbol{w}\end{bmatrix}\end{aligned} \tag{2.37}$$

In contrast to (2.33), (2.34), the matrices $\boldsymbol{L}_P, \boldsymbol{F}_P$ are now independent parameters so that the separation principle is now valid and thus the controller is of the estimator corrector type. The controller can then be given explicitly

$$\begin{aligned}\boldsymbol{A}'_K &= \boldsymbol{A}_P - \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{C}_P + \boldsymbol{B}_P\boldsymbol{E}_{KP}^{-1}\boldsymbol{F}_P + \boldsymbol{L}_P\boldsymbol{E}_{PK}^{-1}\boldsymbol{C}_P + \boldsymbol{L}_P\boldsymbol{D}_P\boldsymbol{E}_{KP}^{-1}\boldsymbol{F}_P \\ \boldsymbol{B}'_K &= \boldsymbol{B}_P\boldsymbol{D}_K - \boldsymbol{L}_P\boldsymbol{E}_{PK}^{-1} \\ \boldsymbol{C}'_K &= \boldsymbol{E}_{KP}^{-1}\boldsymbol{F}_P - \boldsymbol{D}_K\boldsymbol{C}_P\end{aligned} \tag{2.38}$$

as a function of the design parameters $\boldsymbol{L}_P, \boldsymbol{F}_P$. The dynamics of the controller can then be given explicitly as well

$$\begin{aligned}\text{estimator} &\begin{cases}\dot{\boldsymbol{x}}_K = \boldsymbol{A}_P\boldsymbol{x}_K + \boldsymbol{B}_P(\boldsymbol{u} - \boldsymbol{v}) + \boldsymbol{L}(\boldsymbol{y} - \boldsymbol{y}_K) \\ \boldsymbol{y}_K = \boldsymbol{C}_P\boldsymbol{x}_K + \boldsymbol{D}_P\boldsymbol{u}_K\end{cases} \\ \text{corrector} &\begin{cases}\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u}_K + \boldsymbol{D}_K(\boldsymbol{y} - \boldsymbol{y}_K) \\ \boldsymbol{u}_K = \boldsymbol{F}\boldsymbol{x}_K\end{cases}\end{aligned}$$

with $\boldsymbol{F} = \boldsymbol{F}_P, \boldsymbol{L} = -\boldsymbol{L}_P\boldsymbol{E}_{PK}^{-1}$ and it can be seen that $\boldsymbol{K}$ is the generic estimator corrector type controller, consisting of a Luenberger state estimator with arbitrary parameter $\boldsymbol{L}$ and a static state corrector with arbitrary parameter $\boldsymbol{F}$.

Now suppose $\boldsymbol{K}$ is of this generic estimator corrector type

$$\boldsymbol{K} = \left[\begin{array}{c|c}\boldsymbol{A}'_K & \boldsymbol{B}'_K \\ \hline \boldsymbol{C}'_K & \boldsymbol{D}_K\end{array}\right]$$

with $\boldsymbol{A}'_K, \boldsymbol{B}'_K, \boldsymbol{C}'_K$ as given in (2.38) for arbitrary $\boldsymbol{F}_P, \boldsymbol{L}_P, \boldsymbol{D}_K$. Then according to (2.19) and using this generic estimator corrector type controller $\boldsymbol{K}$ the

dynamic matrix of the loop $\boldsymbol{H}$ gives

$$\boldsymbol{A}_H = \begin{bmatrix} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{B}_P\boldsymbol{F}_P - \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P \\ \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P - \boldsymbol{L}_P\boldsymbol{C}_P & \boldsymbol{A}_P + \boldsymbol{B}_P(\boldsymbol{F}_P - \boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P) + \boldsymbol{L}_P\boldsymbol{C}_P \end{bmatrix}$$

Then clearly

$$\begin{bmatrix} -\mathbf{1} & \mathbf{1} \end{bmatrix} \boldsymbol{A}_H \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} = \mathbf{0} = \begin{bmatrix} \mathbf{1} & -\mathbf{1} \end{bmatrix} \boldsymbol{A}_H \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$$

so that the Riccati equations (2.21) and (2.22) thus have the trivial solution $\boldsymbol{R}_{KP} = \mathbf{1} = \boldsymbol{R}_{PK}$. This means that the generic estimator corrector type controller, which is given explicitly in (2.38) and which can therefore be constructed for any given plant $\boldsymbol{P}$, is a factorizing controller of state dimension $p_K = p_P$, hence (2.35) follows.

Remark If the state space representation of a given plant $\boldsymbol{P}$ of dimension p_P is minimal, which is equivalent to it having full McMillan degree $\delta(\boldsymbol{P}) = p_P$, then there always exists a factorizing controller of state space dimension $p_K = p_P$ independent of the selected underlying ring R_Z so that the Riccati equations (2.21) and (2.22) have a solution and

$$\boldsymbol{P} \text{ minimal} \Rightarrow K_{p_P}(\boldsymbol{A}_P, \boldsymbol{B}_P, \boldsymbol{C}_P, \boldsymbol{D}_P) \text{ is nonempty} \tag{2.39}$$

which due to (2.20) implies that $K(\boldsymbol{P})$ is nonempty and the doubly coprime factorization (2.7) has a solution. This stems from the fact that if the state space representation of $\boldsymbol{P}$ is minimal, it is controllable so that the eigenvalues of $\boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}_P$ can be assured to lie in a zone $Z \subset \mathbb{C}$ and it is observable so that the eigenvalues of $\boldsymbol{A}_P + \boldsymbol{L}_P\boldsymbol{C}_P$ can be assured to lie in the same zone Z. Hence, due to the block triangular structure of the loop dynamic matrix in (2.36) and (2.37), $\boldsymbol{H} \in R_Z^{m+n,m+n}$, with R_Z being the ring of proper transfer functions with all poles contained in a predefined zone Z so that (2.39) follows.

2.4.5 Parametrization of Fixed Order Controllers

There is a reversible relation between the controller $\boldsymbol{K}$ and the factors $\boldsymbol{X}_L, \boldsymbol{X}_R$ in the form of $\boldsymbol{K} = \boldsymbol{X}_R\boldsymbol{Y}_R^{-1} = \boldsymbol{Y}_L^{-1}\boldsymbol{X}_L$ which leads to the idea of the following change of variables.

Definition The factors $\boldsymbol{X}_L, \boldsymbol{X}_R$ of the controller $\boldsymbol{K}$ which appear in the fixed order loop factorizations (2.23), (2.24) can be made independent of $\boldsymbol{U}_K, \boldsymbol{V}_K$ by defining

$$\begin{aligned} \boldsymbol{Q}_L = \boldsymbol{U}_P\boldsymbol{X}_L &= \left[\begin{array}{c|c} \boldsymbol{A}_{QL} & \boldsymbol{B}_{QL} \\ \hline \boldsymbol{C}_{QL} & \boldsymbol{D}_{QL} \end{array}\right] \text{ with} \\ \begin{bmatrix} \boldsymbol{A}_{QL} & \boldsymbol{B}_{QL} \\ \boldsymbol{C}_{QL} & \boldsymbol{D}_{QL} \end{bmatrix} &= \begin{bmatrix} \mathbf{1} & \boldsymbol{L}_K \\ \mathbf{0} & \boldsymbol{E}_{KP} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix} \end{aligned} \tag{2.40}$$

$$\begin{aligned} \boldsymbol{Q}_R = \boldsymbol{X}_R\boldsymbol{V}_P &= \left[\begin{array}{c|c} \boldsymbol{A}_{QR} & \boldsymbol{B}_{QR} \\ \hline \boldsymbol{C}_{QR} & \boldsymbol{D}_{QR} \end{array}\right] \text{ with} \\ \begin{bmatrix} \boldsymbol{A}_{QR} & \boldsymbol{B}_{QR} \\ \boldsymbol{C}_{QR} & \boldsymbol{D}_{QR} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{F}_K & \boldsymbol{E}_{PK} \end{bmatrix} \end{aligned} \tag{2.41}$$

Since they are based on the controller's factors $\boldsymbol{X}_L, \boldsymbol{X}_R$ there is a relation between these variables and the controller $\boldsymbol{K}$ in the form of

$$\boldsymbol{K} = \boldsymbol{Q}_R(\boldsymbol{Y}_R\boldsymbol{V}_P)^{-1} = (\boldsymbol{U}_P\boldsymbol{Y}_L)^{-1}\boldsymbol{Q}_L \tag{2.42}$$

Proposition The new variables $\boldsymbol{Q}_L$ and $\boldsymbol{Q}_R$ that parametrize the controller $\boldsymbol{K}$ as defined in (2.40) and (2.41) respectively, lead to the loop's state space matrices $\boldsymbol{A}_H, \boldsymbol{B}_H, \boldsymbol{C}_H, \boldsymbol{D}_H$ becoming linear in $\boldsymbol{A}_{Qi}, \boldsymbol{B}_{Qi}, \boldsymbol{C}_{Qi}, \boldsymbol{D}_{Qi}$ with $i \in \{L, R\}$. In addition to that the quadratic Riccati equations (2.21) and (2.22) turn into linear Sylvester equations for $\boldsymbol{R}_{KP}$ and $\boldsymbol{R}_{PK}$ respectively.

The following tables translate recurring terms appearing in the state space matrices of the loop (2.19) and the Riccati conditions (2.21), (2.22) in the new parametrizations.

Original Term	Term using $\boldsymbol{Q}_L$
Riccati (2.21)	$\boldsymbol{R}_{KP}\boldsymbol{A}_P = \boldsymbol{A}_{QL}\boldsymbol{R}_{KP} + \boldsymbol{B}_{QL}\boldsymbol{C}_P$
$\boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K$	$\boldsymbol{A}_{QL} + \boldsymbol{R}_{KP}\boldsymbol{B}_P\boldsymbol{C}_{QL}$
$\boldsymbol{B}_K\boldsymbol{E}_{PK}$	$\boldsymbol{B}_{QL} + \boldsymbol{R}_{KP}\boldsymbol{B}_P\boldsymbol{D}_{QL}$
$\boldsymbol{E}_{KP}\boldsymbol{C}_K$	$\boldsymbol{C}_{QL}$
$\boldsymbol{E}_{KP}\boldsymbol{D}_K = \boldsymbol{D}_K\boldsymbol{E}_{PK}$	$\boldsymbol{D}_{QL}$
$\boldsymbol{E}_{KP} = \boldsymbol{U}_P\boldsymbol{V}_K$	$\boldsymbol{1} + \boldsymbol{D}_{QL}\boldsymbol{D}_P$
$\boldsymbol{E}_{PK} = \boldsymbol{U}_K\boldsymbol{V}_P$	$\boldsymbol{1} + \boldsymbol{D}_P\boldsymbol{D}_{QL}$
$\boldsymbol{L}_K$	$\boldsymbol{B}_{QL}\boldsymbol{D}_P - \boldsymbol{R}_{KP}\boldsymbol{B}_P$
$\boldsymbol{F}_P$	$\boldsymbol{D}_{QL}\boldsymbol{C}_P + \boldsymbol{C}_{QL}\boldsymbol{R}_{KP}$
$\boldsymbol{A}_K + \boldsymbol{L}_K\boldsymbol{C}_K$	$\boldsymbol{A}_{QL}$
$\boldsymbol{B}_K + \boldsymbol{L}_K\boldsymbol{D}_K$	$\boldsymbol{B}_{QL}$

(2.43)

Original Term	Term using $\boldsymbol{Q}_R$
Riccati (2.22)	$\boldsymbol{A}_P\boldsymbol{R}_{PK} = \boldsymbol{R}_{PK}\boldsymbol{A}_{QR} - \boldsymbol{B}_P\boldsymbol{C}_{QR}$
$\boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K$	$\boldsymbol{A}_{QR} - \boldsymbol{B}_{QR}\boldsymbol{C}_P\boldsymbol{R}_{PK}$
$\boldsymbol{E}_{KP}\boldsymbol{C}_K$	$\boldsymbol{C}_{QR} - \boldsymbol{D}_{QR}\boldsymbol{C}_P\boldsymbol{R}_{PK}$
$\boldsymbol{B}_K\boldsymbol{E}_{PK}$	$\boldsymbol{B}_{QR}$
$\boldsymbol{D}_K\boldsymbol{E}_{PK} = \boldsymbol{E}_{KP}\boldsymbol{D}_K$	$\boldsymbol{D}_{QR}$
$\boldsymbol{E}_{PK} = \boldsymbol{U}_K\boldsymbol{V}_P$	$\boldsymbol{1} + \boldsymbol{D}_P\boldsymbol{D}_{QR}$
$\boldsymbol{E}_{KP} = \boldsymbol{U}_P\boldsymbol{V}_K$	$\boldsymbol{1} + \boldsymbol{D}_{QR}\boldsymbol{D}_P$
$\boldsymbol{F}_K$	$\boldsymbol{D}_P\boldsymbol{C}_{QR} + \boldsymbol{C}_P\boldsymbol{R}_{PK}$
$\boldsymbol{L}_P$	$\boldsymbol{B}_P\boldsymbol{D}_{QR} - \boldsymbol{R}_{PK}\boldsymbol{B}_{QR}$
$\boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{F}_K$	$\boldsymbol{A}_{QR}$
$\boldsymbol{C}_K + \boldsymbol{D}_K\boldsymbol{F}_K$	$\boldsymbol{C}_{QR}$

(2.44)

Proof This follows if the definitions (2.40) and (2.41) are applied on the loop's state space matrices (2.19) and the Riccati conditions (2.21), (2.22).

2.4.6 Controllability and Observability

Proposition The right and left parametrizations respectively preserve controllability and observability of the controller

$$\begin{aligned} \boldsymbol{K} \text{ controllable} &\Leftrightarrow \boldsymbol{Q}_R \text{ controllable} \\ \boldsymbol{K} \text{ observable} &\Leftrightarrow \boldsymbol{Q}_L \text{ observable} \end{aligned} \tag{2.45}$$

Proof This can be verified by application of (C.33) on $\boldsymbol{K}$, so that $\boldsymbol{K}$ is unobservable if and only if there exists an eigenvector $\boldsymbol{x}$ with $\boldsymbol{A}_K\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\boldsymbol{C}_K\boldsymbol{x} = \boldsymbol{0}$ which is equivalent to $\boldsymbol{A}_K\boldsymbol{x} + \boldsymbol{L}_K(\boldsymbol{C}_K\boldsymbol{x}) = \lambda\boldsymbol{x}$ and $\boldsymbol{E}_{KP}\boldsymbol{C}_K\boldsymbol{x} = \boldsymbol{0}$ since $\boldsymbol{E}_{KP}$ is invertible. This again is equivalent to $\boldsymbol{A}_{QL}\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\boldsymbol{C}_{QL}\boldsymbol{x} = \boldsymbol{0}$ which is true if and only if $\boldsymbol{Q}_L$ is unobservable and thus verifies the second line in (2.45). The proof for $\boldsymbol{Q}_R$ is similar.

2.4.7 Necessary Condition for Loop Ring Membership

Proposition Ring membership of the parameters is a necessary condition for loop ring membership

$$\boldsymbol{K} \in K(\boldsymbol{P}) \Rightarrow \boldsymbol{Q}_i \in R_Z^{n,m} \tag{2.46}$$

for $i \in \{L, R\}$.

Proof This is an immediate consequence of (2.7) which for $\boldsymbol{K} \in K(\boldsymbol{P})$ requires $\boldsymbol{X}_i \in R_Z^{n,m}$. Since $\boldsymbol{Q}_L = \boldsymbol{U}_P\boldsymbol{X}_L$ and $\boldsymbol{Q}_R = \boldsymbol{X}_R\boldsymbol{V}_P$ with $\boldsymbol{U}_P, \boldsymbol{V}_P$ being constant matrices. This can also be deduced from the triangular structure of the base transformed dynamic matrix $\boldsymbol{A}_H$ of the loop $\boldsymbol{H}$ in (2.33) and (2.34) where the dynamic matrices $\boldsymbol{A}_{QL} = \boldsymbol{A}_K + \boldsymbol{L}_K\boldsymbol{C}_K$ and $\boldsymbol{A}_{QR} = \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{F}_K$ of the parameters $\boldsymbol{Q}_L$ and $\boldsymbol{Q}_R$ are blocks on the diagonal.

Chapter 3

Stable Observable Pairs

This chapter introduces the notion of square lossless systems and provides a parametrization of stable observable pairs by solving a Nudelman interpolation problem. It has been shown in [2] that lossless systems and the stable observable pairs that can be extracted from them form a differentiable manifold. The controller $\boldsymbol{K}$ is in general not stable, so this parametrization can not be applied to $\boldsymbol{K}$ directly. However, it has been shown at the end of chapter 2 that the left parameter $\boldsymbol{Q}_L$ needs to be stable for the loop $\boldsymbol{H}$ to be stable and it needs to be observable for the controller $\boldsymbol{K}$ to be observable. The parametrization of the stable observable pair contained in $\boldsymbol{Q}_L$ then allows to vary the entire parameter $\boldsymbol{Q}_L$ infinitesimally. This enables differential calculus to be carried out with $\boldsymbol{Q}_L$ which will be important in the optimization stage in chapter 4.

3.1 Lossless Systems

This section develops representations of discrete square lossless systems from an energy conservation law. It further shows that any stable observable pair can be completed to a discrete square lossless system.

3.1.1 Transfer Function Representation

If the output energy of a system $\boldsymbol{L}$ equals its input energy then it is energy conserving and $\boldsymbol{L}$ is called a *lossless system*.

Proposition A transfer function $\boldsymbol{L}$ represents a discrete stable lossless system if and only if

$$\boldsymbol{L}^H(e^{j\mathbb{R}})\boldsymbol{L}(e^{j\mathbb{R}}) = \mathbf{1} \tag{3.1}$$

with its analytic continuation

$$\boldsymbol{L}^H(\overline{z}^{-1})\boldsymbol{L}(z) = \mathbf{1} \text{ almost everywhere} \tag{3.2}$$

Proof To verify this, the energy conserving property has to be exploited. The H2 norm of a signal provides a measure of the energy contained in the signal.

A linear time invariant system $\boldsymbol{y} = \boldsymbol{L}\boldsymbol{u}$ is lossless if the energy of the output signal equals the energy of the input signal

$$||\boldsymbol{y}||_{H2} = ||\boldsymbol{u}||_{H2}$$

According to (B.16) and (B.13) the H2 norm of a signal $\boldsymbol{y}$ analytic in $D = \{z \in \mathbb{C} : |z| \geq 1\}$ is given by

$$||\boldsymbol{y}(D)||_{H2} = \left|\left|\boldsymbol{y}(e^{j2\pi\mathbb{R}})\right|\right|_{L2} = \sqrt{\int_{\mathbb{R}} ||\boldsymbol{y}(e^{j2\pi t})||^2 \, \mathrm{d}t}$$

The simplest result will be obtained when using the matrix 2 norm (B.7) which can be given by

$$||\boldsymbol{x}||_2 = \sqrt{\boldsymbol{x}^H \boldsymbol{x}}$$

for any vector $\boldsymbol{x}$. The H2 norm of the signal $\boldsymbol{y}$ is then given by

$$||\boldsymbol{y}(D)||_{H2} = \sqrt{\int_{\mathbb{R}} \boldsymbol{y}^H(e^{j2\pi t})\boldsymbol{y}(e^{j2\pi t})\mathrm{d}t}$$

Applied to the energy conservation condition of a lossless system provides

$$\begin{aligned} ||\boldsymbol{y}(D)||_{H2} &= \sqrt{\int_{\mathbb{R}} \boldsymbol{y}^H(e^{j2\pi t})\boldsymbol{y}(e^{j2\pi t})\mathrm{d}t} = \\ &= \sqrt{\int_{\mathbb{R}} \boldsymbol{u}^H(e^{j2\pi t})\boldsymbol{L}^H(e^{j2\pi t})\boldsymbol{L}(e^{j2\pi t})\boldsymbol{u}(e^{j2\pi t})\mathrm{d}t} \stackrel{!}{=} \\ &\stackrel{!}{=} \sqrt{\int_{\mathbb{R}} \boldsymbol{u}^H(e^{j2\pi t})\boldsymbol{u}(e^{j2\pi t})\mathrm{d}t} = ||\boldsymbol{u}(D)||_{H2} \end{aligned}$$

The general solution for all inputs $\boldsymbol{u}$ is

$$\boldsymbol{L}^H(e^{j2\pi t})\boldsymbol{L}(e^{j2\pi t}) = \boldsymbol{1};\ \forall t \in \mathbb{R}$$

so that (3.1) follows. In order to show the analytic continuation, define

$$\boldsymbol{F}(z) = \boldsymbol{L}^H(\overline{z}^{-1})\boldsymbol{L}(z) - \boldsymbol{1}$$

which is analytic except at its poles and by (3.1) satisfies $\boldsymbol{F}(e^{j\mathbb{R}}) = \boldsymbol{0}$ on the unit circle. Since $\boldsymbol{L}(z)$ is stable, it has all its poles strictly inside the unit circle and consequently $\boldsymbol{L}^H(\overline{z}^{-1})$ has all its poles strictly outside the unit circle so that $\boldsymbol{F}(z)$ is analytic on a small open neighborhood of the unit circle. Then (B.4) applies and states that if $\boldsymbol{F}(z) = \boldsymbol{0}$ is true on the open neighborhood of the unit circle, then it is also true almost everywhere, that is everywhere except at its poles. This then verifies (3.2).

3.1.2 State Space Representation

A discrete lossless system $\boldsymbol{L}(s)$ is called *square* if its transfer matrix has an equal number of columns (inputs) and lines (outputs).

Proposition A minimal realization $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ represents a discrete stable minimal lossless system if and only if

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \text{ with } \boldsymbol{G} > 0 \tag{3.3}$$

where $\boldsymbol{G}, \boldsymbol{G}^{-1}$ are the system's Gram observability and controllability matrices respectively.

Proof This can be verified by applying the lossless condition (3.1) on a transfer function $\boldsymbol{L}(z)$ given in terms of the state space matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ as in (C.7):

$$\boldsymbol{1} = \boldsymbol{L}^H(z)\boldsymbol{L}(z) = \big(\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}\big)^H \big(\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}\big)$$

for $|z| = 1 \Leftrightarrow \overline{z} = z^{-1}$. This develops into

$$\begin{aligned} \boldsymbol{1} = \boldsymbol{L}^H(z)\boldsymbol{L}(z) = \boldsymbol{B}^H(z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{C}^H\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \\ + \boldsymbol{B}^H(z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{C}^H\boldsymbol{D} + \boldsymbol{D}^H\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D} \end{aligned} \tag{3.4}$$

Since $\boldsymbol{L}(z)$ is required to be stable, the inverse $(z\boldsymbol{1} - \boldsymbol{A})^{-1} = z^{-1}(\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1}$ and thus also $(\overline{z}\boldsymbol{1} - \boldsymbol{A}^H)^{-1} = (z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1} = z(\boldsymbol{1} - z\boldsymbol{A}^H)^{-1}$ exist for $|z| = 1 \Leftrightarrow \overline{z} = z^{-1}$. And since the realization of $\boldsymbol{L}(z)$ is furthermore required to be minimal and thus observable, according to (C.34) there is a unique Gram observability matrix $\boldsymbol{G} > 0$ that solves

$$\boldsymbol{G} - \boldsymbol{A}^H\boldsymbol{G}\boldsymbol{A} = \boldsymbol{C}^H\boldsymbol{C} \tag{3.5}$$

In order to simplify the lossless condition (3.4), the expansion of the following product is helpful:

$$\begin{aligned} (z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)\boldsymbol{G}(z\boldsymbol{1} - \boldsymbol{A}) = (\boldsymbol{1} - z\boldsymbol{A}^H)\boldsymbol{G}(\boldsymbol{1} - z^{-1}\boldsymbol{A}) = \\ = \boldsymbol{G}(\boldsymbol{1} - z^{-1}\boldsymbol{A}) + (\boldsymbol{1} - z\boldsymbol{A}^H)\boldsymbol{G} + \boldsymbol{A}^H\boldsymbol{G}\boldsymbol{A} - \boldsymbol{G} = \\ = \boldsymbol{G}(\boldsymbol{1} - z^{-1}\boldsymbol{A}) + (\boldsymbol{1} - z\boldsymbol{A}^H)\boldsymbol{G} - \boldsymbol{C}^H\boldsymbol{C} \end{aligned}$$

using (3.5). Left multiplication by $(\boldsymbol{1} - z\boldsymbol{A}^H)^{-1}$ and right multiplication by $(\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1}$ leads to

$$\begin{aligned} (\boldsymbol{1} - z\boldsymbol{A}^H)^{-1}\boldsymbol{C}^H\boldsymbol{C}(\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1} = (z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{C}^H\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1} = \\ = \boldsymbol{G}(\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1} + (\boldsymbol{1} - z\boldsymbol{A}^H)^{-1}\boldsymbol{G} - \boldsymbol{G} \end{aligned}$$

The matrix relation (A.56) can be used to obtain

$$\begin{aligned} (\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1} &= \boldsymbol{1} + z^{-1}\boldsymbol{A}(\boldsymbol{1} - z^{-1}\boldsymbol{A})^{-1} = \boldsymbol{1} + \boldsymbol{A}(z\boldsymbol{1} - \boldsymbol{A})^{-1} \\ (\boldsymbol{1} - z\boldsymbol{A}^H)^{-1} &= \boldsymbol{1} + (\boldsymbol{1} - z\boldsymbol{A}^H)^{-1}z\boldsymbol{A}^H = \boldsymbol{1} + (z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{A}^H \end{aligned}$$

and thus

$$(z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{C}^H\boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1} = \boldsymbol{G}\boldsymbol{A}(z\boldsymbol{1} - \boldsymbol{A})^{-1} + (z^{-1}\boldsymbol{1} - \boldsymbol{A}^H)^{-1}\boldsymbol{A}^H\boldsymbol{G} + \boldsymbol{G}$$

Substituting this in (3.4) gives

$$\boldsymbol{M}^H(z\mathbf{1}-\boldsymbol{A})^{-1}\boldsymbol{B}+\boldsymbol{B}^H(z^{-1}\mathbf{1}-\boldsymbol{A}^H)^{-1}\boldsymbol{M}=\boldsymbol{N}$$

with $\boldsymbol{M}=\boldsymbol{A}^H\boldsymbol{G}\boldsymbol{B}+\boldsymbol{C}^H\boldsymbol{D}$ and $\boldsymbol{N}=\mathbf{1}-\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B}-\boldsymbol{D}^H\boldsymbol{D}$. According to (A.67), the inverses can now be developed in geometric series:

$$(z\mathbf{1}-\boldsymbol{A})^{-1}=z^{-1}(\mathbf{1}-z^{-1}\boldsymbol{A})^{-1}=z^{-1}\sum_{k=0}^{\infty}z^{-k}\boldsymbol{A}^k=\sum_{k=-\infty}^{-1}z^k\boldsymbol{A}^{-k-1}$$

$$(z^{-1}\mathbf{1}-\boldsymbol{A}^H)^{-1}=z(\mathbf{1}-z\boldsymbol{A}^H)^{-1}=z\sum_{k=0}^{\infty}z^k{\boldsymbol{A}^H}^k=\sum_{k=1}^{\infty}z^k\boldsymbol{A}^{H^{-k-1}}$$

Now the lossless condition writes as

$$\boldsymbol{M}^H\left(\sum_{k=-\infty}^{-1}z^k\boldsymbol{A}^{-k-1}\right)\boldsymbol{B}+\boldsymbol{B}^H\left(\sum_{k=1}^{\infty}z^k\boldsymbol{A}^{H^{k-1}}\right)\boldsymbol{M}=\boldsymbol{N}$$

which is equal to

$$\sum_{k=-\infty}^{\infty}z^k\boldsymbol{F}_k=\boldsymbol{N}\text{ with }\boldsymbol{F}_k=\begin{cases}\boldsymbol{M}^H\boldsymbol{A}^{k-1}\boldsymbol{B}; & k\le -1\\ \mathbf{0}; & k=0\\ \boldsymbol{B}^H{\boldsymbol{A}^H}^{k-1}\boldsymbol{M}; & k\ge 1\end{cases}$$

Comparing the coefficients provides

$$\begin{aligned}&\boldsymbol{M}^H\begin{bmatrix}\boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \boldsymbol{A}^2\boldsymbol{B}\cdots\end{bmatrix}=\mathbf{0}\text{ for }k\neq 0\\ &\boldsymbol{N}=\mathbf{0}\text{ for }k=0\end{aligned}$$

But since the realization of $\boldsymbol{L}(z)$ is required to be minimal and thus controllable, according to (C.31) the matrix $\begin{bmatrix}\boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \boldsymbol{A}^2\boldsymbol{B}\cdots\end{bmatrix}$ has full line rank and thus a discrete square stable minimal lossless system is characterized by

$$\begin{aligned}\boldsymbol{M}&=\boldsymbol{A}^H\boldsymbol{G}\boldsymbol{B}+\boldsymbol{C}^H\boldsymbol{D}=\mathbf{0}\\ \boldsymbol{N}&=\mathbf{1}-\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B}-\boldsymbol{D}^H\boldsymbol{D}=\mathbf{0}\end{aligned}$$

in addition to (3.5). Assembling these three equations leads to the block matrix form of (3.3).

3.1.3 Discrete Lossless Completion

Proposition A stable observable pair $(\boldsymbol{A},\boldsymbol{C})$ with Gram observability matrix $\boldsymbol{G}$ can be completed by a pair $(\boldsymbol{B},\boldsymbol{D})$ so that the realization $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D})$ represents a discrete stable minimal lossless system

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}=\begin{bmatrix}\boldsymbol{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\\ &\text{parametrized by }\begin{bmatrix}\boldsymbol{B}\\ \boldsymbol{D}\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}_0\\ \boldsymbol{D}_0\end{bmatrix}\boldsymbol{U}\end{aligned}\tag{3.6}$$

with $\boldsymbol{U}$ unitary.

Proof This is due to the fact that observability of the pair $(\boldsymbol{A} \in \mathbb{C}^{n,n}, \boldsymbol{C} \in \mathbb{C}^{m,n})$ implies the existence of a Gram observability matrix $\boldsymbol{G} > 0$ so that

$$\boldsymbol{A}^H \boldsymbol{G} \boldsymbol{A} + \boldsymbol{C}^H \boldsymbol{C} = \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} = \boldsymbol{G} \geq 0$$

according to (C.34) and (A.77). This shows that

$$\operatorname{rnk} \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} = n$$

so that according to (A.80) there exist orthogonal completions

$$\begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}_\perp = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix} \boldsymbol{T}$$

parametrized by invertible $\boldsymbol{T}$ so that

$$\begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \underbrace{\left(\begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \right)^{-1}}_{\boldsymbol{G}^{-1}} \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}^H + \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}^H = \mathbf{1} \tag{3.7}$$

The restriction

$$\boldsymbol{T} = \left(\begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix}^H \begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix} \right)^{-\frac{1}{2}} \boldsymbol{U}$$

with $\boldsymbol{U}$ unitary, reduces the set of orthogonal completions to

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_0 \\ \boldsymbol{Y}_0 \end{bmatrix} \boldsymbol{U} \tag{3.8}$$

with

$$\begin{bmatrix} \boldsymbol{X}_0 \\ \boldsymbol{Y}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix}^H \begin{bmatrix} \boldsymbol{X}'_0 \\ \boldsymbol{Y}'_0 \end{bmatrix} \right)^{-\frac{1}{2}}$$

so that the constraint

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \mathbf{1}$$

is fulfilled. This together with (3.7) gives

$$\begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{A}^H & \boldsymbol{A}\boldsymbol{G}^{-1}\boldsymbol{C}^H \\ \boldsymbol{C}\boldsymbol{G}^{-1}\boldsymbol{A}^H & \boldsymbol{C}\boldsymbol{G}^{-1}\boldsymbol{C}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^H & \boldsymbol{X}\boldsymbol{Y}^H \\ \boldsymbol{Y}\boldsymbol{X}^H & \boldsymbol{Y}\boldsymbol{Y}^H \end{bmatrix} = \mathbf{1}$$

which can be written as

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X} \\ \boldsymbol{C} & \boldsymbol{Y} \end{bmatrix} \begin{bmatrix} \boldsymbol{G}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X} \\ \boldsymbol{C} & \boldsymbol{Y} \end{bmatrix}^H = \begin{bmatrix} \boldsymbol{G}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{3.9}$$

With the definition of $\boldsymbol{B} = \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X}, \boldsymbol{D} = \boldsymbol{Y}$ and $\boldsymbol{B}_0 = \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X}_0, \boldsymbol{D}_0 = \boldsymbol{Y}_0$ the orthogonal completions (3.8) are given by

$$\begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}} \boldsymbol{X}_0 \\ \boldsymbol{Y}_0 \end{bmatrix} \boldsymbol{U} = \begin{bmatrix} \boldsymbol{B}_0 \\ \boldsymbol{D}_0 \end{bmatrix} \boldsymbol{U}$$

and equation (3.9) can be written in the form of (3.3) so that (3.6) follows.

3.1.4 Link to Conjugate Realization

Proposition If an observable pair $(\boldsymbol{A}, \boldsymbol{C})$ has a matrix $\boldsymbol{I}$ that solves

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{I} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}} \tag{3.10}$$

then $\boldsymbol{I}$ is unique and invertible.

Proof Assume there exists a matrix $\boldsymbol{I}$ that solves (3.10). Since the pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable the observability matrix

$$\boldsymbol{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \end{bmatrix}$$

has full column rank and provides

$$\boldsymbol{O}\boldsymbol{I} = \overline{\boldsymbol{O}}$$

using (3.10). Suppose $\boldsymbol{I}$ not uniquely determined by (3.10) then there should exist another $\boldsymbol{I}' \neq \boldsymbol{I}$ with $\boldsymbol{O}\boldsymbol{I}' = \overline{\boldsymbol{O}}$ hence $\boldsymbol{O}(\boldsymbol{I} - \boldsymbol{I}') = \boldsymbol{0}$ which contradicts $\boldsymbol{O}$ having full column rank so that $\boldsymbol{I}$ must be unique. Suppose $\boldsymbol{I}$ not invertible then there should exist a vector $\boldsymbol{z} \neq \boldsymbol{0}$ with $\boldsymbol{I}\boldsymbol{z} = \boldsymbol{0}$ hence $\boldsymbol{O}\boldsymbol{I}\boldsymbol{z} = \boldsymbol{0}$ which contradicts $\boldsymbol{O}$ having full column rank so that $\boldsymbol{I}$ must be invertible.

Proposition If the observable pair $(\boldsymbol{A}, \boldsymbol{C})$ of a discrete stable minimal lossless system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ has a matrix $\boldsymbol{I}$ that solves (3.10) then it also solves

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \tag{3.11}$$

with unitary $\boldsymbol{U}$ given by

$$\boldsymbol{U} = \overline{\boldsymbol{B}}^H \boldsymbol{I}^H \boldsymbol{G}\boldsymbol{B} + \overline{\boldsymbol{D}}^H \boldsymbol{D} \tag{3.12}$$

and the Gram observability matrix $\overline{\boldsymbol{G}}$ of the conjugate pair $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{C}})$ satisfying

$$\overline{\boldsymbol{G}} = \boldsymbol{I}^H \boldsymbol{G}\boldsymbol{I} \tag{3.13}$$

Hence the lossless system is real in the sense of (C.27) if and only if $\boldsymbol{U} = \boldsymbol{1}$ due to (C.28).

Proof Assume there exists a matrix $\boldsymbol{I}$ that solves (3.10) for the observable pair $(\boldsymbol{A}, \boldsymbol{C})$ of the discrete stable minimal lossless system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$. The Gram observability matrix $\boldsymbol{G} > 0$ is uniquely defined by

$$\boldsymbol{A}^H \boldsymbol{G}\boldsymbol{A} - \boldsymbol{G} + \boldsymbol{C}^H \boldsymbol{C} = \boldsymbol{0}$$

according to (C.34). With (3.10) the conjugate of this equation provides

$$\boldsymbol{A}^H \boldsymbol{I}^{-H} \overline{\boldsymbol{G}} \boldsymbol{I}^{-1} \boldsymbol{A} - \boldsymbol{I}^{-H} \boldsymbol{G} \boldsymbol{I}^{-1} + \boldsymbol{C}^H \boldsymbol{C} = \boldsymbol{0}$$

Subtracting this from (3.10) gives

$$\boldsymbol{A}^H(\boldsymbol{G}-\boldsymbol{I}^{-H}\overline{\boldsymbol{G}}\boldsymbol{I}^{-1})\boldsymbol{A}-(\boldsymbol{G}-\boldsymbol{I}^{-H}\boldsymbol{G}\boldsymbol{I}^{-1})=\boldsymbol{0}$$

This is a Stein equation and according to (A.88) has a unique solution (3.13) since $\boldsymbol{A}$ is stable. With (3.10) and (3.13) the conjugate of the lossless condition (3.3) provides

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}\overline{\boldsymbol{B}}\\ \boldsymbol{C} & \overline{\boldsymbol{D}}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}\overline{\boldsymbol{B}}\\ \boldsymbol{C} & \overline{\boldsymbol{D}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}$$

Thus $(\boldsymbol{B},\boldsymbol{D})$ and $(\boldsymbol{I}\overline{\boldsymbol{B}},\overline{\boldsymbol{D}})$ are two lossless completions to the pair $(\boldsymbol{A},\boldsymbol{C})$. Then according to (3.6) there exists a unitary $\boldsymbol{U}$ so that

$$\begin{bmatrix}\boldsymbol{B}\\ \boldsymbol{D}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{B}\\ \boldsymbol{D}\end{bmatrix}}\boldsymbol{U}$$

Together with (3.10) this can be written in the form (3.11). Due to (3.3) we can rewrite (3.11) in the form of

$$\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{U}^H\end{bmatrix}=\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}}$$

so that $\boldsymbol{U}$ is uniquely defined and (3.12) follows.

3.2 Nudelman Interpolation Problem

This section introduces the Nudelman interpolation problem and provides a solution to it in the form of a constraint on general discrete square lossless systems. Extensive work on this subject has been done in [11] wherein it is also shown that the Nudelman interpolation problem is a matrix generalization of the Nevanlinna-Pick interpolation problem. A parametrization of the solutions of the Nudelman interpolation problem then leads to a parametrization of stable observable pairs.

3.2.1 Interpolation Problem Definition

The *Nudelman interpolation problem* consists of finding a discrete lossless system $\boldsymbol{L}(z)$ which for a given stable and observable pair $(\boldsymbol{A}_0,\boldsymbol{C}_0)$ and interpolation value $\boldsymbol{W}$ solves

$$\boldsymbol{W}=\frac{1}{2\pi j}\oint_C \boldsymbol{L}^H(\overline{z}^{-1})\boldsymbol{C}_0(z\boldsymbol{1}-\boldsymbol{A}_0)^{-1}\mathrm{d}z \tag{3.14}$$

with $C=\{z:|z|=1\}$ being the unit circle and $\delta(\boldsymbol{L})=\operatorname{col}\boldsymbol{A}_0$, which means the McMillan degree of $\boldsymbol{L}$ equals the state space dimension of $\boldsymbol{A}_0$. Note that $\boldsymbol{L}(\overline{z}^{-1})=\boldsymbol{L}^{-1}(z)$ due to (3.2).

Proposition The Nudelman interpolation problem (3.14) has an explicit solution $\boldsymbol{L}(z)$ if the existence condition

$$\exists\boldsymbol{G}>0:\ \boldsymbol{A}_0^H\boldsymbol{G}\boldsymbol{A}_0-\boldsymbol{G}+\boldsymbol{C}_0^H\boldsymbol{C}_0=\boldsymbol{W}^H\boldsymbol{W} \tag{3.15}$$

is satisfied. The solution $\boldsymbol{L}(z)$ then has a stable minimal state space representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$, with $\boldsymbol{A}_0, \boldsymbol{A}$ having same dimension, that is constrained by

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{W} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_0 \\ \boldsymbol{C}_0 \end{bmatrix} \tag{3.16}$$

with $\boldsymbol{L}(z) = \boldsymbol{C}(z\boldsymbol{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}$ being a square stable minimal lossless system with Gram observability matrix $\boldsymbol{G}$ and

$$0 < \boldsymbol{G} \leq \boldsymbol{G}_0 \tag{3.17}$$

wherein $\boldsymbol{G}_0$ is the Gram observability matrix of the pair $(\boldsymbol{A}_0, \boldsymbol{C}_0)$.

Proof In order to verify (3.16), let $(\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{C}', \boldsymbol{D}')$ be another minimal state space representation of $\boldsymbol{L}$ with the dynamic matrices $\boldsymbol{A}_0, \boldsymbol{A}'$ having same dimension, then with $\boldsymbol{L}^H(\overline{z}^{-1}) = \boldsymbol{B}'^H(z^{-1}\boldsymbol{1} - \boldsymbol{A}'^H)^{-1}\boldsymbol{C}'^H + \boldsymbol{D}'^H$, the interpolation problem (3.14) writes as

$$\boldsymbol{W} = \boldsymbol{B}'^H \frac{1}{2\pi j} \oint_C (z^{-1}\boldsymbol{1} - \boldsymbol{A}'^H)^{-1}\boldsymbol{C}'^H\boldsymbol{C}_0(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\mathrm{d}z + \\ + \boldsymbol{D}'^H\boldsymbol{C}_0 \frac{1}{2\pi j} \oint_C (z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\mathrm{d}z$$

For the first integral, using (A.56), we get

$$(z^{-1}\boldsymbol{1} - \boldsymbol{A}'^H)^{-1} = z(\boldsymbol{1} - z\boldsymbol{A}'^H)^{-1}$$
$$(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1} = z^{-1}(\boldsymbol{1} - z^{-1}\boldsymbol{A}_0)^{-1} = z^{-1}((z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\boldsymbol{A}_0 + \boldsymbol{1})$$

For the second integral, using (A.67) on $(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}$ we get

$$(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1} = z^{-1}(\boldsymbol{1} - z^{-1}\boldsymbol{A}_0)^{-1} = z^{-1}\sum_{k=0}^{\infty} z^{-k}\boldsymbol{A}_0^k = \sum_{k=0}^{\infty} z^{-(k+1)}\boldsymbol{A}_0^k$$

which converges on C since $\boldsymbol{A}_0$ is stable. Using (B.3) the residue of this function is $\boldsymbol{1}$ and the second integral can be evaluated as

$$\boldsymbol{D}'^H\boldsymbol{C}_0 \frac{1}{2\pi j} \oint_C (z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\mathrm{d}z = \boldsymbol{D}'^H\boldsymbol{C}_0$$

so that the interpolation problem can now be written as

$$\boldsymbol{W} = \boldsymbol{B}'^H \frac{1}{2\pi j} \oint_C (\boldsymbol{1} - z\boldsymbol{A}'^H)^{-1}\boldsymbol{C}'^H\boldsymbol{C}_0(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\mathrm{d}z\boldsymbol{A}_0 + \\ + \boldsymbol{B}'^H \frac{1}{2\pi j} \oint_C (\boldsymbol{1} - z\boldsymbol{A}'^H)^{-1}\mathrm{d}z\boldsymbol{C}'^H\boldsymbol{C}_0 + \boldsymbol{D}'^H\boldsymbol{C}_0$$

For the first integral, define

$$\boldsymbol{T} = \frac{1}{2\pi j} \oint_C (\boldsymbol{1} - z\boldsymbol{A}'^H)^{-1}\boldsymbol{C}'^H\boldsymbol{C}_0(z\boldsymbol{1} - \boldsymbol{A}_0)^{-1}\mathrm{d}z \tag{3.18}$$

which is a square matrix since $\boldsymbol{A}_0, \boldsymbol{A}'$ are required to have same dimension. For the second integral, using (A.67) on $(\mathbf{1} - z\boldsymbol{A}'^H)^{-1}$ we get

$$(\mathbf{1} - z\boldsymbol{A}'^H)^{-1} = \sum_{k=0}^{\infty} z^k \boldsymbol{A}'^{H\,k}$$

which converges on C since $\boldsymbol{A}'$ is stable. Using (B.3) the residue of this function is $\mathbf{0}$ and the second integral can be evaluated as

$$\boldsymbol{B}'^H \frac{1}{2\pi j} \oint_C (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} \mathrm{d}z \boldsymbol{C}'^H \boldsymbol{C}_0 = \mathbf{0}$$

so that the interpolation problem can now be written as

$$\boldsymbol{W} = \boldsymbol{B}'^H \boldsymbol{T} \boldsymbol{A}_0 + \boldsymbol{D}'^H \boldsymbol{C}_0 \tag{3.19}$$

A matrix equation for $\boldsymbol{T}$ as defined in (3.18) can be constructed by calculating

$$\boldsymbol{A}'^H \boldsymbol{T} \boldsymbol{A}_0 = \frac{1}{2\pi j} \oint_C \boldsymbol{A}'^H (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} \boldsymbol{C}'^H \boldsymbol{C}_0 (z\mathbf{1} - \boldsymbol{A}_0)^{-1} \boldsymbol{A}_0 \mathrm{d}z$$

Again using (A.56), we get

$$\begin{aligned}
&\boldsymbol{A}'^H (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} = z^{-1} z \boldsymbol{A}'^H (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} = z^{-1}((\mathbf{1} - z\boldsymbol{A}'^H)^{-1} - \mathbf{1}) \\
&(z\mathbf{1} - \boldsymbol{A}_0)^{-1} \boldsymbol{A}_0 = (\mathbf{1} - z^{-1}\boldsymbol{A}_0)^{-1} z^{-1} \boldsymbol{A}_0 = (\mathbf{1} - z^{-1}\boldsymbol{A}_0)^{-1} - \mathbf{1}
\end{aligned}$$

so that

$$\begin{aligned}
\boldsymbol{A}'^H \boldsymbol{T} \boldsymbol{A}_0 = \boldsymbol{T} &- \frac{1}{2\pi j} \oint_C z^{-1} (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} \mathrm{d}z \boldsymbol{C}'^H \boldsymbol{C}_0 - \\
&- \boldsymbol{C}'^H \boldsymbol{C}_0 \frac{1}{2\pi j} \oint_C (z\mathbf{1} - \boldsymbol{A}_0)^{-1} \mathrm{d}z + \frac{1}{2\pi j} \oint_C z^{-1} \mathrm{d}z \boldsymbol{C}'^H \boldsymbol{C}_0
\end{aligned}$$

For the first integral, using (A.67) on $z^{-1}(\mathbf{1} - z\boldsymbol{A}'^H)^{-1}$ we get

$$z^{-1}(\mathbf{1} - z\boldsymbol{A}'^H)^{-1} = z^{-1} \sum_{k=0}^{\infty} z^k \boldsymbol{A}'^{H\,k} = \sum_{k=0}^{\infty} z^{k-1} \boldsymbol{A}'^{H\,k}$$

which converges on C since $\boldsymbol{A}'$ is stable. Using (B.3) the residue of this function is $\mathbf{1}$ and the first integral can be evaluated as

$$\frac{1}{2\pi j} \oint_C z^{-1} (\mathbf{1} - z\boldsymbol{A}'^H)^{-1} \mathrm{d}z \boldsymbol{C}'^H \boldsymbol{C}_0 = \boldsymbol{C}'^H \boldsymbol{C}_0$$

For the second integral, using (A.67) on $(z\mathbf{1} - \boldsymbol{A}_0)^{-1}$ we get

$$(z\mathbf{1} - \boldsymbol{A}_0)^{-1} = z^{-1}(\mathbf{1} - z^{-1}\boldsymbol{A}_0)^{-1} = z^{-1} \sum_{k=0}^{\infty} z^{-k} \boldsymbol{A}_0^k = \sum_{k=0}^{\infty} z^{-(k+1)} \boldsymbol{A}_0^k$$

which converges on C since $\boldsymbol{A}_0$ is stable. Using (B.3) the residue of this function is $\mathbf{1}$ and the second integral can be evaluated as

$$\boldsymbol{C}'^H \boldsymbol{C}_0 \frac{1}{2\pi j} \oint_C (z\mathbf{1} - \boldsymbol{A}_0)^{-1} \mathrm{d}z = \boldsymbol{C}'^H \boldsymbol{C}_0$$

For the third integral, using (B.3) the residue of z^{-1} is 1 and the third integral can be evaluated as

$$\frac{1}{2\pi j}\oint_C z^{-1}\mathrm{d}z\boldsymbol{C}'^H\boldsymbol{C}_0 = \boldsymbol{C}'^H\boldsymbol{C}_0$$

Assembling the expressions leads to a Stein equation for $\boldsymbol{T}$

$$\boldsymbol{A}'^H\boldsymbol{T}\boldsymbol{A}_0 - \boldsymbol{T} + \boldsymbol{C}'^H\boldsymbol{C}_0 = \boldsymbol{0} \tag{3.20}$$

which according to (A.88) uniquely determines $\boldsymbol{T}$ since $\boldsymbol{A}_0, \boldsymbol{A}'$ are stable and thus have eigenvalues $|\lambda_A| < 1, |\lambda_{A'}| < 1$. Together, matrix equations (3.20) and (3.19) can be written in block matrix form

$$\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}^H\begin{bmatrix}\boldsymbol{T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{W}\end{bmatrix} \tag{3.21}$$

Since $(\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{C}', \boldsymbol{D}')$ is a minimal state space representation of square lossless $\boldsymbol{L}$, the adjoint Gram observability matrix $\boldsymbol{G}' > 0$ exists and (3.3) applies, so that (3.21) provides

$$\begin{aligned}\boldsymbol{T}^H\boldsymbol{G}'^{-1}\boldsymbol{T} + \boldsymbol{W}^H\boldsymbol{W} &= \begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{W}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G}' & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{W}\end{bmatrix} =\\ &= \begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix}^H\begin{bmatrix}\boldsymbol{T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}\begin{bmatrix}\boldsymbol{G}' & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}^H\begin{bmatrix}\boldsymbol{T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix} =\\ &= \begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix}^H\begin{bmatrix}\boldsymbol{T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G}' & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{T} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix} =\\ &= \boldsymbol{A}_0^H\boldsymbol{T}^H\boldsymbol{G}'^{-1}\boldsymbol{T}\boldsymbol{A}_0 + \boldsymbol{C}_0^H\boldsymbol{C}_0\end{aligned}$$

Comparing this to the given existence condition (3.15) provides

$$\boldsymbol{T}^H\boldsymbol{G}'^{-1}\boldsymbol{T} = \boldsymbol{G} > 0 \tag{3.22}$$

This means, if the existence condition (3.15) is satisfied, the square matrix $\boldsymbol{T}$ has full rank which implies $\boldsymbol{T}$ to be invertible. Hence (3.21) can be written as

$$\left(\begin{bmatrix}\boldsymbol{T}^H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}\begin{bmatrix}\boldsymbol{T}^H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}\right)^H\begin{bmatrix}\boldsymbol{A}_0\\ \boldsymbol{C}_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{W}\end{bmatrix}$$

which makes it clear that $\boldsymbol{T}$ is a state space basis transformation matrix as in (C.36) that naturally leads to a new state space representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ with

$$\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{T}^H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}\begin{bmatrix}\boldsymbol{T}^H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1}$$

so that (3.21) leads to (3.16). The old state space representation $(\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{C}', \boldsymbol{D}')$ being stable minimal lossless with Gram observability matrix $\boldsymbol{G}'$, (3.3) provides

$$\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix}^H\begin{bmatrix}\boldsymbol{G}' & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{B}'\\ \boldsymbol{C}' & \boldsymbol{D}'\end{bmatrix} = \begin{bmatrix}\boldsymbol{G}' & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}$$

Transformed to the new state space representation and using (3.22) gives

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}$$

from which can be seen that $\boldsymbol{G}$ is the Gram observability matrix of the state space representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ of the discrete lossless system $\boldsymbol{L}$.

Condition (3.17) can be readily verified by subtracting the solution existence condition (3.15) from the Stein equation for the Gram observability matrix $\boldsymbol{G}_0$ of the pair $(\boldsymbol{A}_0, \boldsymbol{C}_0)$ in (C.34) which gives

$$\boldsymbol{A}_0^H(\boldsymbol{G}_0 - \boldsymbol{G})\boldsymbol{A}_0 - (\boldsymbol{G}_0 - \boldsymbol{G}) + \boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{0} \tag{3.23}$$

According to (A.89), this new symmetric Stein equation has a unique solution $\boldsymbol{G}_0 - \boldsymbol{G} \geq 0 \Leftrightarrow \boldsymbol{G} \leq \boldsymbol{G}_0$ since $\boldsymbol{W}^H\boldsymbol{W} \geq 0$ and $\boldsymbol{A}_0$ is stable and thus has eigenvalues $|\lambda_A| < 1$.

3.2.2 Interpolation Problem Solution

Proposition If there exists a matrix $\boldsymbol{G} > 0$ that verifies (3.15), the interpolation problem (3.14) has a lossless solution $\boldsymbol{L}$ with state space representation $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ constrained by (3.16). Then the general solution is given by

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\boldsymbol{1} \end{bmatrix} \tag{3.24}$$

with

$$\begin{aligned} \boldsymbol{S} &= (\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 + \boldsymbol{D}_0^H\boldsymbol{D}_0)^{-\frac{1}{2}}(\boldsymbol{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} \\ \boldsymbol{T} &= (\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} \end{aligned} \tag{3.25}$$

and $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ being any lossless completion to the observable pair $(\boldsymbol{A}_0, \boldsymbol{C}_0)$ with Gram observability matrix $\boldsymbol{G}_0$, as in (3.6).

Proof To verify this, first note that according to (3.6), an observable pair $(\boldsymbol{A}_0, \boldsymbol{C}_0)$ with Gram observability matrix $\boldsymbol{G}_0 > 0$ can always be completed by $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ so that

$$\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \tag{3.26}$$

Now the constraint (3.16) can be written in the form

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{X} \\ \boldsymbol{W} & \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \tag{3.27}$$

so that the state space matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ of the interpolation problem solution $\boldsymbol{L}$ are a function of yet unknown matrices $\boldsymbol{X}, \boldsymbol{Y}$ that have to be determined in the following.

Since $\boldsymbol{L}$ must be lossless and thus satisfy (3.3), this condition can be used on (3.27) to obtain a constraint for $\boldsymbol{X}, \boldsymbol{Y}$:

$$\begin{aligned}\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}^{-H} = \\ = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{X}\\ \boldsymbol{W} & \boldsymbol{Y}\end{bmatrix} \boldsymbol{H}^{-1} \begin{bmatrix}\boldsymbol{1} & \boldsymbol{X}\\ \boldsymbol{W} & \boldsymbol{Y}\end{bmatrix}^{H}\end{aligned} \tag{3.28}$$

with

$$\boldsymbol{H} = \begin{bmatrix}\boldsymbol{A}_0 & \boldsymbol{B}_0\\ \boldsymbol{C}_0 & \boldsymbol{D}_0\end{bmatrix}^{H} \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{A}_0 & \boldsymbol{B}_0\\ \boldsymbol{C}_0 & \boldsymbol{D}_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{H}_1 & \boldsymbol{H}_2\\ \boldsymbol{H}_2^H & \boldsymbol{H}_3\end{bmatrix} \tag{3.29}$$

The matrix $\boldsymbol{H}_1$ is invertible which can be seen by

$$\boldsymbol{H}_1 = \boldsymbol{A}_0^H \boldsymbol{G} \boldsymbol{A}_0 + \boldsymbol{C}_0^H \boldsymbol{C}_0 = \boldsymbol{G} + \boldsymbol{W}^H \boldsymbol{W} > 0 \tag{3.30}$$

using (3.15). Thus, according to (A.57), $\boldsymbol{H}$ can be factorized into

$$\boldsymbol{H} = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{0}\\ \boldsymbol{E}^H & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{H}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & \boldsymbol{E}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \tag{3.31}$$

with $\boldsymbol{E} = \boldsymbol{H}_1^{-1}\boldsymbol{H}_2$ and $\boldsymbol{F} = \boldsymbol{H}_3 - \boldsymbol{H}_2^H \boldsymbol{H}_1^{-1} \boldsymbol{H}_2 = \boldsymbol{F}^H$, which is invertible since $\boldsymbol{H}$ is invertible, so that (3.28) can now be written as

$$\begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{X}\\ \boldsymbol{W} & \boldsymbol{Y}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{E}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{H}_1^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}^{-1}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & \boldsymbol{0}\\ -\boldsymbol{E}^H & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & \boldsymbol{X}\\ \boldsymbol{W} & \boldsymbol{Y}\end{bmatrix}^{H}$$

With the definition of

$$\begin{aligned}\boldsymbol{U} &= \begin{bmatrix}\boldsymbol{G}^{\frac{1}{2}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{W}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G}^{\frac{1}{2}}\\ \boldsymbol{W}\end{bmatrix}\\ \boldsymbol{V} &= \begin{bmatrix}\boldsymbol{G}^{\frac{1}{2}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G}^{\frac{1}{2}}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}\end{aligned} \tag{3.32}$$

this gives

$$\begin{bmatrix}\boldsymbol{U} & \boldsymbol{V}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{E}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}(\boldsymbol{U}^H\boldsymbol{U})^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}^{-1}\end{bmatrix} \begin{bmatrix}\boldsymbol{1} & \boldsymbol{0}\\ -\boldsymbol{E}^H & \boldsymbol{1}\end{bmatrix} \begin{bmatrix}\boldsymbol{U}^H\\ \boldsymbol{V}^H\end{bmatrix} = \boldsymbol{1}$$

and thus the constraint for $\boldsymbol{X}, \boldsymbol{Y}$ appearing in $\boldsymbol{V}$ can be written in the form of

$$\boldsymbol{1} - \boldsymbol{U}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H = (\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})\boldsymbol{F}^{-1}(\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})^H \tag{3.33}$$

According to (A.80), there exists a $\boldsymbol{U}_\perp$ with

$$\boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-1}\boldsymbol{U}_\perp^H = \boldsymbol{1} - \boldsymbol{U}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H$$

since $\boldsymbol{U}$ as defined in (3.32) has full rank. In order to find $\boldsymbol{U}_\perp$, calculate

$$\boldsymbol{1} - \boldsymbol{U}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H = \begin{bmatrix}\boldsymbol{1} - \boldsymbol{G}^{\frac{1}{2}}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{G}^{\frac{1}{2}} & -\boldsymbol{G}^{\frac{1}{2}}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{W}^H\\ -\boldsymbol{W}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{G}^{\frac{1}{2}} & \boldsymbol{1} - \boldsymbol{W}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{W}^H\end{bmatrix}$$

With

$$(\boldsymbol{U}^H\boldsymbol{U})^{-1} = (\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} = \boldsymbol{G}^{-1} - \boldsymbol{G}^{-1}\boldsymbol{W}^H\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{G}^{-1}$$

and $\boldsymbol{\Delta} = (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-1}$ and further using (A.56) this can be written as

$$\mathbf{1} - \boldsymbol{U}(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H = \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{G}^{-\frac{1}{2}} & -\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{G}^{-\frac{1}{2}} & \boldsymbol{\Delta} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix}^H =$$

$$= \boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-1}\boldsymbol{U}_\perp^H$$

so that

$$\boldsymbol{U}_\perp = \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix} \text{ with } \boldsymbol{U}^H\boldsymbol{U}_\perp = \begin{bmatrix} \boldsymbol{G}^{\frac{1}{2}} \\ \boldsymbol{W} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix} = \mathbf{0} \tag{3.34}$$

is an orthogonal completion of $\boldsymbol{U}$ with full rank. Now (3.33) can be written as

$$\boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-1}\boldsymbol{U}_\perp^H = (\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})\boldsymbol{F}^{-1}(\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})^H$$

which according to (A.77) can be uniquely expanded to

$$\boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-\frac{1}{2}} \left(\boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-\frac{1}{2}} \right)^H = (\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})\boldsymbol{F}^{-\frac{1}{2}} \left((\boldsymbol{V} - \boldsymbol{U}\boldsymbol{E})\boldsymbol{F}^{-\frac{1}{2}} \right)^H$$

so that according to (A.78), $\boldsymbol{V}$ containing the unknown $\boldsymbol{X}, \boldsymbol{Y}$ can be extracted in the form of

$$\boldsymbol{V} = \boldsymbol{U}\boldsymbol{E} + \boldsymbol{U}_\perp(\boldsymbol{U}_\perp^H\boldsymbol{U}_\perp)^{-\frac{1}{2}}\boldsymbol{Z}\boldsymbol{F}^{\frac{1}{2}}$$

with unitary parameter $\boldsymbol{Z}$. With (3.32) and (3.34) this leads to

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \boldsymbol{W} \end{bmatrix} \boldsymbol{E} + \begin{bmatrix} \boldsymbol{G}^{-1}\boldsymbol{W}^H \\ -\mathbf{1} \end{bmatrix} \boldsymbol{S}^H$$

with

$$\boldsymbol{S}^H = (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}}\boldsymbol{Z}\boldsymbol{F}^{\frac{1}{2}} \tag{3.35}$$

and further to

$$\begin{bmatrix} \mathbf{1} & \boldsymbol{X} \\ \boldsymbol{W} & \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \boldsymbol{G}^{-1}\boldsymbol{W}^H \\ \boldsymbol{W} & -\mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \boldsymbol{E} \\ \mathbf{0} & \boldsymbol{S}^H \end{bmatrix}$$

The solution (3.27) now becomes independent of $\boldsymbol{X}, \boldsymbol{Y}$ and using (3.26) can be written as

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-H} \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^{-H} \begin{bmatrix} \mathbf{1} & \boldsymbol{X} \\ \boldsymbol{W} & \boldsymbol{Y} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{G}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{G}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1} & \boldsymbol{X} \\ \boldsymbol{W} & \boldsymbol{Y} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{A}_0\boldsymbol{G}_0^{-1} & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 \\ \boldsymbol{C}_0\boldsymbol{G}_0^{-1} & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{E}^H & \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\mathbf{1} \end{bmatrix}$$

thus

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{A}_0\boldsymbol{G}_0^{-1} + \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{E}^H & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{S} \\ \boldsymbol{C}_0\boldsymbol{G}_0^{-1} + \boldsymbol{D}_0\boldsymbol{E}^H & \boldsymbol{D}_0\boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\boldsymbol{1} \end{bmatrix} \tag{3.36}$$

The matrix $\boldsymbol{E}$ can be calculated using (3.29) and (3.30) and gives

$$\boldsymbol{E} = \boldsymbol{H}_1^{-1}\boldsymbol{H}_2 = (\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1}(\boldsymbol{A}_0^H\boldsymbol{G}\boldsymbol{B}_0 + \boldsymbol{C}_0^H\boldsymbol{D}_0)$$

which can also be written as

$$\boldsymbol{E} = (\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1}\boldsymbol{A}_0^H(\boldsymbol{G} - \boldsymbol{G}_0)\boldsymbol{B}_0 \tag{3.37}$$

by taking the expression $\boldsymbol{C}_0^H\boldsymbol{D}_0$ from (3.26). Rewriting (3.26) in the form of

$$\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{G}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^H = \begin{bmatrix} \boldsymbol{G}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1}$$

provides the following equations

$$\begin{aligned} \boldsymbol{B}_0\boldsymbol{B}_0^H &= \boldsymbol{G}_0^{-1} - \boldsymbol{A}_0\boldsymbol{G}_0^{-1}\boldsymbol{A}_0^H \\ \boldsymbol{D}_0\boldsymbol{B}_0^H &= -\boldsymbol{C}_0\boldsymbol{G}_0^{-1}\boldsymbol{A}_0^H \end{aligned} \tag{3.38}$$

Due to (3.37) and (3.38), the matrix $\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{E}^H$ appearing in the solution (3.36) can be expressed as

$$\begin{aligned} \boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{E}^H &= \boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{B}_0^H(\boldsymbol{G} - \boldsymbol{G}_0)\boldsymbol{A}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} = \\ &= ((\boldsymbol{G} - \boldsymbol{G}_0)\boldsymbol{A}_0 - \boldsymbol{G}_0\boldsymbol{A}_0\boldsymbol{G}_0^{-1}\boldsymbol{A}_0^H(\boldsymbol{G} - \boldsymbol{G}_0)\boldsymbol{A}_0)(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} \end{aligned}$$

Using (3.23) this leads to

$$\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{E}^H = \boldsymbol{G}\boldsymbol{A}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} - \boldsymbol{G}_0\boldsymbol{A}_0\boldsymbol{G}_0^{-1} \tag{3.39}$$

Due to (3.37) and (3.38), the matrix $\boldsymbol{D}_0\boldsymbol{E}^H$ appearing in the solution (3.36) can be expressed as

$$\begin{aligned} \boldsymbol{D}_0\boldsymbol{E}^H &= \boldsymbol{D}_0\boldsymbol{B}_0^H(\boldsymbol{G} - \boldsymbol{G}_0)\boldsymbol{A}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} = \\ &= \boldsymbol{C}_0\boldsymbol{G}_0^{-1}\boldsymbol{A}_0^H(\boldsymbol{G}_0 - \boldsymbol{G})\boldsymbol{A}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} \end{aligned}$$

Using (3.23) this leads to

$$\boldsymbol{D}_0\boldsymbol{E}^H = \boldsymbol{C}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} - \boldsymbol{C}_0\boldsymbol{G}_0^{-1} \tag{3.40}$$

Substituting the expressions (3.39), (3.40) in (3.36) leads to

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{S} \\ \boldsymbol{C}_0(\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} & \boldsymbol{D}_0\boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\boldsymbol{1} \end{bmatrix}$$

so that (3.24) follows. The matrix $\boldsymbol{F}$ can be calculated by comparing the expressions for $\boldsymbol{H}^{-1}$ obtained by inverting (3.29) and (3.31). The first one gives

$$\begin{aligned} \boldsymbol{H}^{-1} &= \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^{-H} = \\ &= \begin{bmatrix} \boldsymbol{G}_0^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{G}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{G}_0^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \end{aligned}$$

using (3.26). The second one gives

$$\boldsymbol{H}^{-1} = \begin{bmatrix} \mathbf{1} & -\boldsymbol{E} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{F}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\boldsymbol{E}^H & \mathbf{1} \end{bmatrix}$$

Comparing the lower right block provides

$$\boldsymbol{F} = (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-1}$$

which can be substituted in (3.35) so that

$$\boldsymbol{S}(\boldsymbol{B}_0, \boldsymbol{D}_0, \boldsymbol{W}, \boldsymbol{Z}) = (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-\frac{1}{2}} \boldsymbol{Z} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} \tag{3.41}$$

with unitary parameter $\boldsymbol{Z}$. This can also be written as

$$\begin{aligned} &\boldsymbol{S}(\boldsymbol{B}_0, \boldsymbol{D}_0, \boldsymbol{W}, \boldsymbol{Z}) = \\ &\quad = \boldsymbol{Z}\boldsymbol{Z}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-\frac{1}{2}} \boldsymbol{Z} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} = \\ &\quad = \boldsymbol{Z} (\boldsymbol{Z}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0) \boldsymbol{Z})^{-\frac{1}{2}} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} = \\ &\qquad\qquad = \boldsymbol{Z}\boldsymbol{S}(\boldsymbol{B}_0\boldsymbol{Z}, \boldsymbol{D}_0\boldsymbol{Z}, \boldsymbol{W}, \mathbf{1}) \end{aligned}$$

using (A.76). The solution (3.24) then gives

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} &= f(\boldsymbol{A}_0, \boldsymbol{B}_0, \boldsymbol{C}_0, \boldsymbol{D}_0, \boldsymbol{W}, \boldsymbol{Z}) = \\ &= \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{T} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}\boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\mathbf{1} \end{bmatrix} = \\ &= \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{Z} \\ \boldsymbol{C}_0 & \boldsymbol{D}_0\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{T} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\mathbf{1} \end{bmatrix} = \\ &\qquad\qquad = f(\boldsymbol{A}_0, \boldsymbol{B}_0\boldsymbol{Z}, \boldsymbol{C}_0, \boldsymbol{D}_0\boldsymbol{Z}, \boldsymbol{W}, \mathbf{1}) \end{aligned}$$

with $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{B}_0\boldsymbol{Z}, \boldsymbol{D}_0\boldsymbol{Z}, \boldsymbol{W}, \mathbf{1})$ given by (3.41). According to (3.6), this shows that the unitary $\boldsymbol{Z}$ parametrizes all lossless completion pairs $(\boldsymbol{B}_0\boldsymbol{Z}, \boldsymbol{D}_0\boldsymbol{Z})$ and thus the solution can be written in the form of (3.24) with $\boldsymbol{S}$ given by (3.25) and $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ being any lossless completion pair.

Remark Note that for a unitary matrix $\boldsymbol{U}$ and according to (A.76) we can write

$$\begin{aligned} &(\boldsymbol{U}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0) \boldsymbol{U})^{-\frac{1}{2}} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} = \\ &\quad = \boldsymbol{U}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-\frac{1}{2}} \boldsymbol{U} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} \boldsymbol{U}^H \boldsymbol{U} = \\ &\quad = \boldsymbol{U}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-\frac{1}{2}} (\boldsymbol{U} (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H) \boldsymbol{U}^H)^{-\frac{1}{2}} \boldsymbol{U} \end{aligned}$$

so that the functions $\boldsymbol{S}(\boldsymbol{B}_0, \boldsymbol{D}_0, \boldsymbol{W}), \boldsymbol{T}(\boldsymbol{W})$ given in (3.25) satisfy

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{B}_0\boldsymbol{U}, \boldsymbol{D}_0\boldsymbol{U}, \boldsymbol{W}) &= \boldsymbol{U}^H \boldsymbol{S}(\boldsymbol{B}_0, \boldsymbol{D}_0, \boldsymbol{U}\boldsymbol{W}) \boldsymbol{U} \\ \boldsymbol{T}(\boldsymbol{W}) &= \boldsymbol{T}(\boldsymbol{U}\boldsymbol{W}) \end{aligned}$$

hence (3.24) satisfies

$$\begin{bmatrix} \boldsymbol{A}(\boldsymbol{B}_0\boldsymbol{U},\boldsymbol{D}_0\boldsymbol{U},\boldsymbol{W}) & \boldsymbol{B}(\boldsymbol{B}_0\boldsymbol{U},\boldsymbol{D}_0\boldsymbol{U},\boldsymbol{W}) \\ \boldsymbol{C}(\boldsymbol{B}_0\boldsymbol{U},\boldsymbol{D}_0\boldsymbol{U},\boldsymbol{W}) & \boldsymbol{D}(\boldsymbol{B}_0\boldsymbol{U},\boldsymbol{D}_0\boldsymbol{U},\boldsymbol{W}) \end{bmatrix} = \\ = \begin{bmatrix} \boldsymbol{A}(\boldsymbol{B}_0,\boldsymbol{D}_0,\boldsymbol{W}\boldsymbol{U}) & \boldsymbol{B}(\boldsymbol{B}_0,\boldsymbol{D}_0,\boldsymbol{W}\boldsymbol{U})\boldsymbol{U} \\ \boldsymbol{C}(\boldsymbol{B}_0,\boldsymbol{D}_0,\boldsymbol{W}\boldsymbol{U}) & \boldsymbol{D}(\boldsymbol{B}_0,\boldsymbol{D}_0,\boldsymbol{W}\boldsymbol{U})\boldsymbol{U} \end{bmatrix} \tag{3.42}$$

3.2.3 Link to Conjugate Realization

Proposition If the observable pair $(\boldsymbol{A}_0,\boldsymbol{C}_0)$ of the discrete stable minimal lossless system $(\boldsymbol{A}_0,\boldsymbol{B}_0,\boldsymbol{C}_0,\boldsymbol{D}_0)$ has a matrix $\boldsymbol{I}$ that solves

$$\begin{bmatrix} \boldsymbol{A}_0 \\ \boldsymbol{C}_0 \end{bmatrix} \boldsymbol{I} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_0 \\ \boldsymbol{C}_0 \end{bmatrix}} \tag{3.43}$$

then $\boldsymbol{I}$ is unique and invertible and the observable pair $(\boldsymbol{A},\boldsymbol{C})$ of the discrete stable minimal lossless system $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D})$ defined by (3.24) satisfies

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{I} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}} \Leftrightarrow \overline{\boldsymbol{W}} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{I} \tag{3.44}$$

with unitary $\boldsymbol{U}$ defined by

$$\boldsymbol{U} = \overline{\boldsymbol{B}}^H \boldsymbol{I}^H \boldsymbol{G}\boldsymbol{B} + \overline{\boldsymbol{D}}^H \boldsymbol{D} = \overline{\boldsymbol{B}}_0^H \boldsymbol{I}^H \boldsymbol{G}_0\boldsymbol{B}_0 + \overline{\boldsymbol{D}}_0^H \boldsymbol{D}_0 \tag{3.45}$$

Proof First suppose there exists a matrix $\boldsymbol{I}$ solving (3.43) then it is unique and invertible due to (3.10). According to (3.11) the matrix $\boldsymbol{I}$ also solves

$$\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \tag{3.46}$$

with unitary $\boldsymbol{U}$ given by

$$\boldsymbol{U} = \overline{\boldsymbol{B}}_0^H \boldsymbol{I}^H \boldsymbol{G}_0\boldsymbol{B}_0 + \overline{\boldsymbol{D}}_0^H \boldsymbol{D}_0 \tag{3.47}$$

according to (3.12) and Gram observability matrix $\boldsymbol{G}_0$ satisfying

$$\overline{\boldsymbol{G}}_0 = \boldsymbol{I}^H \boldsymbol{G}_0 \boldsymbol{I} \tag{3.48}$$

according to (3.13).

To verify sufficiency suppose there exists a matrix $\boldsymbol{I}'$ that solves

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{I}' = \begin{bmatrix} \boldsymbol{I}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}}$$

then it is unique and invertible due to (3.10). According to (3.11) the matrix $\boldsymbol{I}'$ also solves

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}'^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}' \end{bmatrix} \tag{3.49}$$

with unitary $\boldsymbol{U}'$ given by

$$\boldsymbol{U}' = \overline{\boldsymbol{B}}^H \boldsymbol{I}'^H \boldsymbol{G}\boldsymbol{B} + \overline{\boldsymbol{D}}^H \boldsymbol{D} \tag{3.50}$$

according to (3.12) and Gram observability matrix $\boldsymbol{G}$ satisfying

$$\overline{\boldsymbol{G}} = \boldsymbol{I'}^H \boldsymbol{G} \boldsymbol{I'} \tag{3.51}$$

according to (3.13). The constraint (3.16) can be rearranged in the form of

$$\begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{W} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A}_0 \\ \boldsymbol{C}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_0 \\ \boldsymbol{C}_0 \end{bmatrix}$$

using (3.3) since $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ is discrete stable minimal lossless. This provides

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{A}^H \boldsymbol{G} \boldsymbol{A}_0 + \boldsymbol{C}^H \boldsymbol{C}_0 \\ \boldsymbol{W} &= \boldsymbol{B}^H \boldsymbol{G} \boldsymbol{A}_0 + \boldsymbol{D}^H \boldsymbol{C}_0 \end{aligned} \tag{3.52}$$

With (3.46), (3.48) and (3.49), (3.51) the conjugate of (3.52) provides

$$\begin{aligned} \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1} &= \boldsymbol{A}^H \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1} \boldsymbol{A}_0 + \boldsymbol{C}^H \boldsymbol{C}_0 \\ \boldsymbol{U'}^H \overline{\boldsymbol{W}} \boldsymbol{I}^{-1} &= \boldsymbol{B}^H \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1} \boldsymbol{A}_0 + \boldsymbol{D}^H \boldsymbol{C}_0 \end{aligned}$$

Subtracting this from (3.52) gives

$$\begin{aligned} \boldsymbol{A}^H (\boldsymbol{G} - \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1}) \boldsymbol{A}_0 - (\boldsymbol{G} - \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1}) &= \boldsymbol{0} \\ \boldsymbol{B}^H (\boldsymbol{G} - \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1}) \boldsymbol{A}_0 - (\boldsymbol{W} - \boldsymbol{U'}^H \overline{\boldsymbol{W}} \boldsymbol{I}^{-1}) &= \boldsymbol{0} \end{aligned}$$

The first equation is a Stein equation and according to (A.88) has a unique solution $\boldsymbol{G} = \boldsymbol{G} \boldsymbol{I'} \boldsymbol{I}^{-1}$ since $\boldsymbol{A}, \boldsymbol{A}_0$ are stable. Since $\boldsymbol{G} > 0$ this provides

$$\boldsymbol{I'} = \boldsymbol{I} \tag{3.53}$$

The second equation then provides

$$\overline{\boldsymbol{W}} = \boldsymbol{U'} \boldsymbol{W} \boldsymbol{I} \tag{3.54}$$

With (3.46), (3.48), (3.49), (3.51) and (3.53), (3.54) it follows from (3.25) that $\boldsymbol{T} = \boldsymbol{I} \overline{\boldsymbol{T}} \boldsymbol{I}^H$ and the solution (3.24) can be written as

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U'} \end{bmatrix} = \\ = \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^H \overline{\boldsymbol{S}} \boldsymbol{U'} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\boldsymbol{1} \end{bmatrix} \end{aligned}$$

This can be rewritten in the form

$$\begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^H \overline{\boldsymbol{S}} \boldsymbol{U'} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{W}^H \\ \boldsymbol{W} & -\boldsymbol{1} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix}$$

hence

$$\overline{\boldsymbol{S}} = \boldsymbol{U} \boldsymbol{S} \boldsymbol{U'}^H \tag{3.55}$$

With (3.46), (3.48), (3.49), (3.51) and (3.53), (3.54) the conjugate of $\boldsymbol{S}$ in (3.25) gives

$$\overline{\boldsymbol{S}} = \boldsymbol{U} (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0)^{-\frac{1}{2}} \boldsymbol{U}^H \boldsymbol{U'} (\boldsymbol{1} + \boldsymbol{W} \boldsymbol{G}^{-1} \boldsymbol{W}^H)^{-\frac{1}{2}} \boldsymbol{U'}^H$$

using (A.76). With (3.55) and the fact that the matrices $(\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 + \boldsymbol{D}_0^H\boldsymbol{D}_0)^{-\frac{1}{2}}, (\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}}$ are invertible this provides $\boldsymbol{U}^H\boldsymbol{U}' = \mathbf{1}$ hence

$$\boldsymbol{U}' = \boldsymbol{U} \tag{3.56}$$

Thus (3.45) follows from (3.46) and (3.49) using (3.53), (3.56) and (3.44) follows from (3.54) using (3.56), which verifies sufficiency.

To verify necessity suppose that parameter $\boldsymbol{W}$ satisfies

$$\overline{\boldsymbol{W}} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{I} \tag{3.57}$$

with unitary $\boldsymbol{U}$ given by (3.47). With (3.43) and (3.57) the conjugate of (3.15) provides

$$\boldsymbol{A}_0^H\boldsymbol{I}^{-H}\overline{\boldsymbol{G}}\boldsymbol{I}^{-1}\boldsymbol{A}_0 - \boldsymbol{I}^{-H}\overline{\boldsymbol{G}}\boldsymbol{I}^{-1} + \boldsymbol{C}_0^H\boldsymbol{C}_0 = \boldsymbol{W}^H\boldsymbol{W}$$

Subtracting this from (3.15) gives

$$\boldsymbol{A}_0^H(\boldsymbol{G} - \boldsymbol{I}^{-H}\overline{\boldsymbol{G}}\boldsymbol{I}^{-1})\boldsymbol{A}_0 - (\boldsymbol{G} - \boldsymbol{I}^{-H}\overline{\boldsymbol{G}}\boldsymbol{I}^{-1}) = \mathbf{0}$$

This is a Stein equation and according to (A.88) has a unique solution

$$\overline{\boldsymbol{G}} = \boldsymbol{I}^H\boldsymbol{G}\boldsymbol{I} \tag{3.58}$$

since $\boldsymbol{A}_0$ is stable. With (3.46), (3.48) and (3.58) the definitions (3.25) provide

$$\begin{aligned} \overline{\boldsymbol{S}} &= \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^H \\ \overline{\boldsymbol{T}} &= \boldsymbol{I}^{-1}\boldsymbol{T}\boldsymbol{I}^{-H} \end{aligned} \tag{3.59}$$

using (A.76). With (3.46), (3.48) and (3.57), (3.58), (3.59) it can be seen that the solution (3.24) satisfies

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U} \end{bmatrix}$$

so that (3.44) follows and according to (3.12) and with (3.47) matrix $\boldsymbol{U}$ is given by (3.45), which verifies necessity.

3.2.4 Parametrization of Stable Observable Pairs

Proposition The set of all stable observable pairs $(\boldsymbol{A}, \boldsymbol{C})$ with Gram observability matrix $\boldsymbol{G} > 0$ is a differentiable manifold and can be parametrized by a matrix $\boldsymbol{W}$ of same dimension than $\boldsymbol{C}$ in the way

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{A}_0\boldsymbol{T}\boldsymbol{G} + \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{S}\boldsymbol{W} \\ \boldsymbol{C} &= \boldsymbol{C}_0\boldsymbol{T}\boldsymbol{G} + \boldsymbol{D}_0\boldsymbol{S}\boldsymbol{W} \end{aligned} \tag{3.60}$$

with

$$\begin{aligned} \boldsymbol{S} &= (\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 + \boldsymbol{D}_0^H\boldsymbol{D}_0)^{-\frac{1}{2}}(\mathbf{1} + \boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H)^{-\frac{1}{2}} \\ \boldsymbol{T} &= (\boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W})^{-1} \end{aligned} \tag{3.61}$$

while $(\boldsymbol{A}_0, \boldsymbol{C}_0)$ is a stable observable pair with Gram observability matrix $\boldsymbol{G}_0$ and lossless completion $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ as in (3.6). The Gram observability matrices are uniquely determined by

$$\begin{aligned} \boldsymbol{A}_0^H \boldsymbol{G}_0 \boldsymbol{A}_0 - \boldsymbol{G}_0 + \boldsymbol{C}_0^H \boldsymbol{C}_0 &= \boldsymbol{0} \\ \boldsymbol{A}_0^H \boldsymbol{G} \boldsymbol{A}_0 - \boldsymbol{G} + \boldsymbol{C}_0^H \boldsymbol{C}_0 &= \boldsymbol{W}^H \boldsymbol{W} \end{aligned} \tag{3.62}$$

and $\boldsymbol{G} \leq \boldsymbol{G}_0$ must obey the condition

$$\boldsymbol{G} > 0 \tag{3.63}$$

Proof This parametrization is a particular consequence of the solution to the Nudelman interpolation problem (3.24), (3.25). It further follows from (3.42) that any lossless completion pair $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ can be chosen since they are all parametrized by $(\boldsymbol{B}_0\boldsymbol{U}, \boldsymbol{D}_0\boldsymbol{U})$ according to (3.6). The effect of a different completion pair is only that in order to reach the same pair $(\boldsymbol{A}, \boldsymbol{C})$ a different parameter $\boldsymbol{W}$ has to be used.

Remark Moreover it has been shown in [2] that the set of lossless systems and the set of stable observable pairs are *differentiable manifolds*. A differentiable manifold is a smooth curved surface of a dimension d that locally resembles the well known straight euclidean vector space $\mathbb{R}^d$. This means that locally, calculations can be carried out in the usual straight euclidean space instead of the curved manifold. The local link between the euclidean space and the manifold is provided by a local *chart*. A complete collection of local charts is called an *atlas* and globally describes the curved manifold like a complete collection of planar road maps describe the curved surface of earth. Topologically, these charts are an *homeomorphism* between an open subset of the curved manifold and an open subset of the straight $\mathbb{R}^d$ whereby an homeomorphism is a bicontinuous function, that is a continuous function which has a continuous inverse function. Such an homeomorphism allows for a point on the manifold to find the corresponding point on the local chart and the other way round.

In this context, the stable observable pair $(\boldsymbol{A}_0, \boldsymbol{C}_0)$ with Gram observability matrix $\boldsymbol{G}_0$ is the center of a chart since for $\boldsymbol{W} = \boldsymbol{0}$ we get $\boldsymbol{G} = \boldsymbol{G}_0$ from (3.62) and further $\boldsymbol{S} = \boldsymbol{1}, \boldsymbol{T} = \boldsymbol{G}^{-1}$ from (3.61) and (3.3) which states that $\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0 = \boldsymbol{1}$ so that by (3.60) we get $\boldsymbol{A} = \boldsymbol{A}_0, \boldsymbol{C} = \boldsymbol{C}_0$ for $\boldsymbol{W} = \boldsymbol{0}$. The stable observable pair $(\boldsymbol{A}, \boldsymbol{C})$ is displaced from the center as a continuous function of $\boldsymbol{W}$ while observing the condition (3.63). If this condition is close to be violated, another chart of the atlas has to be selected, which is done by centering on the current pair $(\boldsymbol{A}_0' = \boldsymbol{A}, \boldsymbol{C}_0' = \boldsymbol{C})$ balancing this realization $(\boldsymbol{A}_0 = \boldsymbol{J}\boldsymbol{A}_0'\boldsymbol{J}^{-1}, \boldsymbol{C}_0 = \boldsymbol{C}_0'\boldsymbol{J}^{-1})$ with $\boldsymbol{J}$ defined by $\boldsymbol{G} = \boldsymbol{J}^H\boldsymbol{J}$ so that $\boldsymbol{G}_0 = \boldsymbol{1}$ and finally resetting $\boldsymbol{W} = \boldsymbol{0}$. Balancing is necessary because otherwise the Gram observability matrix of the new center pair remains $\boldsymbol{G}_0 = \boldsymbol{G}$ which is a poor choice since this value triggered the changing of the chart by almost violating condition (3.63). Since condition (3.63) defines an open set, it is fulfilled on the center at $\boldsymbol{W} = \boldsymbol{0}$ by $\boldsymbol{G} = \boldsymbol{G}_0 > 0$ but it will also be fulfilled in some neighborhood of the center, so that this condition does not prevent this parametrization from becoming differentiable.

Chapter 4

Multiobjective Optimization

This chapter sets up a multiobjective optimization problem. For a given plant $\boldsymbol{P}$, a controller $\boldsymbol{K}$ has to be found which is optimal with respect to a given cost function and within given constraints. These optimization objectives are formulated in terms of L1, L2/H2, H∞ system norms. The advantage of using the L1, L2/H2, H∞ norms is that the calculation of these norms can themselves be set up as standard semidefinite optimization problems which allows the local sensitivities and eventually the subgradient of the multiobjective optimization problem to be calculated. Formulating the optimization problem in terms of the parametrization found in chapters 2 and 3 then allows to move the parameter $\boldsymbol{Q}_L$ instead of $\boldsymbol{K}$ infinitesimally in the descending direction indicated by the subgradient. Once the optimization task is complete, the controller $\boldsymbol{K}$ can be regained from the parameter $\boldsymbol{Q}_L$. This enables a local improvement of a provided initial controller $\boldsymbol{K}_0$ to a better one $\boldsymbol{K}^*$ with respect to the cost function and within the constraints.

4.1 Problem Formulation

In this section, an extended version of the plant is constructed that incorporates various filters and the plant itself. Together with the controller to be designed, this leads to an extended feedback loop that provides more freedom in the formulation of the optimization objectives. The objectives can then be defined on certain channels and over certain frequencies of this system.

4.1.1 Extended Loops

The plant $\boldsymbol{P}$ that is to be controlled by the controller $\boldsymbol{K}$ is extended by filter systems $\boldsymbol{P}_{zw_i}, \boldsymbol{P}_{yw_i}, \boldsymbol{P}_{zu_i}$ to form the extended plant $\boldsymbol{P}_i$, which together with the controller $\boldsymbol{K}$ forms the extended loop $\boldsymbol{H}_i$ as shown in figure 4.1. These filters can be used to select performance channels and to emphasize certain frequencies. The fact that these filters become a part of the system to be optimized makes the formulation of objectives easier. A bridge structure (C.56) has been chosen for the extended plant $\boldsymbol{P}_i$ to cover all possible filtering situations. Many different

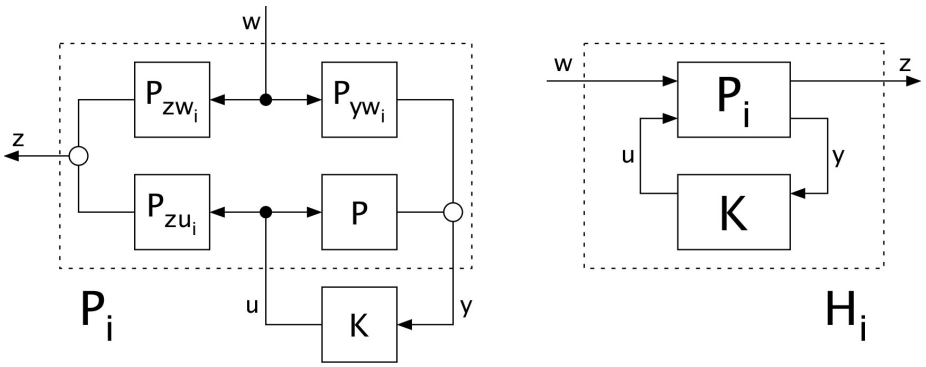


Figure 4.1: The plant $\boldsymbol{P}$ is extended by filters $\boldsymbol{P}_{zw_i}, \boldsymbol{P}_{yw_i}, \boldsymbol{P}_{zu_i}$ to form the extended plant $\boldsymbol{P}_i$, which together with the controller $\boldsymbol{K}$ forms the extended loop $\boldsymbol{H}_i$.

extended plants $\boldsymbol{P}_i$ and loops $\boldsymbol{H}_i$ can be defined and the index i is used to distinguish them.

Proposition The extended loop shown in figure 4.1 has a transfer matrix $\boldsymbol{H}_i$ given by

$$\boldsymbol{H}_i = \boldsymbol{P}_{zw_i} + \boldsymbol{P}_{zu_i}(\mathbf{1} - \boldsymbol{K}\boldsymbol{P})^{-1}\boldsymbol{K}\boldsymbol{P}_{yw_i} \tag{4.1}$$

A state space representation of $\boldsymbol{H}_i$ is given by

$$\boldsymbol{H}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right] =$$

$$= \left[\begin{array}{cccc|cc} \boldsymbol{A}_{zw_i} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{B}_{zw_i} \\ \mathbf{0} & \boldsymbol{A}_{yw_i} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{B}_{yw_i} \\ \mathbf{0} & \boldsymbol{B}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_{yw_i} & \boldsymbol{A}_{zu_i} & \boldsymbol{B}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{B}_{zu_i}\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{B}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \\ \mathbf{0} & \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_{yw_i} & \mathbf{0} & \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \\ \mathbf{0} & \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_{yw_i} & \mathbf{0} & \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \\ \hline \boldsymbol{C}_{zw_i} & \boldsymbol{D}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_{yw_i} & \boldsymbol{C}_{zu_i} & \boldsymbol{D}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \boldsymbol{D}_{zu_i}\boldsymbol{E}_{KP}\boldsymbol{C}_K & \boldsymbol{D}_{zw_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \end{array}\right] \tag{4.2}$$

In the important special case of the filter system matrices satisfying

$$\begin{aligned} \boldsymbol{A}_{zw_i} = \boldsymbol{A}_{yw_i} = \boldsymbol{A}_{zu_i} &= \boldsymbol{A}_P \\ \boldsymbol{B}_{zw_i} = \boldsymbol{B}_{yw_i} &= \boldsymbol{B}_{w_i} \\ \boldsymbol{B}_{zu_i} &= \boldsymbol{B}_P \\ \boldsymbol{C}_{zw_i} = \boldsymbol{C}_{zu_i} &= \boldsymbol{C}_{z_i} \\ \boldsymbol{C}_{yw_i} &= \boldsymbol{C}_P \end{aligned} \tag{4.3}$$

this state space representation reduces to

$$\boldsymbol{H}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right] = \left[\begin{array}{cc|c} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \cdots & \boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K \\ \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \cdots & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K \\ \hline \boldsymbol{C}_{z_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P & \cdots & \boldsymbol{D}_{zu_i}\boldsymbol{E}_{KP}\boldsymbol{C}_K \end{array}\right.$$
$$\left.\begin{array}{c|c} \boldsymbol{B}_{w_i} + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \\ \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \\ \hline \boldsymbol{D}_{zw_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{D}_{yw_i} \end{array}\right] \tag{4.4}$$

Proof The transfer matrix (4.1) follows from figure 4.1. Using $\boldsymbol{H}$ as defined in (2.2) it can be seen that the transfer matrix can also be given by

$$\boldsymbol{H}_i = \boldsymbol{P}_{zw_i} + \boldsymbol{P}_{zu_i}(\mathbf{1} - \boldsymbol{K}\boldsymbol{P})^{-1}\boldsymbol{K}\boldsymbol{P}_{yw_i} = \boldsymbol{P}_{zw_i} + \boldsymbol{P}_{zu_i}\begin{bmatrix}\mathbf{0} & \mathbf{1}\end{bmatrix}\boldsymbol{H}\begin{bmatrix}\mathbf{0} \\ \mathbf{1}\end{bmatrix}\boldsymbol{P}_{yw_i}$$

Now let the state space representations of plant $\boldsymbol{P}$, controller $\boldsymbol{K}$ and loop $\boldsymbol{H}$ be given by (2.18) and (2.19) and the filters $\boldsymbol{P}_{zw_i}, \boldsymbol{P}_{yw_i}, \boldsymbol{P}_{zu_i}$ be given by

$$\boldsymbol{P}_{kl_i} = \left[\begin{array}{c|c} \boldsymbol{A}_{kl_i} & \boldsymbol{B}_{kl_i} \\ \hline \boldsymbol{C}_{kl_i} & \boldsymbol{D}_{kl_i} \end{array}\right]$$

with $k \in \{y, z\}$ and $l \in \{u, w\}$. Then the state space representation (4.2) follows from this transfer matrix using (C.46), (C.44) and (2.19). In the special case given by (4.3) the state space representation (4.2) then further reduces to (4.4) by applying simple state transformations that allow to cut off uncontrollable and unobservable parts.

4.1.2 System Norms

In order to find an optimal controller, a method of comparing the effects of different controllers is required. One such method is to define norms on the input and output signals of the extended loop $\boldsymbol{H}_i$ to measure its performance. The two most common signal norms are the L2 norm that measures the energy, hence a sort of average of the signal, and the L∞ norm that measure the peak value of the signal. The following inequalities taken from [14] relate the norms of input and output

$$\begin{aligned} ||\boldsymbol{z}(t)||_{L2} &\leq ||\boldsymbol{H}_i(s)||_{H\infty}\,||\boldsymbol{w}(t)||_{L2} \\ ||\boldsymbol{z}(t)||_{L\infty} &\leq ||\boldsymbol{H}_i(t)||_{L1}\,||\boldsymbol{w}(t)||_{L\infty} \\ ||\boldsymbol{z}(t)||_{L\infty} &\leq ||\boldsymbol{H}_i(t)||_{L2}\,||\boldsymbol{w}(t)||_{L2} = ||\boldsymbol{H}_i(s)||_{H2}\,||\boldsymbol{w}(t)||_{L2} \end{aligned}$$

The correspondence between the L2 and H2 norm in the last inequality is due to the Parseval theorem (C.9). These inequalities show that $\boldsymbol{H}_i$ needs to be stable for the L1, L2/H2, H∞ norms to remain bounded. If the filters of the extended plant $\boldsymbol{P}_i$ as in figure 4.1 are selected so that the input $\boldsymbol{w}_i$ consists of external disturbances and the output $\boldsymbol{z}_i$ contains error signals that are to be kept small, then the system norms L1, L2/H2, H∞ can be used on $\boldsymbol{H}_i$ to minimize or limit the effect of the disturbances on the error signals. This explains the interest to use the system norms L1, L2/H2, H∞ as objectives in system specifications. These norms or best upper bounds on them are presented in the following in

the form of matrix inequalities for a discrete system $\boldsymbol{H}_i$. To our best knowledge there exists yet no matrix inequality form for the L1 norm. However, an upper bound, called the star norm, has been found by replacing the reachable set of the system state trajectories by the tightest ellipsoid that contains the reachable set, see [1], [9], [12] and section C.1.4. The intermediate variables $\alpha_i, \beta_i, \gamma_i, \boldsymbol{Z}_i$ are called decision variables. The variables α_i, β_i follow from the application of the so called S-lemma on quadratic constraints as demonstrated in [1] for continuous systems. The variables γ_i represent an upper bound on the norm and the hermitian and positive definite matrices $\boldsymbol{Z}_i$ define Lyapunov functions or are Gram controllability or observability matrices.

L1 Norm An upper bound is the star norm derived in [1], [9] and leads to

$$||\boldsymbol{H}_i[k]||_{L1} \leq ||\boldsymbol{H}_i[k]||_* \leq \gamma_i \tag{4.5}$$

if and only if $\exists \boldsymbol{Z}_i > 0, \alpha_i \in [0,1], \beta_i \in [0, \gamma_i^2]$ so that

$$\begin{bmatrix} \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + (\alpha_i - 1)\boldsymbol{Z}_i & \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i \\ \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i & \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i - \alpha_i \boldsymbol{1} \end{bmatrix} \leq 0$$
$$\begin{bmatrix} \boldsymbol{C}_i^H \boldsymbol{C}_i - \beta_i \boldsymbol{Z}_i & \boldsymbol{C}_i^H \boldsymbol{D}_i \\ \boldsymbol{D}_i^H \boldsymbol{C}_i & \boldsymbol{D}_i^H \boldsymbol{D}_i + (\beta_i - \gamma_i^2)\boldsymbol{1} \end{bmatrix} \leq 0 \tag{4.6}$$

L2/H2 Norm According to (C.12), (C.13) direct calculation leads to

$$||\boldsymbol{H}_i[k]||_{L2} = ||\boldsymbol{H}_i(z)||_{H2} < \gamma_i \tag{4.7}$$

if and only if $\exists \boldsymbol{Z}_i > 0$ so that

$$\begin{aligned} \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + \boldsymbol{C}_i^H \boldsymbol{C}_i - \boldsymbol{Z}_i < 0 \\ \mathrm{tr}(\boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i + \boldsymbol{D}_i^H \boldsymbol{D}_i) - \gamma_i^2 < 0 \end{aligned} \tag{4.8}$$

H∞ Norm It follows from the strict bounded real lemma that

$$||\boldsymbol{H}_i(z)||_{H\infty} < \gamma_i \tag{4.9}$$

if and only if $\exists \boldsymbol{Z}_i > 0$ so that

$$\begin{bmatrix} \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + \boldsymbol{C}_i^H \boldsymbol{C}_i - \boldsymbol{Z}_i & \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i + \boldsymbol{C}_i^H \boldsymbol{D}_i \\ \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + \boldsymbol{D}_i^H \boldsymbol{C}_i & \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i + \boldsymbol{D}_i^H \boldsymbol{D}_i - \gamma_i^2 \boldsymbol{1} \end{bmatrix} < 0 \tag{4.10}$$

4.1.3 Optimization Problem

The typical optimization problem is formulated in terms of optimization objectives that have to be met. These objectives consist of a cost function and constraints. The constraints limit the set of all possible controllers to a subset of admissible controllers. Among these admissible controllers, the set of best controllers can then be found by minimizing the scalar cost function. Here, each objective is defined in terms of one of the norms L1, L2/H2, H∞ presented in the previous section, acting on one extended loop $\boldsymbol{H}_i$. There can be as many different extended loops as objectives. Each objective comes with a set of decision variables $\alpha_i \geq 0, \beta_i \geq 0, \gamma_i > 0, \boldsymbol{Z}_i > 0$.

Proposition The multiobjective optimization problem of finding the controller $\boldsymbol{K}$ that minimizes a cost function under constraints formulated in terms of the L1, L2/H2, H∞ norms can be given in the general form

$$\boldsymbol{K}^* = \arg \inf_{\substack{\boldsymbol{\alpha} \geq 0 \\ \boldsymbol{\beta} \geq 0 \\ \boldsymbol{\gamma} > 0 \\ \boldsymbol{\epsilon} > 0 \\ \boldsymbol{Z} > 0 \\ \boldsymbol{K}}} f(\boldsymbol{\gamma}) \text{ subject to } \boldsymbol{G}'(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{K}) \leq 0 \tag{4.11}$$

The vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ contain the decision variables $\alpha_i, \beta_i, \gamma_i$ and the block diagonal matrix $\boldsymbol{Z}$ contains the decision variables $\boldsymbol{Z}_i$ of the different objectives. In addition to these, the vector $\boldsymbol{\epsilon}$ contains variables ϵ_i that are introduced to formally transform definite constraints of the form $\boldsymbol{M} < 0$ to semidefinite constraints by $\boldsymbol{M} + \epsilon \boldsymbol{1} \leq 0$ with $\epsilon > 0$.

Proof The decision variables γ_i in (4.5), (4.7), (4.9) represent upper bounds of the system norms, so the cost f will be formulated as a function of these. The constraints are then given by a combination of (4.5), (4.7), (4.9) together with user defined norm limits such as $\gamma \leq \gamma_{max}$. Aside from the decision variables $\alpha_i, \beta_i, \boldsymbol{Z}_i$, these constraints depend on the realizations $(\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{C}_i, \boldsymbol{D}_i)$ of the extended loops $\boldsymbol{H}_i$ which according to (4.1) are a function of $\boldsymbol{K}$. The ϵ_i are used to formally transform definite constraints $\boldsymbol{M}_i < 0$ into semidefinite constraints by $\boldsymbol{M}_i + \epsilon_i \boldsymbol{1} \leq 0$ with $\epsilon_i > 0$. All constraints can then be combined in the single unified semidefinite form $\boldsymbol{G}' \leq 0$ used in (4.11) by putting the separate semidefinite constraints as diagonal entries $\boldsymbol{G}'_i$ into a block diagonal $\boldsymbol{G}'$. Then $\boldsymbol{G}' \leq 0$ if and only if all $\boldsymbol{G}'_i = [\boldsymbol{G}']_{ii} \leq 0$.

4.2 Parametrized Optimization

In theory, the solution of (4.11) yields the optimal controller $\boldsymbol{K}^*$ but in practice, the problem is difficult to solve since the constraints are bilinear. The solution proposed in this section is therefore based on the idea of a local optimization. The fixed order parametrization developed in chapter 2 parametrizes all stable loops $\boldsymbol{H}$ formed by the controller with a given plant $\boldsymbol{P}$ with stable parameters $\boldsymbol{Q}_L, \boldsymbol{Q}_R$. The parameter $\boldsymbol{Q}_L$ is observable if and only if $\boldsymbol{K}$ is observable. It is therefore sensible to concentrate on stable observable $\boldsymbol{Q}_L$, so that the parametrization of stable observable pairs developed in chapter 3 can be applied. Since stable observable $\boldsymbol{Q} = \boldsymbol{Q}_L$ form a differentiable manifold, a subgradient can be calculated from a local sensitivity information so that a local optimization can be carried out. However, this requires the constraints to be formulated as linear matrix inequalities in order to get access to the sensitivity information. Therefore, a two step optimization is proposed. In an evaluation step the controller parameter $\boldsymbol{Q}$ and the L1 parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are kept constant so that the L1, L2/H2, H∞ norm constraints become linear matrix inequalities in the remaining variables and the sensitivity information can be accessed. In an actualization step the local subgradient is calculated from the sensitivity matrix

so that the parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ can be changed in order to improve the cost function.

4.2.1 Passage between Controller and Parameter

Proposition If the Riccati equation

$$\begin{aligned}\boldsymbol{R}(\boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P) + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K\boldsymbol{R} = \\ = \boldsymbol{B}_K\boldsymbol{E}_{PK}\boldsymbol{C}_P + (\boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{D}_P\boldsymbol{E}_{KP}\boldsymbol{C}_K)\boldsymbol{R}\end{aligned} \tag{4.12}$$

has a solution $\boldsymbol{R} = \boldsymbol{R}_{KP}$ then the left parameter $\boldsymbol{Q} = \boldsymbol{Q}_L$ can be calculated from the controller $\boldsymbol{K}$ by

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix} = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{L}_K \\ \boldsymbol{0} & \boldsymbol{E}_{KP}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K\end{bmatrix} \\ &\boldsymbol{L}_K = (\boldsymbol{B}_K\boldsymbol{D}_P - \boldsymbol{R}\boldsymbol{B}_P)\boldsymbol{E}_{KP} \\ &\boldsymbol{E}_{KP} = (\boldsymbol{1} - \boldsymbol{D}_K\boldsymbol{D}_P)^{-1} \\ &\boldsymbol{E}_{PK} = (\boldsymbol{1} - \boldsymbol{D}_P\boldsymbol{D}_K)^{-1}\end{aligned} \tag{4.13}$$

If the Sylvester equation

$$\boldsymbol{R}\boldsymbol{A}_P = \boldsymbol{A}_Q\boldsymbol{R} + \boldsymbol{B}_Q\boldsymbol{C}_P \tag{4.14}$$

has a solution $\boldsymbol{R} = \boldsymbol{R}_{KP}$ then the controller $\boldsymbol{K}$ can be calculated from the left parameter $\boldsymbol{Q} = \boldsymbol{Q}_L$ by

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K\end{bmatrix} = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{L}_K \\ \boldsymbol{0} & \boldsymbol{E}_{KP}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix} \\ &\boldsymbol{L}_K = \boldsymbol{B}_Q\boldsymbol{D}_P - \boldsymbol{R}\boldsymbol{B}_P \\ &\boldsymbol{E}_{KP} = \boldsymbol{1} + \boldsymbol{D}_Q\boldsymbol{D}_P\end{aligned} \tag{4.15}$$

Proof This is just a summary of (2.21), (2.23), (2.40) and (2.19) with $\boldsymbol{R} = \boldsymbol{R}_{KP}$.

Remark Note that the quadratic Riccati equation (4.12) becomes a linear Sylvester equation (4.14) if it is expressed in terms of the parametrization. The Riccati equation has to be solved at the start of the optimization when an initial $\boldsymbol{Q}_0$ has to be calculated from the initial $\boldsymbol{K}_0$. The Sylvester equation has to be solved at the end of the optimization when a final $\boldsymbol{K}^*$ has to be calculated from the final $\boldsymbol{Q}^*$.

4.2.2 Parametrized Extended Loops

Proposition A state space representation of the extended loop $\boldsymbol{H}_i$ in terms of parameter $\boldsymbol{Q} = \boldsymbol{Q}_L$ is given by

$$\boldsymbol{H}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right] = \left[\begin{array}{cccc|ccc} \boldsymbol{A}_{zw_i} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \multicolumn{3}{c}{\multirow{4}{*}{$\cdots$}} \\ \boldsymbol{0} & \boldsymbol{A}_{yw_i} & \boldsymbol{0} & \boldsymbol{0} & & & \\ \boldsymbol{0} & \boldsymbol{B}_{zu_i}\boldsymbol{D}_Q\boldsymbol{C}_{yw_i} & \boldsymbol{A}_{zu_i} & \boldsymbol{B}_{zu_i}\boldsymbol{D}_Q\boldsymbol{C}_P & & & \\ \boldsymbol{0} & \boldsymbol{B}_P\boldsymbol{D}_Q\boldsymbol{C}_{yw_i} & \boldsymbol{0} & \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_Q\boldsymbol{C}_P & & & \\ \boldsymbol{0} & (\boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{C}_{yw_i} & \boldsymbol{0} & (\boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{C}_P & \multicolumn{3}{c}{\multirow{2}{*}{$\cdots$}} \\ \hline \boldsymbol{C}_{zw_i} & \boldsymbol{D}_{zu_i}\boldsymbol{D}_Q\boldsymbol{C}_{yw_i} & \boldsymbol{C}_{zu_i} & \boldsymbol{D}_{zu_i}\boldsymbol{D}_Q\boldsymbol{C}_P & & & \end{array}\right.$$

$$\left.\begin{array}{ccc|c} \multicolumn{3}{c|}{\multirow{4}{*}{$\cdots$}} & \boldsymbol{B}_{zw_i} \\ & & & \boldsymbol{B}_{yw_i} \\ & & & \boldsymbol{B}_{zu_i}\boldsymbol{D}_Q\boldsymbol{D}_{yw_i} \\ & & & \boldsymbol{B}_P\boldsymbol{D}_Q\boldsymbol{D}_{yw_i} \\ \multicolumn{3}{c|}{\multirow{2}{*}{$\cdots$}} & (\boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{D}_{yw_i} \\ \hline & & & \boldsymbol{D}_{zw_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_Q\boldsymbol{D}_{yw_i} \end{array}\right] \tag{4.16}$$

In the special case (4.3) this state space representation reduces to

$$\boldsymbol{H}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right] = \left[\begin{array}{cc|c} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{D}_Q\boldsymbol{C}_P & \multirow{2}{*}{$\cdots$} & \\ (\boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{C}_P & & \\ \hline \boldsymbol{C}_{z_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_Q\boldsymbol{C}_P & & \end{array}\right.$$

$$\left.\begin{array}{cc|c} \boldsymbol{B}_P\boldsymbol{C}_Q & \multirow{2}{*}{$\cdots$} & \boldsymbol{B}_{w_i} + \boldsymbol{B}_P\boldsymbol{D}_Q\boldsymbol{D}_{yw_i} \\ \boldsymbol{A}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{C}_Q & & (\boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{D}_{yw_i} \\ \hline \boldsymbol{D}_{zu_i}\boldsymbol{C}_Q & & \boldsymbol{D}_{zw_i} + \boldsymbol{D}_{zu_i}\boldsymbol{D}_Q\boldsymbol{D}_{yw_i} \end{array}\right] \tag{4.17}$$

Proof The parametrized state space representations (4.16) and (4.17) respectively follow from (4.2) and (4.4) by replacing all terms containing state matrices of the controller $\boldsymbol{K}$ according to (2.43) with $\boldsymbol{R} = \boldsymbol{R}_{KP}$.

4.2.3 Parametrized Optimization Problem

Proposition If condition (4.12) or (4.14) have a solution $\boldsymbol{R}$ at all times during the optimization process, then the multiobjective optimization problem (4.11) is equivalent to the formulation in terms of the parameter $\boldsymbol{Q}$ which can be given in the general form

$$\begin{aligned} \boldsymbol{Q}^* = \arg \quad \inf_{\substack{\boldsymbol{\alpha} \geq 0 \\ \boldsymbol{\beta} \geq 0 \\ \boldsymbol{\gamma} > 0 \\ \boldsymbol{\epsilon} > 0 \\ \boldsymbol{Z} > 0 \\ \boldsymbol{Q}}} & f(\boldsymbol{\gamma}) \text{ subject to } \boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{Q}) \leq 0 \end{aligned} \tag{4.18}$$

Proof Problem (4.11) can be formulated in terms of $\boldsymbol{Q}$ by formulating the extended loops $\boldsymbol{H}_i$ as given in (4.2) in terms of $\boldsymbol{Q}$ as given in (4.16). This leads to (4.18). If the conditions (4.12) or (4.14) have a solution $\boldsymbol{R}$ at all times, then there is always a reversible relation between the controller $\boldsymbol{K}$ and the parameter $\boldsymbol{Q}$ expressed by (4.13) or (4.15) so that the problems (4.11) and (4.18) become equivalent.

4.2.4 Evaluation Step

However, despite being formulated in terms of the parameter $\boldsymbol{Q}$, problem (4.18) can not be solved directly with the currently known convex optimization techniques since these require the constraints formulated as linear matrix inequalities in all variables which is not the case as can be seen by (4.6), (4.8), (4.10). To circumvent this problem we propose an optimization consisting of two alternating steps, an evaluation and an actualization. The evaluation step calculates the local sensitivity matrix which is the optimal Lagrange multiplier associated to the constraint function for constant parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$.

Proposition For a given constant controller parameter value $\boldsymbol{Q}$ and given constant L1 parameter values $\boldsymbol{\alpha}, \boldsymbol{\beta}$, the problem (4.18) reduces to an optimization problem for given constant extended loops $\boldsymbol{H}_i$. This means a search for the best upper norm bounds γ^* for a given system and may therefore be rather called a norm evaluation problem. If f is a cost function of the form $f(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^T \boldsymbol{J} \boldsymbol{\gamma}$ with diagonal $\boldsymbol{J} > 0$ and if all objectives are formulated in terms of the L1, L2/H2, H∞ norms this evaluation problem is convex with linear matrix inequality constraints. If this convex optimization problem exhibits strong duality then the local sensitivity matrix $\boldsymbol{\Omega}^*$ for the given parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ can be found and there is a relation in the form of

$$f(\boldsymbol{\gamma}^*) = g(\boldsymbol{\Omega}^*) = \inf_{\substack{\boldsymbol{\gamma} > 0 \\ \boldsymbol{\epsilon} > 0 \\ \boldsymbol{Z} > 0}} f(\boldsymbol{\gamma}) \text{ subject to } \boldsymbol{G}_{Q,\alpha,\beta}(\boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}) \leq 0 \tag{4.19}$$

The function g is the Lagrange dual function to the cost function f as defined in (B.20). The sensitivity matrix $\boldsymbol{\Omega}$ is the Lagrange dual to the variable $\boldsymbol{\gamma}$ and is known as the Lagrange multiplier. The sensitivity matrix is hermitian. The evaluation constraint function $\boldsymbol{G}_{Q,\alpha,\beta}$ equals the function $\boldsymbol{G}$ from (4.18) but for $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ kept constant.

Proof For fixed parameters $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and thus fixed $\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{C}_i, \boldsymbol{D}_i$ the bilinear matrix inequalities in the L1, L2/H2, H∞ norm objectives given in section 4.1.2 reduce to linear matrix inequalities in the variables $\boldsymbol{\gamma}_i^2, \boldsymbol{Z}_i$. If the cost function takes the form $f(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^T \boldsymbol{J} \boldsymbol{\gamma}$ with diagonal $\boldsymbol{J} > 0$ then $f(\boldsymbol{\gamma}) = \sum_i [\boldsymbol{J}]_{ii} \gamma_i^2$ becomes linear in γ_i^2. Under these circumstances the optimization problem (4.18) consists of a linear cost function f in the variables γ_i^2 with linear matrix inequality constraints in the variables $\gamma_i^2, \boldsymbol{Z}_i$ and therefore becomes a convex optimization problem. Most often, convex optimization problems exhibit strong duality, see [5] for extended information. The equality in (4.19) then follows if the optimization problem exhibits strong duality (B.22).

4.2.5 Actualization Step

The actualization step then calculates the local subgradient from the sensitivity matrix so that the parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ can be changed in order to improve the cost function.

Proposition If the optimization problem (4.19) exhibits strong duality and the local sensitivity matrix $\boldsymbol{\Omega}^*$ is known then the local subgradient of the cost function f can be calculated. This information can then be used to find the direction in which the parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ have to be actualized so that the cost function f decreases. The subgradient can be given by

$$\mathrm{d}f \leq \boldsymbol{\Omega}^* \bullet \delta \boldsymbol{G} \tag{4.20}$$

The scalar product $\bullet$ of two matrices is as defined in (A.68).

Proof This is a result of the perturbation analysis of optimization problems that can be found in [4]. It is the generalization of a similar result in linear programming applied to a semidefinite programming. A simplified proof is given in section B.3. The key to understand these proofs is to note that the optimization problem (4.19) can be seen as an optimization problem with $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ as optimization parameters such as in (B.23). Then if the problem is convex and exhibits strong duality, the effects of a parameter perturbation on this optimization problem can be studied, which leads to (B.24) with Lagrange function defined as (B.25). In this context, the Lagrange function is given by

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{\Omega}, \boldsymbol{Q}) = f(\boldsymbol{\gamma}) + \boldsymbol{\Omega} \bullet \boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{Q})$$

with $\boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}$ forming the optimization variables, $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ the optimization parameters and $\boldsymbol{\Omega}$ the Lagrange multiplier. Then due to the linearity of the scalar product, the effect of parameter changes on the cost function takes the variational form

$$\mathrm{d}f \leq \delta L = \boldsymbol{\Omega}^* \bullet \delta \boldsymbol{G}$$

for a variation of the parameters $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, which verifies (4.20).

Remark The total variation $\mathrm{d}f$ of the cost function is defined as

$$\mathrm{d}f = f(\boldsymbol{\gamma}_+^*) - f(\boldsymbol{\gamma}^*)$$

and the partial variation $\delta \boldsymbol{G}$ of the constraint function is defined as

$$\delta \boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\alpha}_+, \boldsymbol{\beta}_+, \boldsymbol{\gamma}^*, \boldsymbol{\epsilon}^*, \boldsymbol{Z}^*, \boldsymbol{Q}_+) - \boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}^*, \boldsymbol{\epsilon}^*, \boldsymbol{Z}^*, \boldsymbol{Q})$$

The values $\boldsymbol{\gamma}^*, \boldsymbol{\epsilon}^*, \boldsymbol{Z}^*$ are the optimal values of the decision variables obtained after the last evaluation step (4.19) for the current parameter values $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. The value $\boldsymbol{\gamma}_+^*$ is an optimal value which would be obtained after the next evaluation step if the parameters were actualized to $\boldsymbol{Q}_+, \boldsymbol{\alpha}_+, \boldsymbol{\beta}_+$.

4.2.6 Variation of the Constraint Function

In order to obtain the subgradient in (4.20) the variation $\delta \boldsymbol{G}$ of the constraint function or at least the directional variation $\boldsymbol{\Omega}^* \bullet \delta \boldsymbol{G}$ has to be calculated.

Proposition If all objectives are formulated in terms of the L1, L2/H2, H∞ norms then equation (4.20) can be written in the form

$$\mathrm{d}f \leq \mathrm{Re} \sum_i (\boldsymbol{\Omega}_{A_i} \bullet \delta \boldsymbol{A}_i + \boldsymbol{\Omega}_{B_i} \bullet \delta \boldsymbol{B}_i + \boldsymbol{\Omega}_{C_i} \bullet \delta \boldsymbol{C}_i + \boldsymbol{\Omega}_{D_i} \bullet \delta \boldsymbol{D}_i) + \boldsymbol{\Omega}_E \bullet \delta \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \quad (4.21)$$

with $(\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{C}_i, \boldsymbol{D}_i)$ being a realization of the extended loop $\boldsymbol{H}_i$ and the sensibility matrices $\boldsymbol{\Omega}_{A_i}, \boldsymbol{\Omega}_{B_i}, \boldsymbol{\Omega}_{C_i}, \boldsymbol{\Omega}_{D_i}, \Omega_E$ depending on the actual constraint function $\boldsymbol{G}$.

Proof Variables ϵ are used to formally transform definite constraints of the form $\boldsymbol{M} < 0$ into semidefinite constraints by $\boldsymbol{M} + \epsilon \boldsymbol{1} \leq 0$ with $\epsilon > 0$ and $\delta\epsilon = 0$ since ϵ does not depend on the optimization parameters $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Hence the directional variation of the definite constraint $\boldsymbol{M} < 0$ in semidefinite form is given by $\boldsymbol{\Omega} \bullet \delta(\boldsymbol{M} + \epsilon \boldsymbol{1}) = \boldsymbol{\Omega} \bullet \delta \boldsymbol{M}$. Then using (A.69), (A.70), (A.71), (A.72) and (A.81) the first order variations of the constraint function given in unified semidefinite form $\boldsymbol{G} \leq 0$ formulated in terms of the L1, L2/H2, H∞ norm constraints can be calculated.

The first order directional variations of the L1 norm constraint (4.6) can then be given by

$$\begin{aligned}
\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{L1_i} = \boldsymbol{\Omega}_1 \bullet \delta \begin{bmatrix} \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + (\alpha_i - 1) \boldsymbol{Z}_i & \boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i \\ \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i & \boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i - \alpha_i \boldsymbol{1} \end{bmatrix} + \\
+ \boldsymbol{\Omega}_2 \bullet \delta \begin{bmatrix} \boldsymbol{C}_i^H \boldsymbol{C}_i - \beta_i \boldsymbol{Z}_i & \boldsymbol{C}_i^H \boldsymbol{D}_i \\ \boldsymbol{D}_i^H \boldsymbol{C}_i & \boldsymbol{D}_i^H \boldsymbol{D}_i + (\beta_i - \gamma_i^2) \boldsymbol{1} \end{bmatrix} = \\
= 2 \,\mathrm{Re}(\boldsymbol{\Omega}_1 \bullet (\begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{B}_i \end{bmatrix}^H \boldsymbol{Z}_i \begin{bmatrix} \delta \boldsymbol{A}_i & \delta \boldsymbol{B}_i \end{bmatrix})) + \\
+ 2 \,\mathrm{Re}(\boldsymbol{\Omega}_2 \bullet (\begin{bmatrix} \boldsymbol{C}_i & \boldsymbol{D}_i \end{bmatrix}^H \begin{bmatrix} \delta \boldsymbol{C}_i & \delta \boldsymbol{D}_i \end{bmatrix})) + \\
+ \boldsymbol{\Omega}_1 \bullet \begin{bmatrix} \boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{1} \end{bmatrix} \delta \alpha_i - \boldsymbol{\Omega}_2 \bullet \begin{bmatrix} \boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{1} \end{bmatrix} \delta \beta_i
\end{aligned}$$

which can be written in the form

$$\begin{aligned}
&\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{L1_i} = \\
&= \mathrm{Re}((2 \boldsymbol{Z}_i (\boldsymbol{A}_i \boldsymbol{\Omega}_{1_{11}} + \boldsymbol{B}_i \boldsymbol{\Omega}_{1_{21}})) \bullet \delta \boldsymbol{A}_i + (2 \boldsymbol{Z}_i (\boldsymbol{A}_i \boldsymbol{\Omega}_{1_{12}} + \boldsymbol{B}_i \boldsymbol{\Omega}_{1_{22}})) \bullet \delta \boldsymbol{B}_i + \\
&\quad + 2 (\boldsymbol{C}_i \boldsymbol{\Omega}_{2_{11}} + \boldsymbol{D}_i \boldsymbol{\Omega}_{2_{21}}) \bullet \delta \boldsymbol{C}_i + 2 (\boldsymbol{C}_i \boldsymbol{\Omega}_{2_{12}} + \boldsymbol{D}_i \boldsymbol{\Omega}_{2_{22}}) \bullet \delta \boldsymbol{D}_i) + \\
&\quad + \begin{bmatrix} \begin{bmatrix} \boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{1} \end{bmatrix} \bullet \boldsymbol{\Omega}_1 \\ \begin{bmatrix} -\boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \bullet \boldsymbol{\Omega}_2 \end{bmatrix} \bullet \begin{bmatrix} \delta \alpha_i \\ \delta \beta_i \end{bmatrix} \quad (4.22)
\end{aligned}$$

The first order directional variations of the L2/H2 norm constraint (4.8) can be given by

$$\begin{aligned}
\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{L2/H2_i} = \boldsymbol{\Omega}_3 \bullet \delta (\boldsymbol{A}_i^H \boldsymbol{Z}_i \boldsymbol{A}_i + \boldsymbol{C}_i^H \boldsymbol{C}_i - \boldsymbol{Z}_i) + \\
+ \Omega_4 \bullet \delta (\mathrm{tr}(\boldsymbol{B}_i^H \boldsymbol{Z}_i \boldsymbol{B}_i + \boldsymbol{D}_i^H \boldsymbol{D}_i) - \gamma_i^2) = \\
= 2 \,\mathrm{Re}(\boldsymbol{\Omega}_3 \bullet (\boldsymbol{A}_i^H \boldsymbol{Z}_i \delta \boldsymbol{A}_i + \boldsymbol{C}_i^H \delta \boldsymbol{C}_i)) + \\
+ 2 \,\mathrm{Re}((\Omega_4 \boldsymbol{1}) \bullet (\boldsymbol{B}_i^H \boldsymbol{Z}_i \delta \boldsymbol{B}_i + \boldsymbol{D}_i^H \delta \boldsymbol{D}_i))
\end{aligned}$$

which can be written in the form

$$\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{L2/H2_i} = \mathrm{Re}((2\boldsymbol{Z}_i\boldsymbol{A}_i\boldsymbol{\Omega}_3) \bullet \delta\boldsymbol{A}_i + (2\boldsymbol{Z}_i\boldsymbol{B}_i\Omega_4) \bullet \delta\boldsymbol{B}_i + \\ + (2\boldsymbol{C}_i\boldsymbol{\Omega}_3) \bullet \delta\boldsymbol{C}_i + (2\boldsymbol{D}_i\Omega_4) \bullet \delta\boldsymbol{D}_i) \quad (4.23)$$

The first order directional variations of the H∞ norm constraint (4.10) can be given by

$$\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{H\infty_i} = \boldsymbol{\Omega}_5 \bullet \delta \begin{bmatrix} \boldsymbol{A}_i^H\boldsymbol{Z}_i\boldsymbol{A}_i + \boldsymbol{C}_i^H\boldsymbol{C}_i - \boldsymbol{Z}_i & \boldsymbol{A}_i^H\boldsymbol{Z}_i\boldsymbol{B}_i + \boldsymbol{C}_i^H\boldsymbol{D}_i \\ \boldsymbol{B}_i^H\boldsymbol{Z}_i\boldsymbol{A}_i + \boldsymbol{D}_i^H\boldsymbol{C}_i & \boldsymbol{B}_i^H\boldsymbol{Z}_i\boldsymbol{B}_i + \boldsymbol{D}_i^H\boldsymbol{D}_i - \gamma_i^2\mathbf{1} \end{bmatrix} = \\ = 2\,\mathrm{Re}(\boldsymbol{\Omega}_5 \bullet (\begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{B}_i \end{bmatrix}^H \boldsymbol{Z}_i \begin{bmatrix} \delta\boldsymbol{A}_i & \delta\boldsymbol{B}_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{C}_i & \boldsymbol{D}_i \end{bmatrix}^H \begin{bmatrix} \delta\boldsymbol{C}_i & \delta\boldsymbol{D}_i \end{bmatrix})$$

which can be written in the form

$$\boldsymbol{\Omega} \bullet \delta \boldsymbol{G}_{H\infty_i} = \\ = \mathrm{Re}((2\boldsymbol{Z}_i(\boldsymbol{A}_i\boldsymbol{\Omega}_{5_{11}} + \boldsymbol{B}_i\boldsymbol{\Omega}_{5_{21}})) \bullet \delta\boldsymbol{A}_i + (2\boldsymbol{Z}_i(\boldsymbol{A}_i\boldsymbol{\Omega}_{5_{12}} + \boldsymbol{B}_i\boldsymbol{\Omega}_{5_{22}})) \bullet \delta\boldsymbol{B}_i + \\ + 2(\boldsymbol{C}_i\boldsymbol{\Omega}_{5_{11}} + \boldsymbol{D}_i\boldsymbol{\Omega}_{5_{21}}) \bullet \delta\boldsymbol{C}_i + 2(\boldsymbol{C}_i\boldsymbol{\Omega}_{5_{12}} + \boldsymbol{D}_i\boldsymbol{\Omega}_{5_{22}}) \bullet \delta\boldsymbol{D}_i) \quad (4.24)$$

Due to the linearity of the scalar product, the directional variations of all L1, L2/H2, H∞ norm constraints given in (4.22), (4.23), (4.24) and thus the directional variation of any constraint function $\boldsymbol{G}$ that is formulated in terms of the L1, L2/H2, H∞ norms can be written in the form (4.21).

4.2.7 Variation of the Extended Loops

In order to obtain the subgradient in (4.21) the directional variations $\boldsymbol{\Omega}_{A_i} \bullet \delta\boldsymbol{A}_i$, $\boldsymbol{\Omega}_{B_i} \bullet \delta\boldsymbol{B}_i$, $\boldsymbol{\Omega}_{C_i} \bullet \delta\boldsymbol{C}_i$, $\boldsymbol{\Omega}_{D_i} \bullet \delta\boldsymbol{D}_i$ of the extended loops have to be calculated.

Proposition If the Sylvester equations

$$\boldsymbol{F}_i\boldsymbol{A}_P^H - \boldsymbol{A}_Q^H\boldsymbol{F}_i = \boldsymbol{E}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{R_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H\boldsymbol{D}_Q^H\boldsymbol{B}_P^H) \quad (4.25)$$

have solutions $\boldsymbol{F}_i$ then equation (4.21) can be written in the form

$$\mathrm{d}f \le \mathrm{Re}(\boldsymbol{\Omega}_{A_Q} \bullet \delta\boldsymbol{A}_Q + \boldsymbol{\Omega}_{B_Q} \bullet \delta\boldsymbol{B}_Q + \boldsymbol{\Omega}_{C_Q} \bullet \delta\boldsymbol{C}_Q + \boldsymbol{\Omega}_{D_Q} \bullet \delta\boldsymbol{D}_Q) + \boldsymbol{\Omega}_E \bullet \delta \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \quad (4.26)$$

with

$$\begin{aligned} \boldsymbol{\Omega}_{A_Q} &= \sum_i (\boldsymbol{E}^H\boldsymbol{\Omega}_{A_i}\boldsymbol{E} + \boldsymbol{F}_i\boldsymbol{R}^H) \\ \boldsymbol{\Omega}_{B_Q} &= \sum_i (\boldsymbol{E}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H) + \boldsymbol{F}_i\boldsymbol{C}_P^H) \\ \boldsymbol{\Omega}_{C_Q} &= \sum_i (\boldsymbol{E}_{C_i}^H\boldsymbol{\Omega}_{A_i} + \boldsymbol{D}_{zu_i}^H\boldsymbol{\Omega}_{C_i})\boldsymbol{E} \\ \boldsymbol{\Omega}_{D_Q} &= \sum_i (\boldsymbol{E}_{C_i}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H) + \boldsymbol{D}_{zu_i}^H(\boldsymbol{\Omega}_{C_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{D_i}\boldsymbol{D}_{yw_i}^H)) \end{aligned} \quad (4.27)$$

and

$$\boldsymbol{E} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \boldsymbol{E}_{B_i} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{C}_{yw_i}^H \\ \mathbf{0} \\ \boldsymbol{C}_P^H \\ \mathbf{0} \end{bmatrix}, \boldsymbol{E}_{C_i} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{B}_{zu_i} \\ \boldsymbol{B}_P \\ \boldsymbol{R}\boldsymbol{B}_P \end{bmatrix}, \boldsymbol{E}_{R_i} = (\boldsymbol{E}\boldsymbol{C}_Q^H + \boldsymbol{E}_{B_i}\boldsymbol{D}_Q^H)\boldsymbol{B}_P^H \tag{4.28}$$

which in the special case (4.3) reduce to $\boldsymbol{E}_{B_i} = \boldsymbol{E}_B, \boldsymbol{E}_{C_i} = \boldsymbol{E}_C, \boldsymbol{E}_{R_i} = \boldsymbol{E}_R$ with

$$\boldsymbol{E} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \boldsymbol{E}_B = \begin{bmatrix} \boldsymbol{C}_P^H \\ \mathbf{0} \end{bmatrix}, \boldsymbol{E}_C = \begin{bmatrix} \boldsymbol{B}_P \\ \boldsymbol{R}\boldsymbol{B}_P \end{bmatrix}, \boldsymbol{E}_R = (\boldsymbol{E}\boldsymbol{C}_Q^H + \boldsymbol{E}_B\boldsymbol{D}_Q^H)\boldsymbol{B}_P^H \tag{4.29}$$

Proof The first order variations of the extended loops can be calculated from (4.16) and can then be given by

$$\begin{aligned}
\delta\boldsymbol{A}_i &= \boldsymbol{E}\delta\boldsymbol{A}_Q\boldsymbol{E}^H + \boldsymbol{E}\delta\boldsymbol{B}_Q\boldsymbol{E}_{B_i}^H + \boldsymbol{E}_{C_i}\delta\boldsymbol{C}_Q\boldsymbol{E}^H + \boldsymbol{E}_{C_i}\delta\boldsymbol{D}_Q\boldsymbol{E}_{B_i}^H + \boldsymbol{E}\delta\boldsymbol{R}\boldsymbol{E}_{R_i}^H \\
\delta\boldsymbol{B}_i &= (\boldsymbol{E}\delta\boldsymbol{B}_Q + \boldsymbol{E}_{C_i}\delta\boldsymbol{D}_Q + \boldsymbol{E}\delta\boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q)\boldsymbol{D}_{yw_i} \\
\delta\boldsymbol{C}_i &= \boldsymbol{D}_{zu_i}(\delta\boldsymbol{C}_Q\boldsymbol{E}^H + \delta\boldsymbol{D}_Q\boldsymbol{E}_{B_i}^H) \\
\delta\boldsymbol{D}_i &= \boldsymbol{D}_{zu_i}\delta\boldsymbol{D}_Q\boldsymbol{D}_{yw_i}
\end{aligned}$$

using (A.81) with $\boldsymbol{E}, \boldsymbol{E}_{B_i}, \boldsymbol{E}_{C_i}, \boldsymbol{E}_{R_i}$ as defined in (4.28). In the special case (4.3) the first order variations of the extended loops can be calculated from (4.17) in a similar way and are given by the same expressions but with $\boldsymbol{E}, \boldsymbol{E}_{B_i}, \boldsymbol{E}_{C_i}, \boldsymbol{E}_{R_i}$ as defined in (4.29). Then using (A.70) equation (4.21) can be written as

$$\begin{aligned}
\mathrm{d}f \leq \mathrm{Re}\sum_i ((\boldsymbol{E}^H\boldsymbol{\Omega}_{A_i}\boldsymbol{E}) \bullet \delta\boldsymbol{A}_Q + (\boldsymbol{E}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H)) \bullet \delta\boldsymbol{B}_Q + \\
+ ((\boldsymbol{E}_{C_i}^H\boldsymbol{\Omega}_{A_i} + \boldsymbol{D}_{zu_i}^H\boldsymbol{\Omega}_{C_i})\boldsymbol{E}) \bullet \delta\boldsymbol{C}_Q + \\
+ (\boldsymbol{E}_{C_i}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H) + \boldsymbol{D}_{zu_i}^H(\boldsymbol{\Omega}_{C_i}\boldsymbol{E}_{B_i} + \boldsymbol{\Omega}_{D_i}\boldsymbol{D}_{yw_i}^H)) \bullet \delta\boldsymbol{D}_Q + \\
+ (\boldsymbol{E}^H(\boldsymbol{\Omega}_{A_i}\boldsymbol{E}_{R_i} + \boldsymbol{\Omega}_{B_i}\boldsymbol{D}_{yw_i}^H\boldsymbol{D}_Q^H\boldsymbol{B}_P^H)) \bullet \delta\boldsymbol{R}) + \boldsymbol{\Omega}_E \bullet \delta\begin{bmatrix}\boldsymbol{\alpha} \\ \boldsymbol{\beta}\end{bmatrix}
\end{aligned} \tag{4.30}$$

The variable $\boldsymbol{R}$ is constrained by the Sylvester equation (4.14) whose first order variation is given by

$$\delta\boldsymbol{R}\boldsymbol{A}_P = \delta\boldsymbol{A}_Q\boldsymbol{R} + \boldsymbol{A}_Q\delta\boldsymbol{R} + \delta\boldsymbol{B}_Q\boldsymbol{C}_P$$

so a directional variation for an arbitrary $\boldsymbol{F}_i$ is given by

$$\boldsymbol{F}_i \bullet (\delta\boldsymbol{R}\boldsymbol{A}_P - \boldsymbol{A}_Q\delta\boldsymbol{R}) = \boldsymbol{F}_i \bullet (\delta\boldsymbol{A}_Q\boldsymbol{R} + \delta\boldsymbol{B}_Q\boldsymbol{C}_P)$$

which using (A.70) can be written in the form

$$(\boldsymbol{F}_i\boldsymbol{A}_P^H - \boldsymbol{A}_Q^H\boldsymbol{F}_i) \bullet \delta\boldsymbol{R} = (\boldsymbol{F}_i\boldsymbol{R}^H) \bullet \delta\boldsymbol{A}_Q + (\boldsymbol{F}_i\boldsymbol{C}_P^H) \bullet \delta\boldsymbol{B}_Q$$

Then if the $\boldsymbol{F}_i$ solve the Sylvester equations (4.25) equation (4.30) can be written in the form (4.26) with $\boldsymbol{\Omega}_{A_Q}, \boldsymbol{\Omega}_{B_Q}, \boldsymbol{\Omega}_{C_Q}, \boldsymbol{\Omega}_{D_Q}$ given by (4.27).

4.2.8 Variation of the Parameter

In order to obtain the subgradient in (4.26) the directional variations $\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q$, $\boldsymbol{\Omega}_{B_Q} \bullet \delta \boldsymbol{B}_Q$, $\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q$, $\boldsymbol{\Omega}_{D_Q} \bullet \delta \boldsymbol{D}_Q$ of the parameter have to be calculated. This is done by applying the parametrization of stable observable pairs in (3.2.4) based on lossless systems. Thereby an infinitesimal variation $(\delta \boldsymbol{A}_Q, \delta \boldsymbol{C}_Q)$ can be calculated as a function of the variation $\delta \boldsymbol{W}$ of the stable observable pair parameter and thus an infinitesimal variation $\delta \boldsymbol{Q} = (\delta \boldsymbol{A}_Q, \delta \boldsymbol{B}_Q, \delta \boldsymbol{C}_Q, \delta \boldsymbol{D}_Q)$ of the parameter $\boldsymbol{Q}$ and hence the directional variations required in (4.26) can be calculated as a function of $\delta \boldsymbol{W}$ and the variations $\delta \boldsymbol{B}_Q, \delta \boldsymbol{D}_Q$.

Proposition If the pair $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$ defining the center of the current chart is stable and observable then it has a lossless completion $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ and the Stein and Sylvester equations

$$\begin{aligned}
\boldsymbol{A}_{Q_0} \boldsymbol{L}_A \boldsymbol{A}_{Q_0}^H - \boldsymbol{L}_A &= \boldsymbol{T} \boldsymbol{A}_{Q_0}^H \boldsymbol{\Omega}_{A_Q} - \boldsymbol{A}_{Q_0} \boldsymbol{T} \boldsymbol{A}_{Q_0}^H \boldsymbol{\Omega}_{A_Q} \boldsymbol{G} \boldsymbol{T} \boldsymbol{A}_{Q_0}^H - \\
&\quad - \boldsymbol{G}^{-1} \boldsymbol{\Omega}_{A_Q} \boldsymbol{W}^H \boldsymbol{S}^H \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} + \\
&\quad + \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \boldsymbol{M}_A \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \boldsymbol{W}^H \boldsymbol{N}_A \boldsymbol{W} \boldsymbol{G}^{-1} \\
\boldsymbol{A}_{Q_0} \boldsymbol{L}_C \boldsymbol{A}_{Q_0}^H - \boldsymbol{L}_C &= \boldsymbol{T} \boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} - \boldsymbol{A}_{Q_0} \boldsymbol{T} \boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{G} \boldsymbol{T} \boldsymbol{A}_{Q_0}^H + \\
&\quad + \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \boldsymbol{M}_C \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \boldsymbol{W}^H \boldsymbol{N}_C \boldsymbol{W} \boldsymbol{G}^{-1} \\
\boldsymbol{U}^{\frac{1}{2}} \boldsymbol{M}_A + \boldsymbol{M}_A \boldsymbol{U}^{\frac{1}{2}} &= \boldsymbol{V}^{\frac{1}{2}} \boldsymbol{S}^H \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{\Omega}_{A_Q} \boldsymbol{W}^H \boldsymbol{S}^H \\
\boldsymbol{U}^{\frac{1}{2}} \boldsymbol{M}_C + \boldsymbol{M}_C \boldsymbol{U}^{\frac{1}{2}} &= \boldsymbol{V}^{\frac{1}{2}} \boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H \\
\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{N}_A + \boldsymbol{N}_A \boldsymbol{V}^{\frac{1}{2}} &= \boldsymbol{S}^H \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{\Omega}_{A_Q} \boldsymbol{W}^H \boldsymbol{S}^H \boldsymbol{U}^{\frac{1}{2}} \\
\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{N}_C + \boldsymbol{N}_C \boldsymbol{V}^{\frac{1}{2}} &= \boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H \boldsymbol{U}^{\frac{1}{2}}
\end{aligned} \tag{4.31}$$

have unique solutions $\boldsymbol{L}_A, \boldsymbol{L}_C, \boldsymbol{M}_A, \boldsymbol{M}_C, \boldsymbol{N}_A, \boldsymbol{N}_C$ respectively and equation (4.26) can be written in the form

$$\mathrm{d}f \leq \mathrm{Re}(\boldsymbol{\Omega}_W \bullet \delta \boldsymbol{W} + \boldsymbol{\Omega}_{B_Q} \bullet \delta \boldsymbol{B}_Q + \boldsymbol{\Omega}_{D_Q} \bullet \delta \boldsymbol{D}_Q) + \boldsymbol{\Omega}_E \bullet \delta \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \tag{4.32}$$

with

$$\begin{aligned}
\boldsymbol{\Omega}_W = \boldsymbol{S}^H (\boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{\Omega}_{A_Q} + \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q}) + \boldsymbol{W} (\boldsymbol{L}_A + \boldsymbol{L}_C + \boldsymbol{L}_A^H + \boldsymbol{L}_C^H) - \\
- (\boldsymbol{N}_A + \boldsymbol{N}_C + \boldsymbol{N}_A^H + \boldsymbol{N}_C^H) \boldsymbol{W} \boldsymbol{G}^{-1}
\end{aligned} \tag{4.33}$$

and

$$\begin{aligned}
\boldsymbol{S} &= \boldsymbol{U}^{-\frac{1}{2}} \boldsymbol{V}^{-\frac{1}{2}} \\
\boldsymbol{T} &= (\boldsymbol{G} + \boldsymbol{W}^H \boldsymbol{W})^{-1} \\
\boldsymbol{U} &= \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 + \boldsymbol{D}_0^H \boldsymbol{D}_0 \\
\boldsymbol{V} &= \boldsymbol{1} + \boldsymbol{W} \boldsymbol{G}^{-1} \boldsymbol{W}^H
\end{aligned} \tag{4.34}$$

The Gram observability matrices $\boldsymbol{G}_0$ of the chart center $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$ and $\boldsymbol{G}$ of the current pair $(\boldsymbol{A}_Q, \boldsymbol{C}_Q)$ are uniquely determined by the Stein equations

$$\begin{aligned}
\boldsymbol{A}_{Q_0}^H \boldsymbol{G}_0 \boldsymbol{A}_{Q_0} - \boldsymbol{G}_0 + \boldsymbol{C}_{Q_0}^H \boldsymbol{C}_{Q_0} &= \boldsymbol{0} \\
\boldsymbol{A}_{Q_0}^H \boldsymbol{G} \boldsymbol{A}_{Q_0} - \boldsymbol{G} + \boldsymbol{C}_{Q_0}^H \boldsymbol{C}_{Q_0} &= \boldsymbol{W}^H \boldsymbol{W}
\end{aligned} \tag{4.35}$$

respectively and $\boldsymbol{G} \leq \boldsymbol{G}_0$ must obey the condition

$$\boldsymbol{G} > 0 \tag{4.36}$$

otherwise the current chart is no longer valid and a new one has to be selected by centering on the balanced current pair $(\boldsymbol{A}_{Q_0} = \boldsymbol{J}\boldsymbol{A}_Q\boldsymbol{J}^{-1}, \boldsymbol{C}_{Q_0} = \boldsymbol{C}_Q\boldsymbol{J}^{-1})$ with $\boldsymbol{J}$ defined by $\boldsymbol{G} = \boldsymbol{J}^H\boldsymbol{J}$ and resetting $\boldsymbol{W} = \boldsymbol{0}$ so that due to (4.35) there is $\boldsymbol{G} = \boldsymbol{G}_0 = \boldsymbol{1}$ and condition (4.36) again becomes valid.

Proof Equations (4.34), (4.35) and condition (4.36) are taken from (3.61), (3.62) and (3.63) respectively. Note that $\boldsymbol{G} > 0$ is required so that $\boldsymbol{T} > 0, \boldsymbol{V} > 0$ and since $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ is a lossless completion also $\boldsymbol{U} > 0$. The first order variations of the stable observable pair $(\boldsymbol{A}_Q, \boldsymbol{C}_Q)$ can be calculated from (3.60) and can then be given by

$$\begin{aligned}
\delta\boldsymbol{A}_Q &= \boldsymbol{A}_{Q_0}\delta\boldsymbol{T}\boldsymbol{G} + \boldsymbol{A}_{Q_0}\boldsymbol{T}\delta\boldsymbol{G} - \boldsymbol{G}^{-1}\delta\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{S}\boldsymbol{W} + \\
&\quad + \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\delta\boldsymbol{S}\boldsymbol{W} + \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{S}\delta\boldsymbol{W} \\
\delta\boldsymbol{C}_Q &= \boldsymbol{C}_{Q_0}\delta\boldsymbol{T}\boldsymbol{G} + \boldsymbol{C}_{Q_0}\boldsymbol{T}\delta\boldsymbol{G} + \boldsymbol{D}_0\delta\boldsymbol{S}\boldsymbol{W} + \boldsymbol{D}_0\boldsymbol{S}\boldsymbol{W}
\end{aligned}$$

using (A.81), (A.82). The first order variations of $\boldsymbol{G}, \boldsymbol{S}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{V}$ can be calculated from (4.34) and (4.35). First

$$\boldsymbol{A}_{Q_0}^H\delta\boldsymbol{G}\boldsymbol{A}_{Q_0} = \delta\boldsymbol{G} + \delta\boldsymbol{W}^H\boldsymbol{W} + \boldsymbol{W}^H\delta\boldsymbol{W}$$

then since $\boldsymbol{S}^{-1} = \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{U}^{\frac{1}{2}}$ we get

$$\delta\boldsymbol{S} + \boldsymbol{S}\delta\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{U}^{\frac{1}{2}}\boldsymbol{S} + \boldsymbol{S}\boldsymbol{V}^{\frac{1}{2}}\delta\boldsymbol{U}^{\frac{1}{2}}\boldsymbol{S} = \boldsymbol{0}$$

and since $\boldsymbol{T}^{-1} = \boldsymbol{G} + \boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{A}_{Q_0}^H\boldsymbol{G}\boldsymbol{A}_{Q_0} + \boldsymbol{C}_{Q_0}^H\boldsymbol{C}_{Q_0}$ we get

$$\delta\boldsymbol{T} + \boldsymbol{T}\boldsymbol{A}_{Q_0}^H\delta\boldsymbol{G}\boldsymbol{A}_{Q_0}\boldsymbol{T} = \boldsymbol{0}$$

and finally

$$\begin{aligned}
&\boldsymbol{U}^{\frac{1}{2}}\delta\boldsymbol{U}^{\frac{1}{2}} + \delta\boldsymbol{U}^{\frac{1}{2}}\boldsymbol{U}^{\frac{1}{2}} + \boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\delta\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0 = \boldsymbol{0} \\
&\boldsymbol{V}^{\frac{1}{2}}\delta\boldsymbol{V}^{\frac{1}{2}} + \delta\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{V}^{\frac{1}{2}} + \boldsymbol{W}\boldsymbol{G}^{-1}\delta\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{W}^H = \delta\boldsymbol{W}\boldsymbol{G}^{-1}\boldsymbol{W}^H + \boldsymbol{W}\boldsymbol{G}^{-1}\delta\boldsymbol{W}^H
\end{aligned}$$

using (A.81), (A.82), (A.83). According to (A.77) the matrices $\boldsymbol{U}^{\frac{1}{2}} > 0, \boldsymbol{V}^{\frac{1}{2}} > 0$ are the unique positive matrix square roots of $\boldsymbol{U}, \boldsymbol{V}$. The first order directional

variations of these variables can then be given by

$$\begin{aligned}
&\mathrm{Re}(\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q) = \mathrm{Re}((\boldsymbol{T}\boldsymbol{A}_{Q_0}^H \boldsymbol{\Omega}_{A_Q} - \boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G} + \\
&+ (\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H) \bullet \delta\boldsymbol{S} + (\boldsymbol{A}_{Q_0}^H\boldsymbol{\Omega}_{A_Q}\boldsymbol{G}) \bullet \delta\boldsymbol{T} + \\
&+ (\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}) \bullet \delta\boldsymbol{W}) \qquad (4.37)\\
&\mathrm{Re}(\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q) = \mathrm{Re}((\boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q}) \bullet \delta\boldsymbol{G} + (\boldsymbol{D}_0^H\boldsymbol{\Omega}_{C_Q}\boldsymbol{W}^H) \bullet \delta\boldsymbol{S} + \\
&+ (\boldsymbol{C}_{Q_0}^H\boldsymbol{\Omega}_{C_Q}\boldsymbol{G}) \bullet \delta\boldsymbol{T} + (\boldsymbol{S}^H\boldsymbol{D}_0^H\boldsymbol{\Omega}_{C_Q}) \bullet \delta\boldsymbol{W}) \qquad (4.38)\\
&\mathrm{Re}((\boldsymbol{A}_{Q_0}\boldsymbol{L}\boldsymbol{A}_{Q_0}^H - \boldsymbol{L}) \bullet \delta\boldsymbol{G}) = \mathrm{Re}((\boldsymbol{W}(\boldsymbol{L} + \boldsymbol{L}^H)) \bullet \delta\boldsymbol{W}) \qquad (4.39)\\
&\mathrm{Re}(\boldsymbol{\Omega} \bullet \delta\boldsymbol{S}) = -\mathrm{Re}((\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{S}^H\boldsymbol{\Omega}\boldsymbol{S}^H) \bullet \delta\boldsymbol{U}^{\frac{1}{2}} + (\boldsymbol{S}^H\boldsymbol{\Omega}\boldsymbol{S}^H\boldsymbol{U}^{\frac{1}{2}}) \bullet \delta\boldsymbol{V}^{\frac{1}{2}}) \qquad (4.40)\\
&\mathrm{Re}(\boldsymbol{\Omega} \bullet \delta\boldsymbol{T}) = -\mathrm{Re}((\boldsymbol{A}_{Q_0}\boldsymbol{T}\boldsymbol{\Omega}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H) \bullet \delta\boldsymbol{G}) \qquad (4.41)\\
&\mathrm{Re}((\boldsymbol{U}^{\frac{1}{2}}\boldsymbol{M} + \boldsymbol{M}\boldsymbol{U}^{\frac{1}{2}}) \bullet \delta\boldsymbol{U}^{\frac{1}{2}}) = -\mathrm{Re}((\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{M}\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G}) \qquad (4.42)\\
&\mathrm{Re}((\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{N} + \boldsymbol{N}\boldsymbol{V}^{\frac{1}{2}}) \bullet \delta\boldsymbol{V}^{\frac{1}{2}}) = \mathrm{Re}(((\boldsymbol{N} + \boldsymbol{N}^H)\boldsymbol{W}\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{W} - \\
&- (\boldsymbol{G}^{-1}\boldsymbol{W}^H\boldsymbol{N}\boldsymbol{W}\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G}) \qquad (4.43)
\end{aligned}$$

using (A.69), (A.70) and particularly $\mathrm{Re}(\boldsymbol{A} \bullet \boldsymbol{B}) = \mathrm{Re}(\overline{\boldsymbol{B} \bullet \boldsymbol{A}}) = \mathrm{Re}(\boldsymbol{B} \bullet \boldsymbol{A}) = \mathrm{Re}(\boldsymbol{A}^H \bullet \boldsymbol{B}^H)$. Using (4.40), (4.41) on (4.37) leads to

$$\begin{aligned}
\mathrm{Re}(\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q) = \mathrm{Re}((&\boldsymbol{T}\boldsymbol{A}_{Q_0}^H \boldsymbol{\Omega}_{A_Q} - \boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G} - \\
&- (\boldsymbol{A}_{Q_0}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H\boldsymbol{\Omega}_{A_Q}\boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H) \bullet \delta\boldsymbol{G} + (\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}) \bullet \delta\boldsymbol{W} - \\
&- (\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H) \bullet \delta\boldsymbol{U}^{\frac{1}{2}} - \\
&- (\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{U}^{\frac{1}{2}}) \bullet \delta\boldsymbol{V}^{\frac{1}{2}})
\end{aligned}$$

According to (A.86) the Sylvester equations

$$\begin{aligned}
\boldsymbol{U}^{\frac{1}{2}}\boldsymbol{M}_A + \boldsymbol{M}_A\boldsymbol{U}^{\frac{1}{2}} &= \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H \\
\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{N}_A + \boldsymbol{N}_A\boldsymbol{V}^{\frac{1}{2}} &= \boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{U}^{\frac{1}{2}}
\end{aligned}$$

have unique solutions $\boldsymbol{M}_A, \boldsymbol{N}_A$ since $\boldsymbol{U}^{\frac{1}{2}} > 0, \boldsymbol{V}^{\frac{1}{2}} > 0$. With (4.42) and (4.43) this leads to

$$\begin{aligned}
\mathrm{Re}(\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q) = \mathrm{Re}((&\boldsymbol{T}\boldsymbol{A}_{Q_0}^H \boldsymbol{\Omega}_{A_Q} - \boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G} - \\
&- (\boldsymbol{A}_{Q_0}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H\boldsymbol{\Omega}_{A_Q}\boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H) \bullet \delta\boldsymbol{G} + (\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}) \bullet \delta\boldsymbol{W} + \\
&+ (\boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{M}_A\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1} + \boldsymbol{G}^{-1}\boldsymbol{W}^H\boldsymbol{N}_A\boldsymbol{W}\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{G} - \\
&- ((\boldsymbol{N}_A + \boldsymbol{N}_A^H)\boldsymbol{W}\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{W})
\end{aligned}$$

According to (A.88) the Stein equation

$$\begin{aligned}
\boldsymbol{A}_{Q_0}\boldsymbol{L}_A\boldsymbol{A}_{Q_0}^H - \boldsymbol{L}_A = &\boldsymbol{T}\boldsymbol{A}_{Q_0}^H\boldsymbol{\Omega}_{A_Q} - \boldsymbol{A}_{Q_0}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H\boldsymbol{\Omega}_{A_Q}\boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H - \\
&- \boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q}\boldsymbol{W}^H\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1} + \\
&+ \boldsymbol{G}^{-1}\boldsymbol{G}_0\boldsymbol{B}_0\boldsymbol{M}_A\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1} + \boldsymbol{G}^{-1}\boldsymbol{W}^H\boldsymbol{N}_A\boldsymbol{W}\boldsymbol{G}^{-1}
\end{aligned}$$

has a unique solution $\boldsymbol{L}_A$ since $\boldsymbol{A}_{Q_0}$ is stable. With (4.39) this leads to

$$\begin{aligned}
\mathrm{Re}(\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q) = \mathrm{Re}((&\boldsymbol{S}^H\boldsymbol{B}_0^H\boldsymbol{G}_0\boldsymbol{G}^{-1}\boldsymbol{\Omega}_{A_Q} + \boldsymbol{W}(\boldsymbol{L}_A + \boldsymbol{L}_A^H) - \\
&- (\boldsymbol{N}_A + \boldsymbol{N}_A^H)\boldsymbol{W}\boldsymbol{G}^{-1}) \bullet \delta\boldsymbol{W}) \qquad (4.44)
\end{aligned}$$

Using (4.40), (4.41) on (4.38) leads to

$$\begin{aligned}\mathrm{Re}(\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q) = \mathrm{Re}((\boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q}) \bullet \delta \boldsymbol{G} - \\ - (\boldsymbol{A}_{Q_0} \boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H) \bullet \delta \boldsymbol{G} + (\boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q}) \bullet \delta \boldsymbol{W} - \\ - (\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H) \bullet \delta \boldsymbol{U}^{\frac{1}{2}} - \\ - (\boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H \boldsymbol{U}^{\frac{1}{2}}) \bullet \delta \boldsymbol{V}^{\frac{1}{2}})\end{aligned}$$

According to (A.86) the Sylvester equations

$$\begin{aligned}\boldsymbol{U}^{\frac{1}{2}} \boldsymbol{M}_C + \boldsymbol{M}_C \boldsymbol{U}^{\frac{1}{2}} &= \boldsymbol{V}^{\frac{1}{2}} \boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H \\ \boldsymbol{V}^{\frac{1}{2}} \boldsymbol{N}_C + \boldsymbol{N}_C \boldsymbol{V}^{\frac{1}{2}} &= \boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{W}^H \boldsymbol{S}^H \boldsymbol{U}^{\frac{1}{2}}\end{aligned}$$

have unique solutions $\boldsymbol{M}_C, \boldsymbol{N}_C$ since $\boldsymbol{U}^{\frac{1}{2}} > 0, \boldsymbol{V}^{\frac{1}{2}} > 0$. With (4.42) and (4.43) this leads to

$$\begin{aligned}\mathrm{Re}(\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q) = \mathrm{Re}((\boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q}) \bullet \delta \boldsymbol{G} - \\ - (\boldsymbol{A}_{Q_0} \boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H) \bullet \delta \boldsymbol{G} + (\boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q}) \bullet \delta \boldsymbol{W} + \\ + (\boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \boldsymbol{M}_C \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \boldsymbol{W}^H \boldsymbol{N}_C \boldsymbol{W} \boldsymbol{G}^{-1}) \bullet \delta \boldsymbol{G} - \\ - ((\boldsymbol{N}_C + \boldsymbol{N}_C^H) \boldsymbol{W} \boldsymbol{G}^{-1}) \bullet \delta \boldsymbol{W})\end{aligned}$$

According to (A.88) the Stein equation

$$\begin{aligned}\boldsymbol{A}_{Q_0} \boldsymbol{L}_C \boldsymbol{A}_{Q_0}^H - \boldsymbol{L}_C = \boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} - \boldsymbol{A}_{Q_0} \boldsymbol{T}\boldsymbol{C}_{Q_0}^H \boldsymbol{\Omega}_{C_Q} \boldsymbol{G}\boldsymbol{T}\boldsymbol{A}_{Q_0}^H + \\ + \boldsymbol{G}^{-1} \boldsymbol{G}_0 \boldsymbol{B}_0 \boldsymbol{M}_C \boldsymbol{B}_0^H \boldsymbol{G}_0 \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \boldsymbol{W}^H \boldsymbol{N}_C \boldsymbol{W} \boldsymbol{G}^{-1}\end{aligned}$$

has a unique solution $\boldsymbol{L}_C$ since $\boldsymbol{A}_{Q_0}$ is stable. With (4.39) this leads to

$$\begin{aligned}\mathrm{Re}(\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q) = \mathrm{Re}((\boldsymbol{S}^H \boldsymbol{D}_0^H \boldsymbol{\Omega}_{C_Q} + \boldsymbol{W}(\boldsymbol{L}_C + \boldsymbol{L}_C^H) - \\ - (\boldsymbol{N}_C + \boldsymbol{N}_C^H) \boldsymbol{W} \boldsymbol{G}^{-1}) \bullet \delta \boldsymbol{W}) \quad (4.45)\end{aligned}$$

The above Stein and Sylvester equations are collected in (4.31). Replacing the directional variations $\boldsymbol{\Omega}_{A_Q} \bullet \delta \boldsymbol{A}_Q$ and $\boldsymbol{\Omega}_{C_Q} \bullet \delta \boldsymbol{C}_Q$ in (4.26) by the terms in (4.44) and (4.45) respectively leads to the equation for the subgradient (4.32) with $\boldsymbol{\Omega}_W$ given by (4.33).

4.2.9 Realness Constraint

To be practically usable, the parametrized controller $\boldsymbol{K}$ must remain real in the sense of (C.27) so that additional constraints have to be considered, according to the following idea, see also figure 4.2. If the plant $\boldsymbol{P}$ and the initial controller $\boldsymbol{K}_0$ is real, then it is shown that a real initial parameter $\boldsymbol{Q}_0$ can be found among the possible initial parameters $\boldsymbol{Q}_0$. It is then possible to keep the parameter $\boldsymbol{Q}$ real by satisfying some constraints. If the dynamic matrix $\boldsymbol{A}_P$ of the given plant $\boldsymbol{P}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_Q$ of the current real parameter $\boldsymbol{Q}$, then the reconstructed controller $\boldsymbol{K}$ is real.

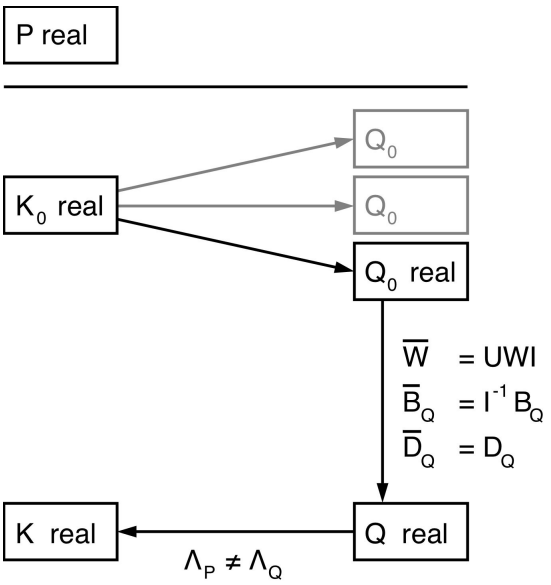


Figure 4.2: Principal idea behind the process to assure realness of the controller $\boldsymbol{K}$ for a real plant $\boldsymbol{P}$. The spectra of the dynamic matrices $\boldsymbol{A}_P, \boldsymbol{A}_Q$ are denoted by $\boldsymbol{\Lambda}_P, \boldsymbol{\Lambda}_Q$ respectively.

Proposition If the Riccati equation (4.12) has solutions and if the plant $\boldsymbol{P}$ is real, then

$$\boldsymbol{K} \text{ real } \Rightarrow \text{ exists a } \boldsymbol{Q} \text{ real} \tag{4.46}$$

and $\boldsymbol{K}, \boldsymbol{Q}$ share the same realness transformation matrix $\boldsymbol{I}$

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix}} \\ \begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix}} \end{aligned} \tag{4.47}$$

Proof Plant $\boldsymbol{P}$ and controller $\boldsymbol{K}$ are real if and only if there exist $\boldsymbol{I}_P, \boldsymbol{I}$ so that

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \boldsymbol{C}_P & \boldsymbol{D}_P \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \boldsymbol{C}_P & \boldsymbol{D}_P \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \\ \begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix}} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \end{aligned} \tag{4.48}$$

according to (C.28). Hence

$$\begin{bmatrix} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \boldsymbol{C}_H & \boldsymbol{D}_H \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_H & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_H & \boldsymbol{B}_H \\ \boldsymbol{C}_H & \boldsymbol{D}_H \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}_H & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \tag{4.49}$$

with $(\boldsymbol{A}_H, \boldsymbol{B}_H, \boldsymbol{C}_H, \boldsymbol{D}_H)$ being the realization of the loop $\boldsymbol{H}$ as given in (2.19) and with

$$\boldsymbol{I}_H = \begin{bmatrix} \boldsymbol{I}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{4.50}$$

which shows that the loop $\boldsymbol{H}$ is real. With (4.48) and (4.50) and by using the substitutions in (4.13) equation (4.49) provides

$$\begin{aligned}\boldsymbol{A}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{C}_Q &= \boldsymbol{I}\overline{\boldsymbol{A}}_Q\boldsymbol{I}^{-1} + \boldsymbol{R}'\boldsymbol{B}_P\overline{\boldsymbol{C}}_Q\boldsymbol{I}^{-1}\\ \boldsymbol{B}_Q + \boldsymbol{R}\boldsymbol{B}_P\boldsymbol{D}_Q &= \boldsymbol{I}\overline{\boldsymbol{B}}_Q + \boldsymbol{R}'\boldsymbol{B}_P\overline{\boldsymbol{D}}_Q\\ \boldsymbol{C}_Q &= \overline{\boldsymbol{C}}_Q\boldsymbol{I}^{-1}\\ \boldsymbol{D}_Q &= \overline{\boldsymbol{D}}_Q\end{aligned}$$

with $\boldsymbol{R}' = \boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1}$. This can be written in block matrix form

$$\begin{bmatrix}\boldsymbol{1} & (\boldsymbol{R}-\boldsymbol{R}')\boldsymbol{B}_P\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q\\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q\\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix}}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}^{-1} \tag{4.51}$$

Matrix $\boldsymbol{R}$ is a solution of the Riccati equation (4.12) which can be written as the Sylvester equation (4.14) using the substitutions in (4.13). With (4.48) the conjugate of (4.14) provides

$$\boldsymbol{R}'\boldsymbol{A}_P = \boldsymbol{I}\overline{\boldsymbol{A}}_Q\boldsymbol{I}^{-1}\boldsymbol{R}' + \boldsymbol{I}\overline{\boldsymbol{B}}_Q\boldsymbol{C}_P$$

Subtracting this from (4.14) gives

$$(\boldsymbol{R}-\boldsymbol{R}')\boldsymbol{A}_P = \boldsymbol{A}_Q\boldsymbol{R} - \boldsymbol{I}\overline{\boldsymbol{A}}_Q\boldsymbol{I}^{-1}\boldsymbol{R}' + \boldsymbol{B}_Q\boldsymbol{C}_P - \boldsymbol{I}\overline{\boldsymbol{B}}_Q\boldsymbol{C}_P$$

Using (4.51) this simplifies to

$$\boldsymbol{A}_Q(\boldsymbol{R}-\boldsymbol{R}') = (\boldsymbol{R}-\boldsymbol{R}')(\boldsymbol{A}_P + \boldsymbol{B}_P(\boldsymbol{C}_Q\boldsymbol{R}' + \boldsymbol{D}_Q\boldsymbol{C}_P))$$

One solution of this Riccati equation is $\boldsymbol{R} = \boldsymbol{R}' = \boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1}$. This solution leads to $\boldsymbol{Q}$ real as can be seen in (4.51) with transformation matrix $\boldsymbol{I}$.

Proposition Realness of the parameter $\boldsymbol{Q}$ is preserved

$$\boldsymbol{Q}_0 \text{ real } \Rightarrow \boldsymbol{Q} \text{ real} \tag{4.52}$$

if and only if the parameter $\boldsymbol{Q}$ that is itself parametrized by $(\boldsymbol{W}, \boldsymbol{B}_Q, \boldsymbol{D}_Q)$ satisfies

$$\begin{aligned}\overline{\boldsymbol{W}} &= \boldsymbol{U}\boldsymbol{W}\boldsymbol{I}\\ \overline{\boldsymbol{B}}_Q &= \boldsymbol{I}^{-1}\boldsymbol{B}_Q\\ \overline{\boldsymbol{D}}_Q &= \boldsymbol{D}_Q\end{aligned} \tag{4.53}$$

with invertible $\boldsymbol{I}$ uniquely defined by

$$\begin{bmatrix}\boldsymbol{A}_{Q_0}\\ \boldsymbol{C}_{Q_0}\end{bmatrix}\boldsymbol{I} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{A}_{Q_0}\\ \boldsymbol{C}_{Q_0}\end{bmatrix}} \tag{4.54}$$

and unitary $\boldsymbol{U}$ defined by

$$\boldsymbol{U} = \overline{\boldsymbol{B}}_0^H\boldsymbol{I}^H\boldsymbol{G}_0\boldsymbol{B}_0 + \overline{\boldsymbol{D}}_0^H\boldsymbol{D}_0 \tag{4.55}$$

while the pair $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ is a lossless completion (3.6) of the stable observable pair $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$. Then $\boldsymbol{Q}_0, \boldsymbol{Q}$ share the same realness transformation matrix $\boldsymbol{I}$

$$\begin{aligned}\begin{bmatrix}\boldsymbol{A}_{Q_0} & \boldsymbol{B}_{Q_0}\\ \boldsymbol{C}_{Q_0} & \boldsymbol{D}_{Q_0}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{A}_{Q_0} & \boldsymbol{B}_{Q_0}\\ \boldsymbol{C}_{Q_0} & \boldsymbol{D}_{Q_0}\end{bmatrix}}\\ \begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q\\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix} &= \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}\overline{\begin{bmatrix}\boldsymbol{A}_Q & \boldsymbol{B}_Q\\ \boldsymbol{C}_Q & \boldsymbol{D}_Q\end{bmatrix}}\end{aligned} \tag{4.56}$$

Proof Parameter $\boldsymbol{Q}_0$ realized by $(\boldsymbol{A}_{Q_0}, \boldsymbol{B}_{Q_0}, \boldsymbol{C}_{Q_0}, \boldsymbol{D}_{Q_0})$ is real if and only if there exists a matrix $\boldsymbol{I}$ so that

$$\begin{bmatrix} \boldsymbol{A}_{Q_0} & \boldsymbol{B}_{Q_0} \\ \boldsymbol{C}_{Q_0} & \boldsymbol{D}_{Q_0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_{Q_0} & \boldsymbol{B}_{Q_0} \\ \boldsymbol{C}_{Q_0} & \boldsymbol{D}_{Q_0} \end{bmatrix}} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1}$$

according to (C.28). The first block column in particular gives (4.54) hence condition (3.43) is satisfied and (3.44) can be applied so that

$$\begin{bmatrix} \boldsymbol{A}_Q \\ \boldsymbol{C}_Q \end{bmatrix} \boldsymbol{I} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_Q \\ \boldsymbol{C}_Q \end{bmatrix}} \Leftrightarrow \overline{\boldsymbol{W}} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{I} \tag{4.57}$$

with unitary $\boldsymbol{U}$ defined by (3.45). Now $\boldsymbol{Q}$ is real if and only if there exists a matrix $\boldsymbol{T}$ so that

$$\begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix}} \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \tag{4.58}$$

according to (C.28). Since the pair $(\boldsymbol{A}_Q, \boldsymbol{C}_Q)$ is observable $\boldsymbol{I}$ is uniquely defined by (4.57) according to (3.10). Comparing this to (4.58) shows that $\boldsymbol{T} = \boldsymbol{I}$ since $\boldsymbol{I}$ is unique, so that realness of $\boldsymbol{Q}$ follows from realness of $\boldsymbol{Q}_0$ if and only if the conditions (4.53) are satisfied.

Proposition If the dynamic matrix $\boldsymbol{A}_P$ of the given plant $\boldsymbol{P}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_Q$ of the current parameter $\boldsymbol{Q}$ and if the plant $\boldsymbol{P}$ is real, then

$$\boldsymbol{Q} \text{ real } \Rightarrow \boldsymbol{K} \text{ real} \tag{4.59}$$

and $\boldsymbol{Q}, \boldsymbol{K}$ share the same realness transformation matrix $\boldsymbol{I}$

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix}} \\ \begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix}} \end{aligned} \tag{4.60}$$

Proof Plant $\boldsymbol{P}$ and parameter $\boldsymbol{Q}$ are real if and only if there exist $\boldsymbol{I}_P, \boldsymbol{I}$ so that

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \boldsymbol{C}_P & \boldsymbol{D}_P \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \boldsymbol{C}_P & \boldsymbol{D}_P \end{bmatrix}} \begin{bmatrix} \boldsymbol{I}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \\ \begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix} &= \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix}} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \end{aligned} \tag{4.61}$$

according to (C.28). With (4.61) the conjugate of (4.14) provides

$$\boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1}\boldsymbol{A}_P = \boldsymbol{A}_Q\boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1} + \boldsymbol{B}_Q\boldsymbol{C}_P$$

Subtracting this from (4.14) gives

$$\boldsymbol{A}_Q(\boldsymbol{R} - \boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1}) - (\boldsymbol{R} - \boldsymbol{I}\overline{\boldsymbol{R}}\boldsymbol{I}_P^{-1})\boldsymbol{A}_P = \boldsymbol{0}$$

This is a Sylvester equation and according to (A.86) has a unique solution

$$\overline{\boldsymbol{R}} = \boldsymbol{I}^{-1}\boldsymbol{R}\boldsymbol{I}_P \tag{4.62}$$

if and only if $\lambda_{A_P} \neq \lambda_{A_Q}$ for all eigenvalues $\lambda_{A_P}, \lambda_{A_Q}$ of $\boldsymbol{A}_P, \boldsymbol{A}_Q$ respectively. With (4.61) and (4.62) it follows from (4.15) that

$$\overline{\boldsymbol{L}}_K = \boldsymbol{I}^{-1}\boldsymbol{L}_K \tag{4.63}$$

and with (4.63) further

$$\overline{\begin{bmatrix} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \boldsymbol{C}_K & \boldsymbol{D}_K \end{bmatrix}} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{L}_K \\ \boldsymbol{0} & \boldsymbol{E}_{KP} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_Q & \boldsymbol{B}_Q \\ \boldsymbol{C}_Q & \boldsymbol{D}_Q \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}$$

so that $\boldsymbol{K}$ is real with transformation matrix $\boldsymbol{I}$.

4.3 Sufficient Conditions

The proposed multiobjective optimization method presented in the preceding sections roughly consists of first constructing a controller parameter $\boldsymbol{Q}$ for a given initial controller $\boldsymbol{K}_0$ and constructing initial L1 norm parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$, then calculating the local subgradient and thereby constantly improving the parameters $(\boldsymbol{Q}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and finally reconstructing a final controller $\boldsymbol{K}^*$ from the improved parameter $\boldsymbol{Q}$. While a detailed algorithm is presented in section 4.4, this section provides some sufficient conditions for these constructions, calculations and reconstructions to work.

4.3.1 Sufficient Condition for Parameter Construction

A necessary and sufficient condition for the ability to construct a parameter $\boldsymbol{Q}$ from an initial controller $\boldsymbol{K}_0$ is the existence of a solution $\boldsymbol{R}$ to the Riccati equation (4.12). A sufficient condition for the existence of $\boldsymbol{R}$ can be given.

Proposition If the dynamic matrix $\boldsymbol{A}_H$ of the closed loop $\boldsymbol{H}$ formed by the given plant $\boldsymbol{P}$ and a given controller $\boldsymbol{K}$ is diagonalizable, then there exists a not necessarily unique $\boldsymbol{R}$ so that a parameter $\boldsymbol{Q}$ can be constructed from the given controller $\boldsymbol{K}$. If in addition to that $\boldsymbol{H}$ is stable and $\boldsymbol{K}$ is observable, then $\boldsymbol{Q}$ is stable and observable.

Proof If $\boldsymbol{A}_H$ is diagonalizable then according to (2.30) a not necessarily unique solution $\boldsymbol{R} = \boldsymbol{R}_{KP}$ to the Riccati equation (2.21), which is equal to the Riccati equation (4.12), exists. Then a parameter $\boldsymbol{Q}$ can be constructed from controller $\boldsymbol{K}$ by application of (4.13). If the closed loop $\boldsymbol{H}$ is stable then $\boldsymbol{K}$ is a stabilizing controller and thus $\boldsymbol{Q}$ is stable due to (2.46). If $\boldsymbol{K}$ is observable then $\boldsymbol{Q} = \boldsymbol{Q}_L$ is observable due to (2.45).

4.3.2 Sufficient Condition for Subgradient Calculation

A necessary and sufficient condition for the ability to write the subgradient (4.20) in the explicit form (4.32) is the existence of solutions $\boldsymbol{F}_i$ to the Sylvester equations (4.25) and the existence of solutions $\boldsymbol{L}_A, \boldsymbol{L}_C, \boldsymbol{M}_A, \boldsymbol{M}_C, \boldsymbol{N}_A, \boldsymbol{N}_C$ to the Stein and Sylvester equations (4.31). A sufficient condition for the existence of $\boldsymbol{F}_i, \boldsymbol{L}_A, \boldsymbol{L}_C, \boldsymbol{M}_A, \boldsymbol{M}_C, \boldsymbol{N}_A, \boldsymbol{N}_C$ can be given.

Proposition If the dynamic matrix $\boldsymbol{A}_P$ of the given plant $\boldsymbol{P}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_Q$ of the parameter $\boldsymbol{Q}$ and if $\boldsymbol{Q}$ is stable and observable, then there exist unique $\boldsymbol{F}_i, \boldsymbol{L}_A, \boldsymbol{L}_C, \boldsymbol{M}_A, \boldsymbol{M}_C, \boldsymbol{N}_A, \boldsymbol{N}_C$ and the subgradient of the cost function f can be given in the explicit form of (4.32) so that it can be calculated.

Proof The subgradient in (4.20) can be written in the form (4.21). Due to (A.86) the Sylvester equations (4.25) have unique solutions $\boldsymbol{F}_i$ if and only if $\lambda_{A_P} \neq \lambda_{A_Q}$ for all eigenvalues $\lambda_{A_P}, \lambda_{A_Q}$ of $\boldsymbol{A}_P, \boldsymbol{A}_Q$ respectively. Then the subgradient (4.21) can be written in the form (4.26). If the pair $(\boldsymbol{A}_Q, \boldsymbol{C}_Q)$ of $\boldsymbol{Q}$ is stable and observable it can be set as the center of a chart ($\boldsymbol{A}_{Q_0} = \boldsymbol{A}_Q, \boldsymbol{C}_{Q_0} = \boldsymbol{C}_Q$). Hence the Stein and Sylvester equations (4.31) have unique solutions and thus the subgradient (4.26) can be written in the form (4.32).

4.3.3 Sufficient Condition for Controller Reconstruction

A necessary and sufficient condition for the ability to reconstruct a final controller $\boldsymbol{K}^*$ from a parameter $\boldsymbol{Q}$ is the existence of a solution $\boldsymbol{R}$ to the Sylvester equation (4.14). A sufficient condition for the existence of $\boldsymbol{R}$ can be given.

Proposition If the dynamic matrix $\boldsymbol{A}_P$ of the given plant $\boldsymbol{P}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_Q$ of the current parameter $\boldsymbol{Q}$, then there exists a unique $\boldsymbol{R}$ so that a controller $\boldsymbol{K}$ can be uniquely reconstructed from the current parameter $\boldsymbol{Q}$.

Proof Due to (A.86) the Sylvester equation (4.14) has a unique solution $\boldsymbol{R}$ if and only if $\lambda_{A_P} \neq \lambda_{A_Q}$ for all eigenvalues $\lambda_{A_P}, \lambda_{A_Q}$ of $\boldsymbol{A}_P, \boldsymbol{A}_Q$ respectively. Then a controller $\boldsymbol{K}$ can be uniquely reconstructed from parameter $\boldsymbol{Q}$ by application of (4.15).

4.3.4 Sufficient Condition for L1 Parameter Construction

In order to construct initial L1 parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ for extended loops $\boldsymbol{H}_i$ controlled by an initial controller $\boldsymbol{K}_0$, upper bounds of the L1 norms of the extended loops $\boldsymbol{H}_i$ have to be found by solving the bilinear matrix inequalities in (4.6) which is difficult. A method for iteratively solving the bilinear matrix inequalities is to alternately fix α_i, β_i and $\boldsymbol{Z}_i$ and to solve the resulting linear matrix inequalities at each step in order to decrease the upper bound γ_i. That way the linear matrix inequalities for fixed α_i, β_i are solved for $\boldsymbol{Z}_i$ and provide a potentially better $\boldsymbol{Z}_i$ for the following linear matrix inequality for fixed $\boldsymbol{Z}_i$ and so on. A sufficient condition for the existence of initial $\alpha_i, \beta_i, \boldsymbol{Z}_i$ can be given so that the iterative process can be started, eventually resulting in the initial L1 parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$.

Proposition If the extended loop $\boldsymbol{H}_i$ is stable, then valid initial $\alpha_i, \beta_i, \boldsymbol{Z}_i$ can be derived from the spectral radius ρ_i of the extended loop dynamic matrix $\boldsymbol{A}_i$ by selecting an $\alpha_i \in]0, 1 - \rho_i^2[$, solving the first linear matrix inequality in (4.6) for $\boldsymbol{Z}_i$ and then solving the second linear matrix inequality in (4.6) for β_i. The spectral radii ρ_i can be collected in a radius vector $\boldsymbol{\rho}$.

Proof The proof follows the ideas presented in [1] applied to the discrete case. If $\boldsymbol{A}_i$ is stable then it has a spectra radius bounded by $\rho_i = \max_k |\lambda_k| < 1$ with λ_k denoting the eigenvalues of $\boldsymbol{A}_i$. Now select $\alpha_i \in]0, 1-\rho_i^2[$ then $\frac{1}{\sqrt{1-\alpha_i}}\boldsymbol{A}_i$ is still stable with spectral radius $\frac{1}{\sqrt{1-\alpha_i}}\rho_i < 1$. Since $\frac{1}{\sqrt{1-\alpha_i}}\boldsymbol{A}_i$ is stable the Stein equation in $\boldsymbol{Z}_i^{-1}$

$$\frac{1}{1-\alpha_i}\boldsymbol{A}_i\boldsymbol{Z}_i^{-1}\boldsymbol{A}_i^H - \boldsymbol{Z}_i^{-1} + \frac{1}{\alpha_i}\boldsymbol{B}_i\boldsymbol{B}_i^H + \boldsymbol{E}_i = \boldsymbol{0}$$

with $\boldsymbol{E}_i > 0$ so that $\frac{1}{\alpha_i}\boldsymbol{B}_i\boldsymbol{B}_i^H + \boldsymbol{E}_i > 0$ has a unique solution $\boldsymbol{Z}_i^{-1} > 0$ according to (A.89). Thus there always exists a $\boldsymbol{Z}_i^{-1} > 0$ for any $\alpha_i \in]0, 1-\rho_i^2[$ so that

$$\frac{1}{1-\alpha_i}\boldsymbol{A}_i\boldsymbol{Z}_i^{-1}\boldsymbol{A}_i^H - \boldsymbol{Z}_i^{-1} + \frac{1}{\alpha_i}\boldsymbol{B}_i\boldsymbol{B}_i^H < 0$$

which is equivalent to

$$\boldsymbol{Z}_i^{-1} - \begin{bmatrix}\boldsymbol{A}_i & \boldsymbol{B}_i\end{bmatrix}\begin{bmatrix}\frac{1}{1-\alpha_i}\boldsymbol{Z}_i^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{\alpha_i}\boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_i^H \\ \boldsymbol{B}_i^H\end{bmatrix} > 0$$

Using (A.65) this is equivalent to

$$\begin{bmatrix}\boldsymbol{Z}_i^{-1} & \boldsymbol{A}_i & \boldsymbol{B}_i \\ \boldsymbol{A}_i^H & (1-\alpha_i)\boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{B}_i^H & \boldsymbol{0} & \alpha_i\boldsymbol{1}\end{bmatrix} > 0$$

Using (A.64) this is equivalent to

$$\begin{bmatrix}\boldsymbol{Z}_i & \boldsymbol{Z}_i\boldsymbol{A}_i & \boldsymbol{Z}_i\boldsymbol{B}_i \\ \boldsymbol{A}_i^H\boldsymbol{Z}_i & (1-\alpha_i)\boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{B}_i^H\boldsymbol{Z}_i & \boldsymbol{0} & \alpha_i\boldsymbol{1}\end{bmatrix} > 0$$

Using (A.65) this is equivalent to

$$\begin{bmatrix}(1-\alpha_i)\boldsymbol{Z}_i & \boldsymbol{0} \\ \boldsymbol{0} & \alpha_i\boldsymbol{1}\end{bmatrix} - \begin{bmatrix}\boldsymbol{A}_i^H\boldsymbol{Z}_i \\ \boldsymbol{B}_i^H\boldsymbol{Z}_i\end{bmatrix}\boldsymbol{Z}_i^{-1}\begin{bmatrix}\boldsymbol{Z}_i\boldsymbol{A}_i & \boldsymbol{Z}_i\boldsymbol{B}_i\end{bmatrix} > 0$$

This finally gives

$$\begin{bmatrix}\boldsymbol{A}_i^H\boldsymbol{Z}_i\boldsymbol{A}_i + (\alpha_i - 1)\boldsymbol{Z}_i & \boldsymbol{A}_i^H\boldsymbol{Z}_i\boldsymbol{B}_i \\ \boldsymbol{B}_i^H\boldsymbol{Z}_i\boldsymbol{A}_i & \boldsymbol{B}_i^H\boldsymbol{Z}_i\boldsymbol{B}_i - \alpha_i\boldsymbol{1}\end{bmatrix} < 0$$

which is a slightly stricter version of the first matrix inequality in (4.6). This means that for any $\alpha_i \in]0, 1-\rho_i^2[$ this linear matrix inequality in $\boldsymbol{Z}_i$ has at least one solution $\boldsymbol{Z}_i > 0$ which can then be put as a constant matrix into the second matrix inequality in (4.6) that then becomes a linear matrix inequality in β_i.

4.4 Complete Algorithm

This section summarizes the steps to carry out a multiobjective optimization based on the parametrization of fixed order controllers and the parametrization of observable pairs with objectives formulated in terms of the L1, L2/H2, H∞ norms. The proposed algorithm consists of the three stages initialization, optimization, finalization. An overview of this algorithm is presented in figure 4.3 while a more detailed description is provided below.

- Get realization of the plant $\mathbf{P}$, get observable realization of a working controller $\mathbf{K}_0$
- Calculate parameter $\mathbf{Q}$ from $\mathbf{K}_0$
- Define extended loops $\mathbf{H}_i$
- Define cost $f(\boldsymbol{\gamma})=\boldsymbol{\gamma}^T\mathbf{I}\boldsymbol{\gamma}$ with diagonal $\mathbf{I}\geq 0$, define semidefinite constraint $\mathbf{G}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\varepsilon},\mathbf{Z},\mathbf{Q})\leq 0$ in terms of L1,L2/H2,H∞ norms on the extended loops $\mathbf{H}_i$
- Initialize $\boldsymbol{\alpha},\boldsymbol{\beta}$ using the spectral radius $\boldsymbol{\rho}$ by solving L1 linear matrix inequalities
- Extract stable observable pair $(\mathbf{A}_{Q0},\mathbf{C}_{Q0})$ as initial chart center from parameter $\mathbf{Q}$, calculate Gram observability matrix $\mathbf{G}_0$ of pair $(\mathbf{A}_{Q0},\mathbf{C}_{Q0})$, find lossless completion $(\mathbf{B}_0,\mathbf{D}_0)$ of pair $(\mathbf{A}_{Q0},\mathbf{C}_{Q0})$, set $\mathbf{W}=\mathbf{0}$ and Gram observability matrix $\mathbf{G}=\mathbf{G}_0$

Repeat as long as cost f improvement is above threshold

- Obtain sensibility matrix $\boldsymbol{\Omega}^*$ by evaluation of Lagrange dual function for current parameter values $\mathbf{Q},\boldsymbol{\alpha},\boldsymbol{\beta}$
- Actualize the current parameter values $\mathbf{Q},\boldsymbol{\alpha},\boldsymbol{\beta}$ in direction provided by the subgradient of cost f as a function of the variations of $\mathbf{W},\mathbf{B}_Q,\mathbf{D}_Q$
- If smallest eigenvalue of $\mathbf{G}$ below threshold set current pair $(\mathbf{A}_Q,\mathbf{C}_Q)$ as new center $(\mathbf{A}_{Q0},\mathbf{C}_{Q0})$, balance it so that $\mathbf{G}_0=\mathbf{1}$, set $\mathbf{W}=\mathbf{0}$ and $\mathbf{G}=\mathbf{G}_0$

- Calculate locally optimized observable controller realization $\mathbf{K}^*$ from current parameter value $\mathbf{Q}$

Figure 4.3: Overview of the optimization algorithm.

4.4.1 Initialization

In the initialization stage, the optimization problem is set up for a given plant $\boldsymbol{P}$ that is to be optimally controlled by a controller $\boldsymbol{K}$ in a feedback loop $\boldsymbol{H}$. The optimization objectives are defined in form of a scalar cost function and a negative semidefinite matrix constraint function. An initial parameter $\boldsymbol{Q}$ is constructed from an admissible controller $\boldsymbol{K}_0$.

- Get a realization of the plant $\boldsymbol{P}$ and an observable realization of an initial working controller $\boldsymbol{K}_0$.

- Define filters $\boldsymbol{P}_{zw_i}, \boldsymbol{P}_{yw_i}, \boldsymbol{P}_{zu_i}$ and thereby formally construct the extended loops $\boldsymbol{H}_i$ as defined in section 4.1.1 for each optimization objective which allows to refine the objectives by selecting certain performance channels or by emphasizing certain frequencies.

- Define the optimization objectives in form of a scalar cost function $f(\boldsymbol{\gamma})$ of the form $f(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^T \boldsymbol{I} \boldsymbol{\gamma}$ with diagonal $\boldsymbol{I} > 0$ and a constraint function $\boldsymbol{G}'(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{K}) \leq 0$ formulated in terms of the L1, L2/H2, H∞ norms given in section 4.1.2 while assuring the initial working controller $\boldsymbol{K}_0$ to be admissible. It is admissible if there exist $\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta} \geq 0, \boldsymbol{\gamma} > 0, \boldsymbol{\epsilon} > 0, \boldsymbol{Z} > 0$ so that $\boldsymbol{G}'(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\epsilon}, \boldsymbol{Z}, \boldsymbol{K}_0) \leq 0$. The optimization problem is now set up and can be written in the form (4.11).

- Try to construct initial L1 parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$. It has been shown in section 4.3.4 that if the extended loop $\boldsymbol{H}_i$ is stable then this is possible by first selecting an initial $\alpha_i \in]0, 1 - \rho_i^2[$ then calculating the initial $\boldsymbol{Z}_i$ and finally the initial β_i, with ρ_i being the spectral radius of the extended loop dynamic matrix $\boldsymbol{A}_i$. Alternately fixing α_i, β_i and $\boldsymbol{Z}_i$ and solving the resulting linear matrix inequalities at each step in order to decrease the upper L1 norm bound γ_i should then lead to usable initial L1 parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$.

- Try to build $\boldsymbol{Q}$ from $\boldsymbol{K}_0$ using the relations provided in section 4.2.1. If equation (4.12) has a solution $\boldsymbol{R}$ then $\boldsymbol{Q}$ can be calculated by (4.13). It has been shown in section 4.3.1 that if the closed loop dynamic matrix $\boldsymbol{A}_H$ is diagonalizable then this is possible. Since $\boldsymbol{K}_0$ is required to be observable, also $\boldsymbol{Q} = \boldsymbol{Q}_L$ is observable due to (2.45). Further the extended loops $\boldsymbol{H}_i$ need to be stable for the L1, L2/H2, H∞ norms to remain bounded so that $\boldsymbol{K}_0$ is a stabilizing controller and thus $\boldsymbol{Q}$ is stable due to (2.46). Hence $\boldsymbol{Q}$ is stable and observable. The optimization problem is now ready for optimization and can be written in the form (4.18).

- Extract the stable observable pair $(\boldsymbol{A}_Q, \boldsymbol{C}_Q)$ from parameter $\boldsymbol{Q}$ and set it as the center $(\boldsymbol{A}_{Q_0} = \boldsymbol{A}_Q, \boldsymbol{C}_{Q_0} = \boldsymbol{C}_Q)$ of the chart for the parametrization of stable observable pairs. Reset $\boldsymbol{W} = \boldsymbol{0}$ and let $\boldsymbol{G}_0$ be the Gram observability matrix of the center stable observable pair $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$. Find a lossless completion pair $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ of the center $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$ as defined in (3.6).

4.4.2 Optimization

In the optimization stage, the parameter $\boldsymbol{Q}$ along with $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are improved, which means that they are moved from their initial positions so that the cost function f decreases and the constraint function $\boldsymbol{G}$ remains negative semidefinite. The whole optimization stage is repeated until the cost function reaches a local minimum.

- Evaluate the Lagrange dual function of the cost function f for the current parameter value $\boldsymbol{Q}$ and current values $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and obtain the local sensitivity matrix $\boldsymbol{\Omega}^*$ as described in section 4.2.4.

- Actualize the parameter $\boldsymbol{Q}$ and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ so that f decreases, the decreasing direction being indicated by the subgradient of the cost function f as described in section 4.2.5. Try to calculate the subgradient as a function of the variation of the parameter state space matrices $\boldsymbol{B}_Q, \boldsymbol{D}_Q$ and the stable observable pair parameter $\boldsymbol{W}$ as shown in the sections 4.2.6, 4.2.7, 4.2.8. It has been shown in section 4.3.2 that if the dynamic matrix $\boldsymbol{A}_Q$ of the stable and observable parameter $\boldsymbol{Q}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_P$ of the plant $\boldsymbol{P}$ then this is possible.

- If the condition $\boldsymbol{G} > 0$ is close to be violated, center on the current balanced stable observable pair $(\boldsymbol{A}_{Q_0} = \boldsymbol{J}\boldsymbol{A}_Q\boldsymbol{J}^{-1}, \boldsymbol{C}_{Q_0} = \boldsymbol{C}_Q\boldsymbol{J}^{-1})$ with $\boldsymbol{J}$ defined by $\boldsymbol{G} = \boldsymbol{J}^H\boldsymbol{J}$ and reset $\boldsymbol{W} = \boldsymbol{0}$ and $\boldsymbol{G}_0 = \boldsymbol{1}$. Find a lossless completion pair $(\boldsymbol{B}_0, \boldsymbol{D}_0)$ of the center $(\boldsymbol{A}_{Q_0}, \boldsymbol{C}_{Q_0})$ as defined in (3.6).

4.4.3 Finalization

In the finalization stage, an optimal controller $\boldsymbol{K}^*$ is reconstructed from the improved parameter $\boldsymbol{Q}$.

- Try to build $\boldsymbol{K}^*$ from $\boldsymbol{Q}$ using the relations provided in section 4.2.1. If equation (4.14) has a solution $\boldsymbol{R}$ then $\boldsymbol{K}^*$ can be calculated by (4.15). It has been shown in section 4.3.3 that if the dynamic matrix $\boldsymbol{A}_Q$ of the parameter $\boldsymbol{Q}$ has no eigenvalues in common with the dynamic matrix $\boldsymbol{A}_P$ of the plant $\boldsymbol{P}$ then this is possible.

Chapter 5

Conclusion

In this report, a way to improve an existing controller with respect to a multiobjective specification, written in terms of the L1, L2/H2, H∞ system norms, has been presented. Three key elements distinguish the proposed approach. First, a parametrization of fixed order controllers by a stable and observable controller parameter, based on the work in [13]. Second, a parametrization of all stable observable pairs that allows an infinitesimal adjustment of the controller parameter, based on the work in [2]. And third, the use of a sensitivity information to calculate the local subgradient of the optimization problem so that the controller parameter can be adjusted the right way, based on the work in [4].

Although many performance measures to quantify the specifications of a multiobjective optimization problem exist, the L1, L2/H2, H∞ system norms are possibly the most important ones. They have been chosen in this report due to their practical relevance and also to illustrate the scope of the proposed approach. For a given system, the H2 and H∞ norms stand exemplarily for linear matrix inequality constraints and the L1 norm stands exemplarily for a bilinear matrix inequality constraint. Despite these inherent differences, these constraints are treated in a unified way, as part of a more general semidefinite constraint. The necessity for the proposed approach to gain access to the sensitivity information currently limits the possible constraints to these three normes.

The proposed solution is based on the left controller parameter $\boldsymbol{Q}_L$ on which a parametrization of stable observable pairs is applied. However, the dual approach is also possible, based on the right controller parameter $\boldsymbol{Q}_R$ on which a parametrization of stable controllable pairs is applied. In practical applications one could imagine to use $\boldsymbol{Q}_L$ and the stable observable pair parametrization in case of $\boldsymbol{K}$ having more inputs than outputs since the output matrices $\boldsymbol{C}_K, \boldsymbol{C}_{QL}$ are then of smaller dimension than the input matrices $\boldsymbol{B}_K, \boldsymbol{B}_{QR}$ and thus the stable observable pair parametrization of $(\boldsymbol{A}_{QL}, \boldsymbol{C}_{QL})$ is then simpler than the stable controllable pair parametrization of $(\boldsymbol{A}_{QR}, \boldsymbol{B}_{QR})$ and to use $\boldsymbol{Q}_R$ and the stable controllable pair parametrization in the other case.

The Riccati equations (2.21), (2.22) may have multiple solutions $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ for a given plant $\boldsymbol{P}$ and controller $\boldsymbol{K}$. The implications on the proposed algorithm are yet unknown. It may be possible that the descending paths of the algorithm are different for different $\boldsymbol{R}_{KP}, \boldsymbol{R}_{PK}$ so that in the case of the de-

scending algorithm starting from a saddle point or a maximum of the cost function, the outcome of the local optimization may be totally different.

Appendix A

Algebra

A.1 Ring Structure

This section serves as the mathematical foundation on which the main text is built upon and provides in particular the coprime factorization and coprime completion that are the starting point in the search for parametrizations.

A.1.1 Rings and Units

This section introduces the algebraic structure of a ring and provides the laws that follow from the definitions. Some special rings are given and the notion of unit is defined.

Ring

A ring R is a set with two binary relations "+" called addition and "." called multiplication with multiplication having priority over addition. Every element of a ring has an inverse with respect to addition, but not all elements have inverses with respect to multiplication. A ring is defined by the properties

$$R + R = R \qquad \text{(completeness)} \tag{A.1}$$
$$\forall a, b, c \in R : a + (b + c) = (a + b) + c \qquad \text{(associativity)} \tag{A.2}$$
$$\forall a, b \in R : a + b = b + a \qquad \text{(commutativity)} \tag{A.3}$$
$$\forall a \in R : a + 0 = 0 + a = a \qquad \text{(identity)} \tag{A.4}$$
$$\forall a \in R : a + (-a) = (-a) + a = a - a = 0 \qquad \text{(inverse)} \tag{A.5}$$

$$RR = R \qquad \text{(completeness)} \tag{A.6}$$
$$\forall a, b, c \in R : a(bc) = (ab)c \qquad \text{(associativity)} \tag{A.7}$$
$$\forall a \in R : a1 = 1a = a \qquad \text{(identity)} \tag{A.8}$$

$$\forall a, b, c \in R : a(b + c) = ab + ac,\ (a + b)c = ac + bc \qquad \text{(distributivity)} \tag{A.9}$$

Remark 1 Multiplication with the additive identity

$$\forall a \in R: \ a0 = 0a = 0 \tag{A.10}$$

since $a = a1 = a(1+0) = a1 + a0 = a + a0 \Leftrightarrow a - a = 0 = a0$ and by analogy $a = 1a$ leads to $0 = 0a$.

Remark 2 Multiplication with an additive inverse

$$\forall a, b \in R: \ a(-b) = (-a)b = -ab \tag{A.11}$$

since $ab + a(-b) = a[b + (-b)] = a0 = 0 \Leftrightarrow a(-b) = -ab$ and by analogy $ab + (-a)b$ leads to $(-a)b = -ab$.

Trivial Ring

The ring $R = \{0 = 1\}$ is called the *trivial ring*. In a nontrivial ring R that has other elements $a \in R: \ a \neq 0$ we always have

$$0 \neq 1 \Leftrightarrow \text{nontrivial ring} \tag{A.12}$$

Proof Let $0 = 1$ and take an arbitrary $a \in R$ then $a = a1 = a0 = 0$ and thus $a = 0$ for all $a \in R$ so that R must be the trivial ring.

Commutative Ring

If in addition to laws (A.1) to (A.9) the elements of ring R further obey

$$\forall a, b \in R: \ ab = ba \tag{A.13}$$

then R is said to be a *commutative ring*.

Integral Ring

If in addition to laws (A.1) to (A.9) there is no solution to

$$\forall a, b \in R \setminus \{0\}: \ ab = 0 \tag{A.14}$$

then R has no zero divisors and is said to be an *integral ring*.

Remark On an integral ring, the equation $ax = a$ with $x \in R, a \in R \setminus \{0\}$ has a unique solution $x = 1$ since $ax - a = a(x-1) = 0$ and $a \neq 0$.

Unit

If an element u in a ring R is invertible with respect to multiplication, then it is called a unit and obeys the rule

$$\forall u \in R: \ uu^{-1} = u^{-1}u = 1 \tag{A.15}$$

This leads to a special property for units $u \in R$

$$Ru = uR = R \tag{A.16}$$

Proof since on the one hand $Ru \subset R$ and $uR \subset R$ and on the other hand $R = R1 = Ru^{-1}u \subset Ru$ and $R = 1R = uu^{-1}R \subset uR$ and therefore being subsets of each other the two sets are equal and $Ru = uR = R$. This is in contrast to $Rr \subset R$, $rR \subset R$ for an arbitrary $r \in R$ due to multiplicative completeness (A.6).

Field

If every nonzero element $r \neq 0$ in a ring R is a unit then the ring R is called a *field*.

A.1.2 Ideals and Equivalence

This section introduces important subsets of a ring, the ideals. A measure for the size of ideals is given and the smallest ideals containing a number of elements are calculated. Finally, the notion of equivalence on a ring is defined and its relation to ideals is shown.

Ideal

An ideal I is an additive subgroup of a ring R that is invariant under multiplication by R. Since multiplication on a ring is not necessarily commutative, there exist left and right ideals. On a commutative ring they are both equal. A left ideal I is defined by the properties

$$I \subset R \qquad \text{(subset)} \tag{A.17}$$

$$I - I = I \qquad \text{(additive subgroup)} \tag{A.18}$$

$$RI = I \qquad \text{(multiplicative invariance)} \tag{A.19}$$

For a right ideal, equation (A.19) changes to $IR = I$. Equation (A.17) assures that I inherits the laws of the ring except completeness (A.1) and (A.6). Furthermore, since I is a subset, it not necessarily contains the identities 0 and 1 and the additive inverse $-a$ of an element $a \in I$. Equation (A.18) assures additive completeness of I and provides the ideal with the additive identity 0 and the additive inverse $-a$ of an element $a \in I$. With these two equations I becomes an additive subgroup of R. Equation (A.19) provides the multiplicative invariance property and since $I \subset R \Rightarrow II \subset RI = I$ assures multiplicative completeness.

Remark That means an ideal I is almost a ring. If I also contained the multiplicative identity 1, then I would indeed be a ring. Equation (A.20) will show that this ring is R.

Ideal containing a Unit

If any ideal I contains a unit $u \in R$ then the ideal is the ring itself

$$I = R \Leftrightarrow u \in I \tag{A.20}$$

Proof Since $R = Ru = uR \subset I$ using (A.16) and (A.19) and $I \subset R$ due to (A.17) and therefore being subsets of each other the two sets are equal and $I = R$. The ideals 0 and R are called *trivial ideals*.

Intersection of Ideals

The intersection of two same sided ideals is again an ideal of the same side

$$I = I_1 \cap I_2 \tag{A.21}$$

Proof This is because first $I_1 \subset R$ and $I_2 \subset R$ so $I = I_1 \cap I_2 \subset R$ which verifies (A.17), second $I_1 - I_1 = I_1$ and $I_2 - I_2 = I_2$ so $I - I \subset I_1 - I_1 = I_1$ and $I - I \subset I_2 - I_2 = I_2$ so that $I - I \subset I_1 \cap I_2 = I$ and thus $I - I = I$ which verifies (A.18) and third $RI_1 = I_1$ and $RI_2 = I_2$ so $RI \subset RI_1 = I_1$ and $RI \subset RI_2 = I_2$ so that $RI \subset I_1 \cap I_2 = I$ and with $1 \in R$ this gives $RI = I$ which verifies (A.19). The same is true for right ideals.

Sum of Ideals

The sum of two same sided ideals is again an ideal of the same side

$$I = I_1 + I_2 \tag{A.22}$$

Proof This is because first $I_1 \subset R$ and $I_2 \subset R$ so $I = I_1 + I_2 \subset R + R = R$ which verifies (A.17), second $I_1 - I_1 = I_1$ and $I_2 - I_2 = I_2$ so $I - I = (I_1 + I_2) - (I_1 + I_2) = (I_1 - I_1) + (I_2 - I_2) = I_1 + I_2 = I$ which verifies (A.18) and third $RI_1 = I_1$ and $RI_2 = I_2$ so $RI \subset RI_1 + RI_2 = I_1 + I_2 = I$ and with $1 \in R$ this gives $RI = I$ which verifies (A.19). The same is true for right ideals.

Principal Ideal

A principal ideal $I(x)$ is an ideal with one additional property

$$Rx = I(x) \qquad \text{(generation)} \tag{A.23}$$

This defines a left principal ideal. For a right principal ideal, this equation changes to $xR = I(x)$. This equation means that there is one element $x \in I(x)$ that generates the ideal. A principal ideal is therefore the set of all multiples of x by ring elements.

Size of an Ideal

Two same sided ideals can be compared in size with respect to mutual inclusion

$$I_1 \subset I_2 \Leftrightarrow I_1 \text{ smaller than } I_2 \tag{A.24}$$

and an ideal $I \neq R$ is said to be *maximal* if it is not contained in any other ideal of the same side.

Smallest Ideal containing two Ideals

The smallest ideal that contains two ideals $I_1, I_2 \subset R$ is the sum of the same sided ideals $I_1 + I_2$ generated by these ideals

$$I_1 + I_2 \text{ is the smallest ideal that contains } I_1, I_2 \tag{A.25}$$

Proof Let I be the smallest ideal that contains I_1, I_2, thus $I \subset I_1 + I_2$. But I as an ideal verifies (A.18) and since $I_1 \subset I$ and $I_2 \subset I$ this leads to $I_1 + I_2 \subset I$ and therefore being subsets of each other the two sets are equal and $I = I_1 + I_2$.

Smallest Ideal containing one Element

The smallest ideal that contains an element $a \in R$ is the principal ideal $I(a)$ generated by this element

$$I(a) \text{ is the smallest ideal that contains } a \tag{A.26}$$

Proof Let I be the smallest ideal that contains a, thus $I \subset I(a)$. But for left ideals $I(a) = Ra \subset RI = I$ and therefore being subsets of each other the two sets are equal and $I = I(a)$. The same is true for right ideals.

Smallest Ideal containing two Elements

The smallest ideal that contains two elements $a, b \in R$ is the sum of the same sided principal ideals $I(a) + I(b)$ generated by these elements

$$I(a) + I(b) \text{ is the smallest ideal that contains } a, b \tag{A.27}$$

Proof Let I be the smallest ideal that contains a, b, thus $I \subset I(a) + I(b)$. But for left ideals $I(a) = Ra \subset RI = I$ and $I(b) = Rb \subset RI = I$ with (A.18) leads to $I(a) + I(b) \subset I$ and therefore being subsets of each other the two sets are equal and $I = I(a) + I(b)$. The same is true for right ideals.

Equivalence

Two elements $x, y \in R$ are said to be left equivalent if there exists a unit $u \in R$ for which $x = uy$. By analogy, they are right equivalent if there exists a unit $v \in R$ for which $x = yv$. Two equivalent elements generate the same principal ideal

$$I(x) = I(y) \Leftrightarrow x \sim y \qquad \text{(equivalence)} \tag{A.28}$$

Proof Left equivalence $x \sim y \Leftrightarrow x = uy$ leads to $I(x) = Rx = Ruy = Ry = I(y)$. To verify the reverse direction, let $x \nsim y$. If $\exists a \in I(y) : a \notin I(x)$ then clearly $I(x) \neq I(y)$. If this is not the case then $\forall a \in I(y) : a \in I(x)$ and therefore $I(y) \subset I(x) \Rightarrow y \in Rx$. So $\exists m \in R : y = mx$, but m is not a unit since $x \nsim y$ so that $I(m) \neq R$. Hence $\exists p \in R : p \notin I(m)$ and thus $\nexists q \in R : p = qm$. This allows the construction of $a = px$ so that $a \in I(x)$ but since $\nexists q \in R : p = qm$ this leads to $\nexists q \in R : a = px = qmx = qy$. This means that $a \notin I(y)$ so indeed $\exists a \in I(x) : a \notin I(y)$ and therefore $x \nsim y \Rightarrow I(x) \neq I(y)$ which is equivalent to $x \sim y \Leftarrow I(x) = I(y)$. The same is true for right equivalence.

A.1.3 Division

This section introduces the notion of divisibility on a general ring, using the concept of ideals. It is shown how the least common multiple and the greatest common divisor of two ring elements can be found.

Divisibility

Since a principal ideal is the set of all multiples of its generator, any element of the ideal can be written as a product of the generator and a ring element. This introduces the notion of divisibility

$$a \in I(b) \Leftrightarrow a \text{ can be divided by } b \qquad \text{(divisibility)} \tag{A.29}$$

Proof Let $I(b)$ be a left ideal then $a \in I(b) = Rb$ and hence $\exists m \in R: \ a = mb$ and thus b is a right divisor of a. If it is a right ideal then, by analogy, its generator is a left divisor.

Remark A unit $u \in R$ is a divisor to every element of the ring since following equation (A.20) any ideal containing a unit is the ring itself. Especially $I(u) = R$ and thus every ring element is contained in this ideal and therefore divisible by its generator u.

Least Common Multiple

If the intersection of two principal ideals is itself a principal ideal, then the least common multiples are defined by

$$\text{The generators of } I(a) \cap I(b) \text{ are the least common multiples of } a, b \tag{A.30}$$

Proof If the intersection of two principal ideals $I(a) \cap I(b)$ is itself a principal ideal, then we have $I(a) \cap I(b) = I(m)$. Since $I(m) \subset I(a), I(b)$ and especially $m \in I(a), I(b)$ divisibility (A.29) shows that m can be divided by a, b or in other words m is a common multiple of a, b. If m' is another common multiple of a, b of the same side as m, then $m' \in I(a), I(b)$ so that $m' \in I(a) \cap I(b) = I(m)$. This means m' is itself a multiple of m and therefore m is the least common multiple if $I(a) \cap I(b)$ is principal.

Remark Due to (A.28) all generators of $I(a) \cap I(b)$ and thus all least common multiples are equivalent.

Greatest Common Divisor

If the sum of two principal ideals is itself a principal ideal, then the greatest common divisors are defined by

$$\text{The generators of } I(a) + I(b) \text{ are the greatest common divisors of } a, b \tag{A.31}$$

Proof If the sum of two principal ideals $I(a) + I(b)$ is itself a principal ideal, then we have $I(a)+I(b) = I(d)$. Since $I(a), I(b) \subset I(d)$ and especially $a, b \in I(d)$ divisibility (A.29) shows that d is a common divisor of a and b. If d' is another common divisor of a, b of the same side as d, then $a, b \in I(d')$. But following (A.27) the ideal $I(d) = I(a) + I(b)$ is the smallest ideal containing a, b, we have $I(d) \subset I(d')$ and thus d can be divided by d' and therefore d is the greatest common divisor if $I(a) + I(b)$ is principal.

Remark Due to (A.28) all generators of $I(a) + I(b)$ and thus all greatest common divisors are equivalent.

Coprime Elements

Two elements $a, b \in R$ are called *coprime* if their greatest common divisor is a unit $u \in R$

$$I(a) + I(b) = R \Leftrightarrow a, b \text{ are coprime} \tag{A.32}$$

since $R = I(u)$. For a greatest common right divisor, they are right coprime with $I(a)$ and $I(b)$ being left ideals. They are left coprime in the other case.

Coprime Quotients

Since the greatest common divisor d of a, b contains all other common divisors, a, b divided by d do not share any common divisors and thus the quotients are coprime

$$\begin{aligned} &a = md \text{ and } b = nd \Rightarrow m, n \text{ are right coprime} \\ &\text{if } d \text{ is the greatest common right divisor of } a, b \end{aligned} \tag{A.33}$$

If it is the greatest common left divisor, then we have $a = dm$ and $b = dn$ with m, n left coprime.

Proof Let m, n have a greatest common right divisor z so there are $v, w \in R$ with $m = vz$ and $n = wz$ so that $a = md = v(zd)$ and $b = nd = w(zd)$. So zd is a common right divisor of a, b so $a, b \in I(zd)$ and because (A.27) and (A.31) show that $I(d)$ is the smallest ideal to contain a, b thus $I(d) \subset I(zd)$. But $I(zd) = Rzd \subset RRd = Rd = I(d)$ and therefore being subsets of each other the two sets are equal and $I(d) = I(zd) \Leftrightarrow d \sim zd$ using (A.28). Therefore z must be a unit and thus m, n are right coprime. The same is true for left division and left coprimeness.

Bezout Identity

The greatest common divisor d of $a, b \in R$ is a generator of $I(a) + I(b)$ and thus $d \in I(a) + I(b)$. If the ideals are left sided then $d \in Ra + Rb$ and therefore

$$\exists x, y \in R: \; xa + yb = d, \text{ the greatest common right divisor of } a, b \tag{A.34}$$

which is called the *Bezout identity*. For right ideals and left divisors, this equation changes to $ax + by = d$. On integral rings, the Bezout identity (A.34) can be reduced to

$$xv + yw = 1 \text{ with } x, y \text{ left coprime and } v, w \text{ right coprime} \tag{A.35}$$

Proof According to (A.33) there are right coprime $v, w \in R$ with $a = vd$ and $b = wd$ so that $xvd + ywd = d$ leads to $xv + yw = 1$ on an integral ring. Let x, y have a greatest common left divisor z so there are $m, n \in R$ with $x = zm$ and $y = zn$ so that $zmv + znw = 1$. By $z(mv + nw)z = z$ we have $z(mv+nw) = (mv+nw)z = 1$ since the ring is integral. So the greatest common left divisor z is a unit and therefore x, y are left coprime.

A.1.4 Factorization

This section develops the factorization of elements on a special case of a ring, the principal commutative ring. On such a ring, it is shown that the factorization is essentially unique.

Principal Ring

A principal ring is a ring where every ideal is principal. So if in addition to laws (A.1) to (A.9) the elements and ideals of R further obey

$$\forall I \subset R, \exists x \in R : \; I = I(x) \tag{A.36}$$

then R is said to be a *principal ring*.

Irreductible Elements

Every element $p \in R$ is at least contained in two principal ideals, namely the ideal $I(p)$ generated by p itself and the trivial ideal $I(1) = I(u) = R$. If p is not contained in any other principal ideal, then p is said to be *irreductible*. Hence, irreductible elements are only divisible by itself and the units $u \in R$. On a principal ring R the following equivalence holds

$$I(p) \text{ is maximal} \Leftrightarrow p \text{ is irreductible} \tag{A.37}$$

Proof If on the one hand $p \in R$ is not irreductible, there is a $q \in R$ that divides p so that $p \in I(q)$ and using (A.26) this means $I(p) \subset I(q)$ and thus $I(p)$ is not maximal. If on the other hand $I(p)$ is not maximal then there is an ideal I so that $I(p) \subset I$ and since the ring is principal, I has a generator $q \in R$ so that $I(p) \subset I(q)$ and p can be divided by q and thus p is not irreductible.

Irreductibility and Coprimeness

For any element $x \in R$, the following equivalence holds if $p \in R$ is irreductible:

$$I(x) + I(p) = R \Leftrightarrow x \notin I(p) \tag{A.38}$$

with the two ideals being same sided. In other words, if an irreductible $p \in R$ does not divide $x \in R$, then with (A.32) x, p are coprime and their greatest common divisor is a unit $u \in R$.

Proof If on the one hand $x \in I(p)$ then using (A.26) this leads to $I(x) \subset I(p)$ and thus $I(x) + I(p) = I(p) \neq R$ so that $I(x) + I(p) = R \Rightarrow x \notin I(p)$. If on the other hand $I(x) + I(p) \neq R$ then clearly $I(p) \subset I(x) + I(p)$ and because p is irreductible and with (A.37) $I(p)$ is maximal, this gives $I(x) + I(p) = I(p)$ and thus $I(x) \subset I(p)$ and especially $x \in I(p)$ so that $I(x) + I(p) = R \Leftarrow x \notin I(p)$.

Irreductibility and Divisibility

If a product ab of elements of a commutative ring $a, b \in R$ is not divisible by an irreductible $p \in R$, then neither of the factors a, b is divisible by p

$$ab = ba \notin I(p) \Leftrightarrow a, b \notin I(p) \tag{A.39}$$

Proof If $a \in I(p)$ then with (A.26) this gives $I(a) \subset I(p)$ and because $I(ab) = Rab \subset Ra = I(a)$ we have $I(ab) \subset I(p)$ hence $ab \in I(p)$ and thus $ab \notin I(p) \Rightarrow a, b \notin I(p)$. Now let $ab \in I(p)$, then at least one of the factors a, b must be divisible by p to verify the reverse direction. If $a \notin I(p)$ then with (A.38) this gives $I(a) + I(p) = R$ and can be written as $\exists x, y \in R : xa + yb = 1$. Multiplying this by b leads to $xab + ypb = b$ and since $ab \in I(p)$ can be written as $\exists z \in R : ab = zp$ this gives $xzp + ypb = b$ so that $(xz + yb)p = b$ and thus $b \in I(p)$ which verifies $ab \notin I(p) \Leftarrow a, b \notin I(p)$.

Factorization on Principal Commutative Rings

Any element of a principal commutative ring $x \in R$ has an essentially unique factorization in irreductible elements

$$x = u \prod_{k=1}^{n} p_k \text{ where } u \in R \text{ is a unit} \tag{A.40}$$

Proof This factorization is due to an algorithm that reduces x to a unit u. If x is not a unit, there must be a maximal ideal $I(p_1)$, which is principal since R is principal, with $x \in I(p_1)$ so that $\exists x_1 \in R : x = x_1 p_1$. If x_1 is not a unit, there must be a maximal ideal $I(p_2)$ with $x_1 \in I(p_2)$ so that $\exists x_2 \in R : x_1 = x_2 p_2$ and so on, until $x_n = u$ is a unit. Since the ideals $I(p_k)$ are maximal, with (A.37) the p_k are irreductible. It can be proven by Zorn's lemma that this algorithm and hence n in (A.40) is finite.

Now let $x = u \prod_{k=1}^{m} p_k = v \prod_{k=1}^{n} q_k$ be two factorizations of x. Since a q_j divides x we have $x = u \prod_{k=1}^{m} p_k \in I(q_j)$ and using (A.39) there is at least one p_i with $p_i \in I(q_j)$. Using (A.26) this gives $I(p_i) \subset I(q_j)$ but $I(p_i)$ is maximal so that $I(p_i) = I(q_j)$ and with (A.28) the two generators are equivalent $p_i \sim q_j$. So for every p_i there is an equivalent q_j and thus the number of irreductible factors is the same $m = n$ and the factorization (A.40) is essentially unique, with the word essentially referring to the possibility of exchanging a factor by an equivalent factor.

A.1.5 Extension to Matrices

This section treats matrices that are built from elements of a principal commutative integral ring.

Unit Matrices

A square matrix $\boldsymbol{U}$ built from elements of a principal commutative ring R is a unit if and only if its determinant is a unit in R

$$\boldsymbol{U} \text{ is a unit in } R^{n,n} \Leftrightarrow \det \boldsymbol{U} \text{ is a unit in } R \tag{A.41}$$

Proof since on the one hand if $\boldsymbol{U}$ is a unit in $R^{n,n}$ then by definition (A.15) there exists a $\boldsymbol{U}^{-1}$ with $\boldsymbol{U}\boldsymbol{U}^{-1} = \boldsymbol{U}^{-1}\boldsymbol{U} = \boldsymbol{1}$ where $\boldsymbol{1}$ denotes the $R^{n,n}$ identity matrix. Therefore $\det(\boldsymbol{U}\boldsymbol{U}^{-1}) = \det \boldsymbol{U} \det \boldsymbol{U}^{-1} = \det(\boldsymbol{U}^{-1}\boldsymbol{U}) = \det \boldsymbol{U}^{-1} \det \boldsymbol{U} = \det \boldsymbol{1} = 1$ shows that $\det \boldsymbol{U}$ is a unit in R. If on the other hand $\det \boldsymbol{U}$ is a unit in R then $(\det \boldsymbol{U})^{-1} \in R$. Cramer's theorem states that

for every matrix $\boldsymbol{M} \in R^{n,n}$ there is a unique adjoint matrix $\operatorname{adj} \boldsymbol{M} \in R^{n,n}$ that satisfies $\boldsymbol{M} \operatorname{adj} \boldsymbol{M} = \operatorname{adj} \boldsymbol{M} \boldsymbol{M} = \mathbf{1} \det \boldsymbol{M}$. Since $(\det \boldsymbol{U})^{-1} \in R$, the matrix $\operatorname{adj} \boldsymbol{U} (\det \boldsymbol{U})^{-1} \in R^{n,n}$ is a valid inverse of $\boldsymbol{U}$ which therefore is a unit in $R^{n,n}$.

Bezout Identity for two Elements

The relations among two elements of a principal commutative integral ring $a, b \in R$ and their greatest common divisor $d \in R$ can be brought into the matrix representation of the Bezout identity

$$\exists \text{ unit } \boldsymbol{U} \in R^{2,2} : \ \boldsymbol{U} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix} \tag{A.42}$$

with d being the greatest common divisor of $\{a, b\}$

Proof Since R is integral, the Bezout identity (A.34) can be reduced to (A.35), so for any two elements $a, b \in R$ with greatest common divisor $d \in R$ there exist $v, w, x, y \in R$ so that $xa + yb = d, xv + yw = 1, wa = vb$ with the third equation stemming from the fact that d is a common divisor of a, b so that $a = vd, b = wd$ hence $wa = wvd = vwd = vb$ since R is commutative. These three equations can then be brought into the matrix form (A.42)

$$\begin{bmatrix} x & y \\ \mp w & \pm v \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix} \text{ with } \det \begin{bmatrix} x & y \\ \mp w & \pm v \end{bmatrix} = \pm xv \pm yw = \pm 1$$

Because of (A.41) the matrix $\boldsymbol{U}$ is a unit in $R^{2,2}$ since $\det \boldsymbol{U} = \pm 1$ is a unit in R.

Bezout Identity for multiple Elements

This form of the Bezout identity can be generalized to a relation between a set of elements $\{a_1, a_2, \ldots, a_n\} \subset R$ and their greatest common divisor $d \in R$

$$\exists \text{ unit } \boldsymbol{U} \in R^{n,n} : \ \boldsymbol{U} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} d \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{A.43}$$

with d being the greatest common divisor of $\{a_1, a_2, \ldots, a_n\}$

Proof This is proven by induction. Following (A.42) this equation is correct for $n = 2$. Suppose (A.43) correct for $n - 1$ so that there exists a unit $\boldsymbol{U} \in R^{n-1,n-1}$ with

$$\boldsymbol{U} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} d \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U} \end{bmatrix} \begin{bmatrix} a_n \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} a_n \\ d \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where d is the greatest common divisor of $a_1, a_2, \ldots, a_{n-1}$. In order to get from $n - 1$ to n, equation (A.42) may be used for the greatest common divisor d' of

elements a_n, d and thus the greatest common divisor of all $a_1, a_2, \ldots, a_n$ so that there exists a unit $\boldsymbol{V} \in R^{2,2}$ with:

$$\boldsymbol{V}\begin{bmatrix} a_n \\ d \end{bmatrix} = \begin{bmatrix} d' \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n-2} \end{bmatrix} \begin{bmatrix} a_n \\ d \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} d' \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Hence the equation can be formulated for n:

$$\begin{bmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n-2} \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \begin{bmatrix} a_n \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n-2} \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \begin{bmatrix} \boldsymbol{0} & 1 \\ \boldsymbol{1}_{n-1} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{U}'} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} d' \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

It is now left to verify that $\boldsymbol{U}'$ is still a unit in $R^{n,n}$ what according to (A.41) is equal to $\det \boldsymbol{U}'$ being a unit in R:

$$\det \boldsymbol{U}' = \det \begin{bmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{n-2} \end{bmatrix} \det \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \det \begin{bmatrix} \boldsymbol{0} & 1 \\ \boldsymbol{1}_{n-1} & \boldsymbol{0} \end{bmatrix} = (-1)^{n+1} \det \boldsymbol{V} \det \boldsymbol{U}$$

Since $(-1)^{n+1} = \pm 1$ is a unit in R and $\det \boldsymbol{U}, \det \boldsymbol{V}$ are units in R, $\det \boldsymbol{U}'$ is also a unit in R and thus the matrix $\boldsymbol{U}'$ is a unit in $R^{n,n}$.

Equivalence to Triangular Matrix

Any matrix $\boldsymbol{M} \in R^{m,n}$ is left equivalent to an upper triangular matrix $\boldsymbol{T} \in R^{m,n}$ that has zeros below the diagonal

$$\exists \text{ unit } \boldsymbol{U} : \ \boldsymbol{U}\boldsymbol{M} = \boldsymbol{T} \text{ upper triangular} \tag{A.44}$$

For right equivalence this equation changes to $\boldsymbol{M}^T\boldsymbol{U}^T = \boldsymbol{T}^T$ when transposed, so that any matrix $\boldsymbol{M}^T$ is right equivalent to $\boldsymbol{T}^T$ which is a lower triangular matrix with zeros above the diagonal.

Proof This is proven by induction. For $m = 1$ matrix $\boldsymbol{M}$ is a single line and thus already an upper triangular matrix. For $n = 1$ matrix $\boldsymbol{M}$ is a single column that according to (A.43) is left equivalent to an upper triangular matrix. Suppose (A.44) correct for $m-1, n-1$ so that there exists a unit $\boldsymbol{U} \in R^{m-1,m-1}$ with

$$\boldsymbol{U}\boldsymbol{M} = \boldsymbol{T} \text{ upper triangular}$$

For m, n the first column of a matrix $\boldsymbol{M}' \in R^{m,n}$ can be brought in upper triangular form using (A.43):

$$\boldsymbol{V}\boldsymbol{M}' = \begin{bmatrix} d & \boldsymbol{m} \\ \boldsymbol{0} & \boldsymbol{M} \end{bmatrix} \Rightarrow \underbrace{\begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} \end{bmatrix} \boldsymbol{V}}_{\boldsymbol{U}'} \boldsymbol{M}' = \begin{bmatrix} d & \boldsymbol{m} \\ \boldsymbol{0} & \boldsymbol{U}\boldsymbol{M} \end{bmatrix} = \underbrace{\begin{bmatrix} d & \boldsymbol{m} \\ \boldsymbol{0} & \boldsymbol{T} \end{bmatrix}}_{\boldsymbol{T}'}$$

with $\boldsymbol{V}$ being a unit and $\det \boldsymbol{U}' = \det \boldsymbol{U} \det \boldsymbol{V}$ shows that $\boldsymbol{U}'$ is a unit in $R^{m,m}$ so that $\boldsymbol{M}'$ is equivalent to the resulting upper triangular matrix $\boldsymbol{T}'$ and thus (A.44) is valid for m, n. It is interesting to note that due to (A.43) the entry d in $\boldsymbol{T}'$ is the greatest common divisor of the first column of $\boldsymbol{M}'$.

Equivalence to Diagonal Matrix

Any matrix $\boldsymbol{M} \in R^{m,n}$ is equivalent to a diagonal matrix $\boldsymbol{D} \in R^{m,n}$ that has zeros below and above the diagonal

$$\exists \text{ units } \boldsymbol{U}, \boldsymbol{V} : \ \boldsymbol{UMV} = \boldsymbol{D} \text{ diagonal} \tag{A.45}$$

Proof This is proven by induction. For $m = 1$ matrix $\boldsymbol{M}$ is a single line and thus by (A.43) already equivalent to the diagonal matrix $\begin{bmatrix} d & \mathbf{0} \end{bmatrix}$. For $n = 1$ matrix $\boldsymbol{M}$ is a single column and thus the transpose of a single line and thus also equivalent to a diagonal matrix. Suppose (A.45) correct for $m-1, n-1$ so that there exist units $\boldsymbol{U} \in R^{m-1,m-1}$ and $\boldsymbol{V} \in R^{n-1,n-1}$ with

$$\boldsymbol{UMV} = \boldsymbol{D} \text{ diagonal}$$

For m, n the first column of a matrix $\boldsymbol{M'} \in R^{m,n}$ can be brought in upper triangular form using (A.43):

$$\boldsymbol{U}_1\boldsymbol{M'} = \begin{bmatrix} d_1 & \boldsymbol{m}_1 \\ \mathbf{0} & \boldsymbol{M}_1 \end{bmatrix}$$

with $\boldsymbol{U}_1$ being a unit. If d_1 divides all elements of vector $\boldsymbol{m}_1$ so that $\boldsymbol{m}_1 = \boldsymbol{n}_1 d_1$ then

$$\boldsymbol{U}_1\boldsymbol{M'} \underbrace{\begin{bmatrix} 1 & -\boldsymbol{n}_1 \\ \mathbf{0} & \mathbf{1} \end{bmatrix}}_{\boldsymbol{V}_1} = \begin{bmatrix} d_1 & \boldsymbol{m}_1 \\ \mathbf{0} & \boldsymbol{M}_1 \end{bmatrix} \begin{bmatrix} 1 & -\boldsymbol{n}_1 \\ \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} d_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_1 \end{bmatrix}$$

with $\boldsymbol{V}_1$ being a unit since $\det \boldsymbol{V}_1 = 1$. If d_1 does not divide all elements of vector $\boldsymbol{m}_1$ so that $\boldsymbol{m}_1 \neq \boldsymbol{n}_1 d_1$ then again, using (A.44), there exists a unit $\boldsymbol{V}_2 \in R^{n,n}$ with

$$\boldsymbol{V}_2(\boldsymbol{U}_1\boldsymbol{M'})^T = \begin{bmatrix} d_2 & \boldsymbol{m}_2 \\ \mathbf{0} & \boldsymbol{M}_2 \end{bmatrix}$$

where d_2 is the greatest common divisor of the first column of $(\boldsymbol{U}_1\boldsymbol{M'})^T$ thus there exists a $r_2 \in R$ so that $d_1 = r_2 d_2$. Here the algorithm repeats until eventually a d_k divides all elements of $\boldsymbol{m}_k$ so that $\boldsymbol{m}_k = \boldsymbol{n}_k d_k$. It can be proven by Zorn's lemma that this algorithm and hence k is finite so that there exist units $\boldsymbol{U'} \in R^{m,m}$ and $\boldsymbol{V'} \in R^{n,n}$ with

$$\boldsymbol{U'M'V'} = \begin{bmatrix} d & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M} \end{bmatrix}$$

Since (A.45) is supposed to be true for $m-1, n-1$, this leads to

$$\underbrace{\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U} \end{bmatrix} \boldsymbol{U'}}_{\boldsymbol{U''}} \boldsymbol{M'} \underbrace{\boldsymbol{V'} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V} \end{bmatrix}}_{\boldsymbol{V''}} = \begin{bmatrix} d & \mathbf{0} \\ \mathbf{0} & \boldsymbol{UMV} \end{bmatrix} = \underbrace{\begin{bmatrix} d & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix}}_{\boldsymbol{D'}}$$

with $\det \boldsymbol{U''} = \det \boldsymbol{U} \det \boldsymbol{U'}$ and $\det \boldsymbol{V''} = \det \boldsymbol{V} \det \boldsymbol{V'}$ shows that $\boldsymbol{U''}$ and $\boldsymbol{V''}$ are units in $R^{m,m}$ and $R^{n,n}$ respectively so that $\boldsymbol{M'}$ is equivalent to the resulting diagonal matrix $\boldsymbol{D'}$ and thus (A.45) is valid for m, n.

Smith Normal Form

Any matrix of elements of a principal commutative integral ring $\boldsymbol{M} \in R^{m,n}$ is equivalent to a diagonal matrix $\boldsymbol{D} \in R^{m,n}$ that has zeros below and above the diagonal and whose ith diagonal element $[\boldsymbol{D}]_{ii} = d_i$ is a divisor of all lower diagonal elements $d_{i+\mathbb{N}}$

$$\begin{aligned} &\exists \text{ units } \boldsymbol{U}, \boldsymbol{V} : \ \boldsymbol{UMV} = \boldsymbol{D} \text{ diagonal} \\ &\text{with } [\boldsymbol{D}]_{ii} = d_i \text{ being a divisor of } d_{i+\mathbb{N}} \end{aligned} \tag{A.46}$$

Proof Due to (A.45) matrix $\boldsymbol{M}$ is equivalent to a diagonal matrix $\boldsymbol{D}$. It is left to verify the sequential divisibility of its elements. This is proven by induction. For $m = 1$ or $n = 1$, matrix $\boldsymbol{M}$ is equivalent to a diagonal matrix with only one diagonal element hence (A.46) is true. Suppose (A.46) correct for $m-1, n-1$ so that there exist units $\boldsymbol{U} \in R^{m-1,m-1}$ and $\boldsymbol{V} \in R^{n-1,n-1}$ with

$$\boldsymbol{UMV} = \boldsymbol{D} \text{ with } [\boldsymbol{D}]_{ii} = d_i \text{ being a divisor of } d_{i+\mathbb{N}}$$

For m, n, using (A.45), there exist units $\boldsymbol{U}' \in R^{m,m}$ and $\boldsymbol{V}' \in R^{n,n}$ so that matrix $\boldsymbol{M} \in R^{m,n}$ is equivalent to diagonal matrix $\boldsymbol{D}' \in R^{m,n}$:

$$\boldsymbol{U}'\boldsymbol{M}'\boldsymbol{V}' = \begin{bmatrix} d & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix}$$

By using another unit $\boldsymbol{V}'' \in R^{n,n}$, matrix $\boldsymbol{D}'$ can be brought into the form

$$\boldsymbol{D}' \underbrace{\begin{bmatrix} 1 & \mathbf{0} \\ \boldsymbol{v} & \mathbf{1} \end{bmatrix}}_{\boldsymbol{V}''} = \begin{bmatrix} d & \mathbf{0} \\ \boldsymbol{Dv} & \boldsymbol{D} \end{bmatrix}$$

with a vector $\boldsymbol{v} \in R^{n-1}$ having identity elements $[\boldsymbol{v}]_k = 1$. Since $\boldsymbol{D}$ is diagonal, $\boldsymbol{Dv}$ gives

$$\boldsymbol{Dv} = \begin{bmatrix} d_1 \\ \vdots \\ d_l \end{bmatrix} \text{ with } l = \min(m-1, n-1)$$

Following (A.43) there exists a unit $\boldsymbol{U}'' \in R^{m,m}$ that reduces the first column of $\boldsymbol{D}'\boldsymbol{V}''$ to upper triangular form:

$$\underbrace{\begin{bmatrix} u_1 & \boldsymbol{u}_2 \\ \boldsymbol{u}_3 & \boldsymbol{U}_4 \end{bmatrix}}_{\boldsymbol{U}''} \begin{bmatrix} d & \mathbf{0} \\ \boldsymbol{Dv} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \delta & \boldsymbol{n} \\ \mathbf{0} & \boldsymbol{N} \end{bmatrix} \Rightarrow u_1 d + \boldsymbol{u}_2 \boldsymbol{Dv} = \delta$$

with δ being the greatest common divisor of d and $\boldsymbol{Dv}$ that is the greatest common divisor of $\{d, d_1, \ldots, d_l\}$. Since $\boldsymbol{U}''$ is a unit, it has an inverse $\boldsymbol{U}''^{-1}$ so that this transformation is reversible:

$$\underbrace{\begin{bmatrix} u_1' & \boldsymbol{u}_2' \\ \boldsymbol{u}_3' & \boldsymbol{U}_4' \end{bmatrix}}_{\boldsymbol{U}''^{-1}} \begin{bmatrix} \delta & \boldsymbol{n} \\ \mathbf{0} & \boldsymbol{N} \end{bmatrix} = \begin{bmatrix} d & \mathbf{0} \\ \boldsymbol{Dv} & \boldsymbol{D} \end{bmatrix} \Rightarrow u_1' \delta = d$$

With $\delta = u_1 d + \boldsymbol{u}_2 \boldsymbol{Dv}$ this gives $u_1'(u_1 d + \boldsymbol{u}_2 \boldsymbol{Dv}) = d$ hence $u_1' \boldsymbol{u}_2 = \mathbf{0}$ and $u_1' u_1 = u_1 u_1' = 1$ so that u_1, u_1' are units since this must hold for all d and $\boldsymbol{D}$.

But $u_1'\delta = d$ and u_1' being a unit means that d, δ are equivalent greatest common divisors of $\{d, d_1, \ldots, d_l\}$ so d divides all elements d_i of diagonal matrix $\boldsymbol{D}$ so that we have

$$\boldsymbol{U}'\boldsymbol{M}'\boldsymbol{V}' = \boldsymbol{D}' \text{ with } [\boldsymbol{D}']_{ii} = d_i \text{ being a divisor of } d_{i+\mathbb{N}}$$

and thus (A.46) is valid for m, n.

Bezout Identity for Matrices

Two matrices $\boldsymbol{A} \in R^{m,k}$ and $\boldsymbol{B} \in R^{n,k}$ with the same number of columns have a common square right divisor matrix $\boldsymbol{D} \in R^{k,k}$

$$\begin{aligned} &\exists \boldsymbol{X}, \boldsymbol{Y}: \ \boldsymbol{X}\boldsymbol{A} + \boldsymbol{Y}\boldsymbol{B} = \boldsymbol{D} \\ &\text{with } \boldsymbol{D} \text{ being the greatest common right divisor of } \boldsymbol{A}, \boldsymbol{B} \end{aligned} \tag{A.47}$$

which is the Bezout Identity for matrices. For left divisors this equation changes to $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{Y} = \boldsymbol{D}$ for $\boldsymbol{A} \in R^{k,m}$ and $\boldsymbol{B} \in R^{k,n}$ having the same number of lines, so this is the transpose of (A.47).

Proof To verify the equation for right divisors, the matrices $\boldsymbol{A}, \boldsymbol{B}$ can be collected in a vertical block matrix since they have the same number of columns. Then by applying (A.45) this leads to

$$\exists \text{ units } \boldsymbol{U}, \boldsymbol{V}: \ \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \\ \boldsymbol{U}_3 & \boldsymbol{U}_4 \end{bmatrix}}_{\boldsymbol{U}} \begin{bmatrix} \boldsymbol{D}' \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{V}$$

with $\boldsymbol{D}'$ diagonal. This gives

$$\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{D}'\boldsymbol{V}, \ \boldsymbol{B} = \boldsymbol{U}_3\boldsymbol{D}'\boldsymbol{V}$$

so that $\boldsymbol{A}, \boldsymbol{B}$ have a common right divisor $\boldsymbol{D} = \boldsymbol{D}'\boldsymbol{V}$. Since $\boldsymbol{U}$ is a unit, it is invertible over R:

$$\boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix}$$

This then provides the following representation:

$$\boldsymbol{U}^{-1} \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}' \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{V} = \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{0} \end{bmatrix}$$

This equation includes the Bezout identity in the first line:

$$\boldsymbol{X}\boldsymbol{A} + \boldsymbol{Y}\boldsymbol{B} = \boldsymbol{D}$$

It is now left to verify that the common right divisor $\boldsymbol{D}$ is also the greatest common right divisor of $\boldsymbol{A}, \boldsymbol{B}$. Let $\boldsymbol{D}''$ be another common right divisor of $\boldsymbol{A}, \boldsymbol{B}$, then there exist matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ so that $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{D}''$ and $\boldsymbol{B} = \boldsymbol{N}\boldsymbol{D}''$. Since $\boldsymbol{X}\boldsymbol{A} + \boldsymbol{Y}\boldsymbol{B} = \boldsymbol{D}$ this leads to $(\boldsymbol{X}\boldsymbol{M} + \boldsymbol{Y}\boldsymbol{N})\boldsymbol{D}'' = \boldsymbol{D}$. This shows that $\boldsymbol{D}''$ is itself a right divisor of $\boldsymbol{D}$ and therefore $\boldsymbol{D}$ is the greatest common right divisor of $\boldsymbol{A}, \boldsymbol{B}$.

Coprimeness

The pairs of matrices $\boldsymbol{A}_R \in R^{m,k}, \boldsymbol{B}_R \in R^{n,k}$ and $\boldsymbol{A}_L \in R^{k,m}, \boldsymbol{B}_L \in R^{k,n}$ are respectively right and left coprime if their respective greatest common right and left divisor is a unit matrix $\boldsymbol{U} \in R^{k,k}$:

$$\boldsymbol{A}_L\boldsymbol{A}_R + \boldsymbol{B}_L\boldsymbol{B}_R = \boldsymbol{U} \tag{A.48}$$

which can be compared to (A.35) for ring elements.

Proof To verify the matrix version, comparison of (A.48) to (A.47) leads to $\boldsymbol{U}$ being the greatest common right divisor of $\boldsymbol{A}_R, \boldsymbol{B}_R$ and by symmetry also to $\boldsymbol{U}$ being the greatest common left divisor of $\boldsymbol{A}_L, \boldsymbol{B}_L$. Since their greatest common divisor is a unit matrix (A.41), the pairs $\boldsymbol{A}_R, \boldsymbol{B}_R$ and $\boldsymbol{A}_L, \boldsymbol{B}_L$ are right and left coprime respectively.

A.1.6 Extension to Fractions

This section extends the preceding equations for elements of a principal commutative integral ring to equations for fractions of two elements of such a ring.

Fractions on a Ring

A fraction $\frac{a}{b}$ of two elements of a principal commutative integral ring $a \in R, b \in R \setminus \{0\}$ is an element of the set $F = \frac{R}{R\setminus\{0\}}$. It can be proven using laws (A.1) to (A.9) and (A.13) and (A.14) that F is itself a commutative integral ring where every nonzero element is a unit and thus F is a field, in this case the field of fractions of R.

McMillan Normal Form

Any matrix $\boldsymbol{M} \in F^{m,n}$ is equivalent to a diagonal matrix $\boldsymbol{D} \in F^{m,n}$ that has zeros below and above the diagonal and whose ith diagonal element is $[\boldsymbol{D}]_{ii} = \frac{a_i}{b_i}$ with a_i, b_i coprime and where a_i divides all lower $a_{i+\mathbb{N}}$ and all lower $b_{i+\mathbb{N}}$ divide b_i

$$\begin{aligned} &\exists \text{ units } \boldsymbol{U}, \boldsymbol{V} : \ \boldsymbol{U}\boldsymbol{M}\boldsymbol{V} = \boldsymbol{D} \text{ diagonal} \\ &\text{with } [\boldsymbol{D}]_{ii} = \frac{a_i}{b_i} \text{ where } a_i \text{ divides } a_{i+\mathbb{N}} \text{ and } b_{i+\mathbb{N}} \text{ divides } b_i \\ &\text{and } a_i, b_i \in R \text{ are coprime} \end{aligned} \tag{A.49}$$

with $\boldsymbol{U} \in R^{m,m}$ and $\boldsymbol{V} \in R^{n,n}$.

Proof By multiplying $\boldsymbol{M} \in F^{m,n}$ with the least common multiple d of all its denominators the resulting product $d\boldsymbol{M} \in R^{m,n}$ is a matrix with elements in R so that (A.46) applies. Hence there exist units $\boldsymbol{U} \in R^{m,m}$ and $\boldsymbol{V} \in R^{n,n}$ so that $\boldsymbol{U}d\boldsymbol{M}\boldsymbol{V} = \boldsymbol{D}$ is diagonal and $[\boldsymbol{D}]_{ii} = d_i \in R$ being a divisor of $d_{i+\mathbb{N}}$ so there exist $r_{i+\mathbb{N}} \in R$ with $d_{i+\mathbb{N}} = d_i r_{i+\mathbb{N}}$. Thus $\boldsymbol{U}\boldsymbol{M}\boldsymbol{V} = \frac{1}{d}\boldsymbol{D}$ is diagonal with $\frac{d_{i+\mathbb{N}}}{d} = \frac{d_i}{d} r_{i+\mathbb{N}}$. In general, a fraction can be reduced if its quotients are both divided by their greatest common divisor so that following (A.33) the reduced quotients are coprime. In this case the reduction of the quotients leads

to $\frac{d_i}{d} = \frac{a_i}{b_i}$ with $a_i, b_i \in R$ coprime and $\frac{d_{i+\mathbb{N}}}{d} = \frac{a_{i+\mathbb{N}}}{b_{i+\mathbb{N}}}$ with $a_{i+\mathbb{N}}, b_{i+\mathbb{N}}$ coprime. This gives $\frac{a_{i+\mathbb{N}}}{b_{i+\mathbb{N}}} = \frac{d_{i+\mathbb{N}}}{d} = \frac{d_i}{d} r_{i+\mathbb{N}} = \frac{a_i}{b_i} r_{i+\mathbb{N}}$ so that $a_{i+\mathbb{N}} b_i = a_i b_{i+\mathbb{N}} r_{i+\mathbb{N}}$. Since a_i, b_i are coprime a_i must be a divisor of $a_{i+\mathbb{N}}$ and since $a_{i+\mathbb{N}}, b_{i+\mathbb{N}}$ are coprime $b_{i+\mathbb{N}}$ must be a divisor of b_i.

Coprime Factorization

Any matrix $\boldsymbol{P} \in F^{m,n}$ has a non unique right and left coprime factorization with matrices $\boldsymbol{M}_R, \boldsymbol{M}_L \in R^{m,n}$, $\boldsymbol{N}_R \in R^{n,n}$, $\boldsymbol{N}_L \in R^{m,m}$ in the form of

$$\begin{aligned} &\boldsymbol{P} = \boldsymbol{M}_R \boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1} \boldsymbol{M}_L \\ &\text{with } \boldsymbol{M}_R, \boldsymbol{N}_R \text{ right coprime and } \boldsymbol{M}_L, \boldsymbol{N}_L \text{ left coprime} \end{aligned} \tag{A.50}$$

Proof Let $\boldsymbol{P}$ be factorized using (A.49) hence there exist unit matrices $\boldsymbol{U}, \boldsymbol{V}$ with $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}$ so that $\boldsymbol{D} \in F^{m,n}$ is diagonal with $[\boldsymbol{D}]_{ii} = \frac{a_i}{b_i}$ and $a_i, b_i \in R$ are coprime. Matrix $\boldsymbol{D}$ can be factorized easily since it is diagonal so that $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{B}_R^{-1} = \boldsymbol{B}_L^{-1}\boldsymbol{A}$ with diagonal matrices $\boldsymbol{A} \in R^{m,n}$, $\boldsymbol{B}_R \in R^{n,n}$, $\boldsymbol{B}_L \in R^{m,m}$ of the shape $[\boldsymbol{A}]_{ii} = a_i$ and $[\boldsymbol{B}_{R,L}]_{ii} = \{b_i \text{ for } i \leq \min(m,n), 1 \text{ for } i > \min(m,n)\}$. Since a_i, b_i are coprime, using (A.34) there exist $x_i, y_i \in R$ so that $x_i a_i + y_i b_i = 1$. In matrix form this means there exist diagonal matrices $\boldsymbol{X} \in R^{n,m}$, $\boldsymbol{Y}_L \in R^{n,n}$, $\boldsymbol{Y}_R \in R^{m,m}$ of the shape $[\boldsymbol{X}]_{ii} = x_i$ and $[\boldsymbol{Y}_{L,R}]_{ii} = \{b_i \text{ for } i \leq \min(m,n), 1 \text{ for } i > \min(m,n)\}$ with $\boldsymbol{X}\boldsymbol{A} + \boldsymbol{Y}_L\boldsymbol{B}_R = \boldsymbol{1}$ and $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}_L\boldsymbol{Y}_R = \boldsymbol{1}$. Following (A.48) this means that $\boldsymbol{A}, \boldsymbol{B}_R$ are right coprime and $\boldsymbol{A}, \boldsymbol{B}_L$ are left coprime.

With $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{B}_R^{-1} = \boldsymbol{B}_L^{-1}\boldsymbol{A}$ the factorization $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}$ can be written in the form

$$\boldsymbol{P} = (\boldsymbol{U}\boldsymbol{A})(\boldsymbol{V}^{-1}\boldsymbol{B}_R)^{-1} = (\boldsymbol{B}_L\boldsymbol{U}^{-1})^{-1}(\boldsymbol{A}\boldsymbol{V}) = \boldsymbol{M}_R\boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1}\boldsymbol{M}_L$$

with $\boldsymbol{M}_R = \boldsymbol{U}\boldsymbol{A}, \boldsymbol{N}_R = \boldsymbol{V}^{-1}\boldsymbol{B}_R$ right coprime since

$$\boldsymbol{1} = \boldsymbol{X}\boldsymbol{A} + \boldsymbol{Y}_L\boldsymbol{B}_R = \boldsymbol{X}\boldsymbol{U}^{-1}\boldsymbol{M}_R + \boldsymbol{Y}_L\boldsymbol{V}\boldsymbol{N}_R$$

and $\boldsymbol{M}_L = \boldsymbol{A}\boldsymbol{V}, \boldsymbol{N}_L = \boldsymbol{B}_L\boldsymbol{U}^{-1}$ left coprime since

$$\boldsymbol{1} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}_L\boldsymbol{Y}_R = \boldsymbol{M}_L\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{N}_L\boldsymbol{U}\boldsymbol{Y}_R$$

using (A.48).

Coprime Completion

A given coprime factorization of $\boldsymbol{P} \in F^{m,n}$ as in (A.50) can be completed by a non unique set of matrices $\boldsymbol{X}_R, \boldsymbol{X}_L \in R^{n,m}$, $\boldsymbol{Y}_R \in R^{m,m}$, $\boldsymbol{Y}_L \in R^{n,n}$ in the form of

$$\begin{aligned} &\boldsymbol{1} = \begin{bmatrix} \boldsymbol{X}_L & \boldsymbol{Y}_L \\ -\boldsymbol{N}_L & \boldsymbol{M}_L \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_R & -\boldsymbol{Y}_R \\ \boldsymbol{N}_R & \boldsymbol{X}_R \end{bmatrix} \\ &\text{with } \boldsymbol{X}_R, \boldsymbol{Y}_R \text{ right coprime and } \boldsymbol{X}_L, \boldsymbol{Y}_L \text{ left coprime} \end{aligned} \tag{A.51}$$

Furthermore all valid matrices $\boldsymbol{X}_i, \boldsymbol{Y}_i$ with $i \in \{R, L\}$ are a function of a free parameter matrix $\boldsymbol{Q} \in R^{n,m}$

$$\begin{aligned} \begin{bmatrix} \boldsymbol{X}_R \\ \boldsymbol{Y}_R \end{bmatrix} &= \begin{bmatrix} \boldsymbol{X}_{R0} & \boldsymbol{N}_R \\ \boldsymbol{Y}_{R0} & -\boldsymbol{M}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{Q} \end{bmatrix} \quad \text{with } \boldsymbol{Y}_R \text{ invertible} \\ \begin{bmatrix} \boldsymbol{X}_L & \boldsymbol{Y}_L \end{bmatrix} &= \begin{bmatrix} \boldsymbol{1} & \boldsymbol{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{L0} & \boldsymbol{Y}_{L0} \\ \boldsymbol{N}_L & -\boldsymbol{M}_L \end{bmatrix} \quad \text{with } \boldsymbol{Y}_L \text{ invertible} \end{aligned} \tag{A.52}$$

with $\boldsymbol{X}_{i0}, \boldsymbol{Y}_{i0}$ being a set of valid initial matrices that satisfy (A.51) and allow the construction of all other solutions $\boldsymbol{X}_i, \boldsymbol{Y}_i$ using (A.52).

Proof According to (A.50) there exists a coprime factorization of $\boldsymbol{P}$ in the form $\boldsymbol{P} = \boldsymbol{M}_R\boldsymbol{N}_R^{-1} = \boldsymbol{N}_L^{-1}\boldsymbol{M}_L$ so that $\boldsymbol{M}_L\boldsymbol{N}_R - \boldsymbol{N}_L\boldsymbol{M}_R = \boldsymbol{0}$. The factors $\boldsymbol{M}_R, \boldsymbol{N}_R$ are right coprime so with (A.48) there exist $\boldsymbol{X}'_L, \boldsymbol{Y}'_L$ and unit $\boldsymbol{U}_L$ so that $\boldsymbol{X}'_L\boldsymbol{M}_R + \boldsymbol{Y}'_L\boldsymbol{N}_R = \boldsymbol{U}_L$. The factors $\boldsymbol{M}_L, \boldsymbol{N}_L$ are left coprime so with (A.48) there exist $\boldsymbol{X}'_R, \boldsymbol{Y}'_R$ and unit $\boldsymbol{U}_R$ so that $\boldsymbol{M}_L\boldsymbol{X}_R + \boldsymbol{N}_L\boldsymbol{Y}_R = \boldsymbol{U}_R$. These equations can be written in block matrix form

$$\begin{bmatrix} \boldsymbol{X}'_L & \boldsymbol{Y}'_L \\ -\boldsymbol{N}_L & \boldsymbol{M}_L \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_R & -\boldsymbol{Y}'_R \\ \boldsymbol{N}_R & \boldsymbol{X}'_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_L & \boldsymbol{Q}' \\ \boldsymbol{0} & \boldsymbol{U}_R \end{bmatrix}$$

with $\boldsymbol{Q}' = \boldsymbol{Y}'_L\boldsymbol{X}'_R - \boldsymbol{X}'_L\boldsymbol{Y}'_R$. The matrix on the right side is a unit matrix which can be inverted on R. This can be done without affecting the given factors $\boldsymbol{M}_i, \boldsymbol{N}_i$ with $i \in \{R, L\}$ by factorizing the unit matrix in the way

$$\begin{bmatrix} \boldsymbol{U}_L & \boldsymbol{Q}' \\ \boldsymbol{0} & \boldsymbol{U}_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_L & \boldsymbol{Q}_L \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{Q}_R \\ \boldsymbol{0} & \boldsymbol{U}_R \end{bmatrix}$$

so that $\boldsymbol{Q}' = \boldsymbol{U}_L\boldsymbol{Q}_R + \boldsymbol{Q}_L\boldsymbol{U}_R$. Hence

$$\begin{bmatrix} \boldsymbol{U}_L & \boldsymbol{Q}_L \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}'_L & \boldsymbol{Y}'_L \\ -\boldsymbol{N}_L & \boldsymbol{M}_L \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_R & -\boldsymbol{Y}'_R \\ \boldsymbol{N}_R & \boldsymbol{X}'_R \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{Q}_R \\ \boldsymbol{0} & \boldsymbol{U}_R \end{bmatrix}^{-1} = \boldsymbol{1}$$

which using (A.59) can be written in new variables in the form of (A.51). The new variables are then defined as

$$\begin{aligned} \boldsymbol{X}_R &= (\boldsymbol{X}'_R - \boldsymbol{N}_R\boldsymbol{Q}_R)\boldsymbol{U}_R^{-1} \\ \boldsymbol{Y}_R &= (\boldsymbol{Y}'_R + \boldsymbol{M}_R\boldsymbol{Q}_R)\boldsymbol{U}_R^{-1} \\ \boldsymbol{X}_L &= \boldsymbol{U}_L^{-1}(\boldsymbol{X}'_L + \boldsymbol{Q}_L\boldsymbol{N}_L) \\ \boldsymbol{Y}_L &= \boldsymbol{U}_L^{-1}(\boldsymbol{Y}'_L - \boldsymbol{Q}_L\boldsymbol{M}_L) \end{aligned}$$

It can be seen that there is a degree of freedom in $\boldsymbol{Q}_L, \boldsymbol{Q}_R$ since these two matrices of $R^{n,m}$ only obey one equation $\boldsymbol{U}_L\boldsymbol{Q}_R + \boldsymbol{Q}_L\boldsymbol{U}_R = \boldsymbol{Q}'$ of $R^{n,m}$. This degree of freedom can be extracted in the form of a free matrix $\boldsymbol{Q}$ so that $\boldsymbol{U}_L\boldsymbol{Q}_R = \frac{1}{2}\boldsymbol{Q}' - \boldsymbol{U}_L\boldsymbol{Q}\boldsymbol{U}_R$ and $\boldsymbol{Q}_L\boldsymbol{U}_R = \frac{1}{2}\boldsymbol{Q}' + \boldsymbol{U}_L\boldsymbol{Q}\boldsymbol{U}_R$. This allows the new variables to be written in the form of (A.52).

It is now left to show that $\boldsymbol{Q}$ is a ring matrix. The right coprime part of the solution (A.52) can be written as

$$\begin{bmatrix} -\boldsymbol{Y}_R \\ \boldsymbol{X}_R \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{M}_R & -\boldsymbol{Y}_{R0} \\ \boldsymbol{N}_R & \boldsymbol{X}_{R0} \end{bmatrix}}_{\boldsymbol{U}} \begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{1} \end{bmatrix}$$

But following (A.51) the matrix $\boldsymbol{U}$ is a unit and hence $\boldsymbol{Q} \in R^{n,m}$.

A.2 Matrix Calculus

This section provides some important definitions and formulae for matrices that are used throughout this text.

A.2.1 Blocks

This section provides basic formulae for the treatment of block matrices.

Factorization of Block Matrices

A block matrix of the following shape can be factorized into

$$\begin{bmatrix} \boldsymbol{AEB} & \boldsymbol{AED} \\ \boldsymbol{CEB} & \boldsymbol{CED} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{E} \begin{bmatrix} \boldsymbol{B} & \boldsymbol{D} \end{bmatrix} \tag{A.53}$$

Determinant of Sums

The determinant of a sum of matrices is given by

$$\det(\boldsymbol{A} + \boldsymbol{BCD}) = \det \boldsymbol{A} \det \boldsymbol{C} \det(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B}) \tag{A.54}$$

Inverse of Sums

The inverse of a sum of matrices is given by

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1} \tag{A.55}$$

Useful Relations

Two useful relations between inverses of sums are

$$\begin{aligned} \boldsymbol{A}(\mathbf{1} + \boldsymbol{BA})^{-1} &= (\mathbf{1} + \boldsymbol{AB})^{-1}\boldsymbol{A} \\ \boldsymbol{AB}(\mathbf{1} + \boldsymbol{AB})^{-1} &= \mathbf{1} - (\mathbf{1} + \boldsymbol{AB})^{-1} \end{aligned} \tag{A.56}$$

Schur Decomposition

A block matrix can be decomposed in its block Schur form

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{CA}^{-1} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D} - \boldsymbol{CA}^{-1}\boldsymbol{B} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \boldsymbol{A}^{-1}\boldsymbol{B} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{A.57}$$

if $\boldsymbol{A}$ is invertible.

Determinant of Block Matrices

The determinant of a block matrix is given by

$$\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{cases} \det \boldsymbol{A} \det(\boldsymbol{D} - \boldsymbol{CA}^{-1}\boldsymbol{B}); & \exists \boldsymbol{A}^{-1} \\ \det \boldsymbol{D} \det(\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}); & \exists \boldsymbol{D}^{-1} \end{cases} \tag{A.58}$$

which follows from (A.57).

Inverse of Block Matrices

The inverse of a block matrix is given by

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{cases} \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{Y} \\ -\boldsymbol{Y}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{Y} \end{bmatrix}; & \exists \boldsymbol{A}^{-1}, \exists \boldsymbol{D}^{-1} \\ \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{X} & \boldsymbol{Y} \end{bmatrix}; & \exists \boldsymbol{A}^{-1}, \exists \boldsymbol{D}^{-1} \\ \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{Y}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{Y} \\ -\boldsymbol{Y}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{Y} \end{bmatrix}; & \exists \boldsymbol{A}^{-1} \\ \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{X} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{X}\boldsymbol{B}\boldsymbol{D}^{-1} \end{bmatrix}; & \exists \boldsymbol{D}^{-1} \end{cases} \tag{A.59}$$

with $\boldsymbol{X} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}$ and $\boldsymbol{Y} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$ which follows from (A.57).

A.2.2 Spectra

This section provides formulae for eigenvalues and definiteness of matrices.

Eigenvalues

If λ is an eigenvalue of $\boldsymbol{A} \in \mathbb{C}^{n,n}$ that is $\det(\lambda\boldsymbol{1} - \boldsymbol{A}) = 0$ then

Matrix	Eigenvalue
$\boldsymbol{A}^T$	λ
$\overline{\boldsymbol{A}}$	$\overline{\lambda}$
$\boldsymbol{A}^{-1}$	λ^{-1}
$\alpha\boldsymbol{A}$	$\alpha\lambda$
$\beta\boldsymbol{1} + \boldsymbol{A}$	$\beta + \lambda$

(A.60)

Proof This is due to $\det(\lambda\boldsymbol{1} - \boldsymbol{A}) = 0 \Leftrightarrow \det(\lambda\boldsymbol{1} - \boldsymbol{A})^T = \det(\lambda\boldsymbol{1} - \boldsymbol{A}^T) = 0 \Leftrightarrow \overline{\det(\lambda\boldsymbol{1} - \boldsymbol{A})} = \det(\overline{\lambda}\boldsymbol{1} - \overline{\boldsymbol{A}}) = 0 \Leftrightarrow \det(\lambda\boldsymbol{A}(\boldsymbol{A}^{-1} - \lambda^{-1}\boldsymbol{1})) = \lambda^n \det\boldsymbol{A} \det(\boldsymbol{A}^{-1} - \lambda^{-1}\boldsymbol{1}) = 0 \Leftrightarrow \det(\lambda^{-1}\boldsymbol{1} - \boldsymbol{A}^{-1}) = 0 \Leftrightarrow \alpha^n \det(\lambda\boldsymbol{1} - \boldsymbol{A}) = \det(\alpha\lambda\boldsymbol{1} - \alpha\boldsymbol{A}) = 0 \Leftrightarrow \det((\beta + \lambda)\boldsymbol{1} - (\beta\boldsymbol{1} + \boldsymbol{A})) = 0$.

Spectral Decomposition of Normal Matrices

A matrix $\boldsymbol{A}$ is normal if and only if $\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{A}^H\boldsymbol{A}$. Then there always exists a spectral decomposition of $\boldsymbol{A}$ in the form of

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H = \sum_k \lambda_k \boldsymbol{a}_k \boldsymbol{a}_k^H \tag{A.61}$$

with unitary $\boldsymbol{U}$ so that $\boldsymbol{U}^{-1} = \boldsymbol{U}^H$. The λ_k are the eigenvalues and the $\boldsymbol{a}_k$ the eigenvectors of $\boldsymbol{A}$ with $\boldsymbol{a}_i^H\boldsymbol{a}_i = 1$ and $\boldsymbol{a}_i^H\boldsymbol{a}_{j\neq i} = 0$.

Remark Hermitian matrices $\boldsymbol{A} = \boldsymbol{A}^H$ are normal since $\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{A}^2 = \boldsymbol{A}^H\boldsymbol{A}$.

Definiteness of Hermitian Matrices

A hermitian matrix $\boldsymbol{A} = \boldsymbol{A}^H$ is called positive semidefinite if

$$\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} \geq 0, \forall \boldsymbol{x} \neq \boldsymbol{0} \Leftrightarrow \lambda_k \geq 0 \tag{A.62}$$

for all its eigenvalues λ_k. A matrix is called positive definite if "$\geq$" can be replaced by "$>$". A matrix $\boldsymbol{B}$ is called negative (semi)definite if $-\boldsymbol{B}$ is positive (semi)definite.

Proof Since $\boldsymbol{A}$ is hermitian it always has a spectral decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H$ due to (A.61). Then we have $\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} = (\boldsymbol{U}^H \boldsymbol{x}) \boldsymbol{A} (\boldsymbol{U}^H \boldsymbol{x}) = \boldsymbol{y}^H \boldsymbol{D} \boldsymbol{y} = \sum_k \lambda_k |\boldsymbol{y}_k|^2 \geq 0, \forall \boldsymbol{y} \neq \boldsymbol{0} \Leftrightarrow \lambda_k \geq 0$ for all eigenvalues λ_k. Finally since $\boldsymbol{y} = \boldsymbol{U}^H \boldsymbol{x}$ with unitary $\boldsymbol{U}$ we have $\boldsymbol{y} \neq \boldsymbol{0} \Leftrightarrow \boldsymbol{x} \neq \boldsymbol{0}$ so that (A.62) follows.

Remark For a hermitian matrix $\boldsymbol{A} = \boldsymbol{A}^H$ the notation $\boldsymbol{A} = 0$ is equivalent to $\boldsymbol{A} = \boldsymbol{0}$. On the one hand $\boldsymbol{A} = \boldsymbol{0} \Rightarrow \boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} = 0, \forall \boldsymbol{x}$ and on the other hand $\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} = 0, \forall \boldsymbol{x} \Rightarrow \boldsymbol{x}\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H \boldsymbol{x} = (\boldsymbol{U}^H \boldsymbol{x})^H \boldsymbol{D} (\boldsymbol{U}^H \boldsymbol{x}) = \boldsymbol{y}^H \boldsymbol{D} \boldsymbol{y} = \sum_k \lambda_k |\boldsymbol{y}_k|^2 = 0, \forall \boldsymbol{y} \Leftrightarrow \lambda_k = 0$ for all eigenvalues λ_k so that $\boldsymbol{D} = \boldsymbol{0}$ and thus $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^H = \boldsymbol{0}$.

Definiteness of Hermitian Block Matrices

A necessary condition for a hermitian block matrix can be given by

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} \geq 0 \Rightarrow \boldsymbol{A} \geq 0 \text{ and } \boldsymbol{C} \geq 0 \tag{A.63}$$

The condition also holds if "$\geq$" is replaced by "$>$".

Proof It follows from (A.62) that

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} \geq 0 \Leftrightarrow \boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} + 2\,\mathrm{Re}(\boldsymbol{x}^H \boldsymbol{B} \boldsymbol{y}) + \boldsymbol{y}^H \boldsymbol{C} \boldsymbol{y} \geq 0$$

for all $\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix}^T \neq \boldsymbol{0}$. Let $\boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{y} = \boldsymbol{0}$ then $\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$ so that $\boldsymbol{A} \geq 0$ and let $\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{y} \neq \boldsymbol{0}$ then $\boldsymbol{y}^H \boldsymbol{C} \boldsymbol{y} \geq 0$ for all $\boldsymbol{y} \neq \boldsymbol{0}$ so that $\boldsymbol{C} \geq 0$. The proof for positive definiteness is similar.

Congruence

Congruence preserves definiteness

$$\boldsymbol{A} \geq 0 \Leftrightarrow \boldsymbol{T}^H \boldsymbol{A} \boldsymbol{T} \geq 0 \tag{A.64}$$

for invertible $\boldsymbol{T}$.

Proof It follows from (A.62) that $\boldsymbol{A} \geq 0 \Leftrightarrow \boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} \geq 0, \forall \boldsymbol{x}$ and with $\boldsymbol{x} = \boldsymbol{T}\boldsymbol{y}$ we have $\boldsymbol{x}^H \boldsymbol{A} \boldsymbol{x} = \boldsymbol{y}^H \boldsymbol{T}^H \boldsymbol{A} \boldsymbol{T} \boldsymbol{y}, \forall \boldsymbol{y} \Leftrightarrow \boldsymbol{T}^H \boldsymbol{A} \boldsymbol{T} \geq 0$. The same is also true for positive and negative (semi)definiteness.

Schur Congruence

If $\boldsymbol{C} > 0$ then the following is true

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} > 0 \Leftrightarrow \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^H > 0 \\ \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} \geq 0 \Leftrightarrow \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^H \geq 0 \end{aligned} \tag{A.65}$$

which is an application of the congruence.

Proof Rearranging the matrix gives

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{C} & \boldsymbol{B}^H \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$$

Then according to (A.57), the block matrix can be factorized

$$\begin{bmatrix} \boldsymbol{C} & \boldsymbol{B}^H \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \boldsymbol{B}\boldsymbol{C}^{-1} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{C} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^H \end{bmatrix} \begin{bmatrix} \mathbf{1} & \boldsymbol{C}^{-1}\boldsymbol{B}^H \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Due to (A.64) this leads to

$$\begin{aligned} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} > 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{C} & \boldsymbol{B}^H \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} > 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{C} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^H \end{bmatrix} > 0 \\ \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^H & \boldsymbol{C} \end{bmatrix} \geq 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{C} & \boldsymbol{B}^H \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \geq 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{C} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^H \end{bmatrix} \geq 0 \end{aligned}$$

Hence (A.65) follows since $\boldsymbol{C} > 0$.

S-Lemma

The so called S-lemma states

$$\begin{aligned} &\exists \sigma \geq 0 : \boldsymbol{Q} \geq \sigma \boldsymbol{P} \text{ is equivalent to} \\ &\forall \boldsymbol{x} : \boldsymbol{x}^H \boldsymbol{P} \boldsymbol{x} \geq 0 \Rightarrow \boldsymbol{x}^H \boldsymbol{Q} \boldsymbol{x} \geq 0 \end{aligned} \tag{A.66}$$

if there exists a $\boldsymbol{y}$ so that $\boldsymbol{y}^H \boldsymbol{P} \boldsymbol{y} > 0$.

A.2.3 Operations

This section provides formulae for special operations on matrices.

Geometric Series

The inverse of $\mathbf{1} - \boldsymbol{A}$ can be expressed in form of the geometric series

$$(\mathbf{1} - \boldsymbol{A})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{A}^k \text{ if } |\lambda_A| < 1 \tag{A.67}$$

for all eigenvalues λ_A of $\boldsymbol{A}$.

Proof Let $\boldsymbol{S} = \sum_{k=0}^{\infty} \boldsymbol{A}^k$ which only converges if $|\lambda_A| < 1$. Then $\boldsymbol{S}$ satisfies the equation $\boldsymbol{AS} = \boldsymbol{S} - \boldsymbol{1}$ so that (A.67) follows.

Scalar Product of Matrices

The trace of the hermitian product defines a scalar product

$$\boldsymbol{A} \bullet \boldsymbol{B} = \operatorname{tr}(\boldsymbol{A}^H \boldsymbol{B}) = \sum_i \sum_j \overline{a}_{ij} b_{ij} \tag{A.68}$$

Proof This is a scalar product since $\boldsymbol{A} \bullet \boldsymbol{B} = \sum_i \sum_j \overline{a}_{ij} b_{ij} = \sum_i \sum_j \overline{a_{ij} \overline{b}_{ij}} = \overline{\boldsymbol{B} \bullet \boldsymbol{A}}$ and $\boldsymbol{C} \bullet (\alpha \boldsymbol{A} + \beta \boldsymbol{B}) = \alpha \operatorname{tr}(\boldsymbol{C}^H \boldsymbol{A}) + \beta \operatorname{tr}(\boldsymbol{C}^H \boldsymbol{B}) = \alpha \boldsymbol{C} \bullet \boldsymbol{A} + \beta \boldsymbol{C} \bullet \boldsymbol{B}$ and $\boldsymbol{A} \bullet \boldsymbol{A} = \sum_i \sum_j \overline{a}_{ij} a_{ij} = \sum_i \sum_j |a|_{ij}^2 \geq 0$.

General Properties

Aside from that, the scalar product has the properties

$$\boldsymbol{A} \bullet \boldsymbol{B} = \boldsymbol{B}^H \bullet \boldsymbol{A}^H = \boldsymbol{A}^T \bullet \boldsymbol{B}^T \tag{A.69}$$

$$\boldsymbol{A} \bullet (\boldsymbol{BC}) = (\boldsymbol{B}^H \boldsymbol{A}) \bullet \boldsymbol{C} = (\boldsymbol{AC}^H) \bullet \boldsymbol{B} \tag{A.70}$$

$$\alpha \bullet \operatorname{tr} \boldsymbol{B} = \alpha \boldsymbol{1} \bullet \boldsymbol{B} \tag{A.71}$$

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} \bullet \begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \boldsymbol{A} \bullet \boldsymbol{C} + \boldsymbol{B} \bullet \boldsymbol{D} \tag{A.72}$$

Proof The first line is due to $\boldsymbol{A} \bullet \boldsymbol{B} = \operatorname{tr}(\boldsymbol{A}^H \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B} \boldsymbol{A}^H) = \boldsymbol{B}^H \bullet \boldsymbol{A}^H = \operatorname{tr}(\boldsymbol{B} \boldsymbol{A}^H)^T = \boldsymbol{A}^T \bullet \boldsymbol{B}^T$. The second line follows from $\boldsymbol{A} \bullet (\boldsymbol{BC}) = \operatorname{tr}(\boldsymbol{A}^H \boldsymbol{BC}) = \operatorname{tr}((\boldsymbol{B}^H \boldsymbol{A})^H \boldsymbol{C}) = (\boldsymbol{B}^H \boldsymbol{A}) \bullet \boldsymbol{C} = \operatorname{tr}((\boldsymbol{AC}^H)^H \boldsymbol{B}) = (\boldsymbol{AC}^H) \bullet \boldsymbol{B}$. The third line is due to $\alpha \bullet \operatorname{tr} \boldsymbol{B} = \operatorname{tr}(\overline{\alpha} \operatorname{tr} \boldsymbol{B}) = \overline{\alpha} \operatorname{tr} \boldsymbol{B} = \operatorname{tr}(\overline{\alpha} \boldsymbol{B}) = \alpha \boldsymbol{1} \bullet \boldsymbol{B}$.

Properties for Hermitian Matrices

For hermitian matrices, the scalar product has the additional properties

$$\boldsymbol{A} = \boldsymbol{A}^H, \boldsymbol{B} = \boldsymbol{B}^H \Rightarrow \boldsymbol{A} \bullet \boldsymbol{B} \in \mathbb{R} \tag{A.73}$$

$$\boldsymbol{A} = \boldsymbol{A}^H \Rightarrow \begin{cases} \boldsymbol{A} \bullet (\boldsymbol{B} + \boldsymbol{B}^H) = 2 \operatorname{Re}(\boldsymbol{A} \bullet \boldsymbol{B}) \\ \boldsymbol{A} \bullet (\boldsymbol{B} - \boldsymbol{B}^H) = 2 \operatorname{Im}(\boldsymbol{A} \bullet \boldsymbol{B}) \end{cases} \tag{A.74}$$

$$\boldsymbol{A} = \boldsymbol{A}^H \geq 0, \boldsymbol{B} = \boldsymbol{B}^H \geq 0 \Rightarrow \boldsymbol{A} \bullet \boldsymbol{B} \geq 0 \tag{A.75}$$

Proof For hermitian matrices $\boldsymbol{A}, \boldsymbol{B}$ the scalar product is commutative since $\boldsymbol{A} \bullet \boldsymbol{B} = \boldsymbol{B}^H \bullet \boldsymbol{A}^H = \boldsymbol{B} \bullet \boldsymbol{A} = \overline{\boldsymbol{A} \bullet \boldsymbol{B}}$ due to (A.69) and the definition of the scalar product. Thus being equal to its conjugate, the scalar product of hermitian matrices is real. The next equality follows from $\boldsymbol{A} \bullet (\boldsymbol{B} \pm \boldsymbol{B}^H) = \boldsymbol{A} \bullet \boldsymbol{B} \pm \boldsymbol{A} \bullet \boldsymbol{B}^H = \boldsymbol{A} \bullet \boldsymbol{B} \pm \boldsymbol{B} \bullet \boldsymbol{A}^H = \boldsymbol{A} \bullet \boldsymbol{B} \pm \boldsymbol{B} \bullet \boldsymbol{A} = \boldsymbol{A} \bullet \boldsymbol{B} \pm \overline{\boldsymbol{A} \bullet \boldsymbol{B}}$ which gives the double of the real and imaginary parts of $\boldsymbol{A} \bullet \boldsymbol{B}$ respectively. Since $\boldsymbol{A}$ is hermitian it always has a spectral decomposition $\boldsymbol{A} = \sum_k \lambda_k \boldsymbol{a}_k \boldsymbol{a}_k^H$ due to (A.61). Applying this on (A.68) gives $\boldsymbol{A} \bullet \boldsymbol{B} = \operatorname{tr}(\boldsymbol{A}^H \boldsymbol{B}) = \sum_k \lambda_k \operatorname{tr}(\boldsymbol{a}_k \boldsymbol{a}_k^H \boldsymbol{B}) = \sum_k \lambda_k \operatorname{tr}(\boldsymbol{a}_k^H \boldsymbol{B} \boldsymbol{a}_k) = \sum_k \lambda_k \boldsymbol{a}_k^H \boldsymbol{B} \boldsymbol{a}_k \geq 0$ since $\lambda_k \geq 0$ due to $\boldsymbol{A} \geq 0$ and $\boldsymbol{a}_k^H \boldsymbol{B} \boldsymbol{a}_k \geq 0$ due to (A.62) and $\boldsymbol{B} \geq 0$.

Matrix Square Root

A matrix $\boldsymbol{B}$ is a square root of matrix $\boldsymbol{A}$ if

$$\boldsymbol{A} = \boldsymbol{B}^2 \Leftrightarrow \boldsymbol{B} = \boldsymbol{A}^{\frac{1}{2}} \Leftrightarrow \boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H = (\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H)^{\frac{1}{2}} \tag{A.76}$$

$$\boldsymbol{A} = \boldsymbol{B}^2 \geq 0 \Rightarrow \boldsymbol{B} = \boldsymbol{A}^{\frac{1}{2}} \geq 0 \text{ is unique} \tag{A.77}$$

$$\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B} \geq 0 \Rightarrow \boldsymbol{B} = \boldsymbol{U}\boldsymbol{A}^{\frac{1}{2}} \tag{A.78}$$

with $\boldsymbol{U}$ unitary, that is $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{1}$.

Proof The first line is the definition of the matrix square root followed by a property that can be verified by $\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^H = \boldsymbol{U}\boldsymbol{B}^2\boldsymbol{U}^H = \boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H$ for unitary $\boldsymbol{U}$. The second line can be verified since for hermitian matrices and thus positive semidefinite matrices there always exists a spectral decomposition which can be written in the form of $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H$ with diagonal $\boldsymbol{D} \geq 0$ and unitary $\boldsymbol{V}$. Then due to (A.76) this leads to $\boldsymbol{B} = \boldsymbol{A}^{\frac{1}{2}} = (\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H)^{\frac{1}{2}} = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{V}^H$ and since $\boldsymbol{D} \geq 0$ is diagonal we have $[\boldsymbol{D}]_{ii} = d_i \geq 0$ and thus its square root $[\boldsymbol{D}^{\frac{1}{2}}]_{ii} = d_i^{\frac{1}{2}}$ is unique and positive semidefinite. The third line again makes use of the spectral decomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H$ which can be written as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{V}^H = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{V}^H\boldsymbol{U}^H\boldsymbol{U}\boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{V}^H$ so that $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{V}^H = \boldsymbol{U}\boldsymbol{A}^{\frac{1}{2}}$.

Pseudoinverse Matrix

A matrix $\boldsymbol{A} \in \mathbb{C}^{m,n}$ with $m \geq n$ and $\operatorname{rnk}\boldsymbol{A} = n$ has a unique pseudoinverse matrix $\boldsymbol{A}^+ \in \mathbb{C}^{n,m}$ with $\operatorname{rnk}\boldsymbol{A}^+ = n$ so that

$$\begin{aligned} &\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{1} \\ &(\boldsymbol{A}\boldsymbol{A}^+)^H = \boldsymbol{A}\boldsymbol{A}^+ \\ &\text{with } \boldsymbol{A}^+ = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H \end{aligned} \tag{A.79}$$

This is the full column rank case of the Moore Penrose pseudoinverse.

Orthogonal Completion

A matrix $\boldsymbol{A} \in \mathbb{C}^{m,n}$ with $m > n$ and $\operatorname{rnk}\boldsymbol{A} = n$ has an orthogonal completion $\boldsymbol{A}_\perp \in \mathbb{C}^{m,m-n}$ with $\operatorname{rnk}\boldsymbol{A}_\perp = m - n$ so that

$$\begin{aligned} &\boldsymbol{A}^H\boldsymbol{A}_\perp = \boldsymbol{0} \\ &\begin{bmatrix}\boldsymbol{A} & \boldsymbol{A}_\perp\end{bmatrix}\begin{bmatrix}\boldsymbol{A}^+ \\ \boldsymbol{A}_\perp^+\end{bmatrix} = \boldsymbol{1} \\ &\text{parametrized by } \boldsymbol{A}_\perp = \boldsymbol{A}_{\perp 0}\boldsymbol{T} \end{aligned} \tag{A.80}$$

with $\boldsymbol{T}$ invertible and the pseudoinverse given by (A.79).

Proof Let $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n\end{bmatrix}$. Since $\boldsymbol{A}$ has full rank, all $\boldsymbol{a}_k \in \mathbb{C}^m$ are linearly independent. Then $m - n$ linearly independent vectors $\boldsymbol{b}_l \in \mathbb{C}^m$ can be found so that $\boldsymbol{a}_k^H\boldsymbol{b}_l = 0; \forall k \in \mathbb{N}_n, \forall l \in \mathbb{N}_{m-n}$, which means that they are all orthogonal on every vector $\boldsymbol{a}_k$. The vectors $\boldsymbol{b}_l$ can then be stacked to form a

matrix $\boldsymbol{A}_\perp = \begin{bmatrix}\boldsymbol{b}_1 & \cdots & \boldsymbol{b}_{m-n}\end{bmatrix} \in \mathbb{C}^{m,m-n}$ with $\operatorname{rnk} \boldsymbol{A}_\perp = m-n$ so that due to their orthogonality on the vectors $\boldsymbol{a}_k$ we get

$$\boldsymbol{A}^H \boldsymbol{A}_\perp = \begin{bmatrix}\boldsymbol{a}_1^H \boldsymbol{b}_1 & \cdots & \boldsymbol{a}_1^H \boldsymbol{b}_{m-n} \\ \vdots & & \vdots \\ \boldsymbol{a}_n^H \boldsymbol{b}_1 & \cdots & \boldsymbol{a}_n^H \boldsymbol{b}_{m-n}\end{bmatrix} = \boldsymbol{0}$$

This property does not depend on the ordering of the column vectors $\boldsymbol{b}_k$ in $\boldsymbol{A}_\perp$ so that an invertible matrix $\boldsymbol{T}$ parametrizes the orthogonal completions in the form of $\boldsymbol{A}_\perp = \boldsymbol{A}_{\perp 0}\boldsymbol{T}$. Since matrices $\boldsymbol{A}, \boldsymbol{A}_\perp$ have full rank, according to (A.79) there exist pseudoinverse matrices $\boldsymbol{A}^+ = (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H, \boldsymbol{A}_\perp^+ = (\boldsymbol{A}_\perp^H\boldsymbol{A}_\perp)^{-1}\boldsymbol{A}_\perp^H$ so that $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{1}, \boldsymbol{A}_\perp^+\boldsymbol{A}_\perp = \boldsymbol{1}$. Along with the orthogonality $\boldsymbol{A}^H\boldsymbol{A}_\perp = \boldsymbol{0}$, this can be written in the form

$$\begin{bmatrix}\boldsymbol{A}^+ \\ \boldsymbol{A}_\perp^+\end{bmatrix}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{A}_\perp\end{bmatrix} = \boldsymbol{1}$$

so that (A.80) follows.

Matrix Variations

The product rule for matrices is given by

$$\delta(\boldsymbol{A}\boldsymbol{B}) = \delta\boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\delta\boldsymbol{B} \tag{A.81}$$

The variation of an inverse matrix is given by

$$\delta(\boldsymbol{A}^{-1}) = -\boldsymbol{A}^{-1}\delta\boldsymbol{A}\boldsymbol{A}^{-1} \tag{A.82}$$

The variation of a matrix square root is determined by

$$\delta\boldsymbol{B} = \delta(\boldsymbol{B}^{\frac{1}{2}})\boldsymbol{B}^{\frac{1}{2}} + \boldsymbol{B}^{\frac{1}{2}}\delta(\boldsymbol{B}^{\frac{1}{2}}) \tag{A.83}$$

Proof The product rule for vectors can be deduced from the scalar product rule

$$\delta(\boldsymbol{a}^T\boldsymbol{b}) = \delta\left(\sum_{i=1}^k a_i b_i\right) = \sum_{i=1}^k \delta a_i b_i + \sum_{i=1}^k a_i \delta b_i = \delta\boldsymbol{a}^T\boldsymbol{b} + \boldsymbol{a}^T\delta\boldsymbol{b}$$

so that the product rule for matrices follows

$$\delta(\boldsymbol{A}\boldsymbol{B}) = \delta\left(\begin{bmatrix}\boldsymbol{a}_1^T \\ \vdots \\ \boldsymbol{a}_m^T\end{bmatrix}\begin{bmatrix}\boldsymbol{b}_1 & \cdots & \boldsymbol{b}_n\end{bmatrix}\right) = \delta\begin{bmatrix}\boldsymbol{a}_1^T\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_1^T\boldsymbol{b}_n \\ \vdots & & \vdots \\ \boldsymbol{a}_m^T\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_m^T\boldsymbol{b}_n\end{bmatrix} =$$
$$= \begin{bmatrix}\delta\boldsymbol{a}_1^T\boldsymbol{b}_1 + \boldsymbol{a}_1^T\delta\boldsymbol{b}_1 & \cdots & \delta\boldsymbol{a}_1^T\boldsymbol{b}_n + \boldsymbol{a}_1^T\delta\boldsymbol{b}_n \\ \vdots & & \vdots \\ \delta\boldsymbol{a}_m^T\boldsymbol{b}_1 + \boldsymbol{a}_m^T\delta\boldsymbol{b}_1 & \cdots & \delta\boldsymbol{a}_m^T\boldsymbol{b}_n + \boldsymbol{a}_m^T\delta\boldsymbol{b}_n\end{bmatrix} = \delta\boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\delta\boldsymbol{B}$$

For the identity matrix, the variation is zero

$$\delta(\boldsymbol{A}\boldsymbol{A}^{-1}) = \delta\boldsymbol{1} = \boldsymbol{0}$$

so that the variation of an inverse matrix follows

$$\delta \boldsymbol{A}\boldsymbol{A}^{-1} + \boldsymbol{A}\delta(\boldsymbol{A}^{-1}) = \delta(\boldsymbol{A}\boldsymbol{A}^{-1}) = \boldsymbol{0} \Leftrightarrow \delta(\boldsymbol{A}^{-1}) = -\boldsymbol{A}^{-1}\delta \boldsymbol{A}\boldsymbol{A}^{-1}$$

The variation of a squared matrix is given by the product rule

$$\delta(\boldsymbol{A}^2) = \delta \boldsymbol{A}\boldsymbol{A} + \boldsymbol{A}\delta \boldsymbol{A}$$

so that the variation of a matrix square root follows by setting $\boldsymbol{A} = \boldsymbol{B}^{\frac{1}{2}}$.

A.3 Matrix Equations

This section discusses some important matrix equations and provides their solutions or solution existence conditions.

A.3.1 Underdetermined Linear Equations

The linear equation

$$\boldsymbol{X}\boldsymbol{A} = \boldsymbol{B} \tag{A.84}$$

for $\boldsymbol{A} \in \mathbb{C}^{m,n}$ with $m \geq n$ and $\operatorname{rnk} \boldsymbol{A} = n$ has solutions

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{A}^+ + \boldsymbol{Y}(\boldsymbol{1} - \boldsymbol{A}\boldsymbol{A}^+) \tag{A.85}$$

with free parameter $\boldsymbol{Y}$.

Proof Define $\boldsymbol{X}\boldsymbol{A}_\perp = \boldsymbol{C}$ using the orthogonal completion $\boldsymbol{A}_\perp$ of $\boldsymbol{A}$. This together with (A.84) gives

$$\boldsymbol{X}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{A}_\perp\end{bmatrix} = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{C}\end{bmatrix} \Leftrightarrow \boldsymbol{X} = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{C}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}^+ \\ \boldsymbol{A}_\perp^+\end{bmatrix}$$

according to (A.80), so that $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{A}^+ + \boldsymbol{C}\boldsymbol{A}_\perp^+$. This is a solution of (A.84) since $\boldsymbol{X}\boldsymbol{A} = \boldsymbol{B}\boldsymbol{A}^+\boldsymbol{A} + \boldsymbol{C}\boldsymbol{A}_\perp^+\boldsymbol{A} = \boldsymbol{B}$ which does not depend on the choice of $\boldsymbol{C}$ so that instead of $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{A}_\perp$ also any other $\boldsymbol{C} = \boldsymbol{Y}\boldsymbol{A}_\perp$ with an arbitrary compatible matrix $\boldsymbol{Y}$ leads to a valid solution $\boldsymbol{X}$. Hence $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{A}^+ + \boldsymbol{Y}\boldsymbol{A}_\perp\boldsymbol{A}_\perp^+ = \boldsymbol{B}\boldsymbol{A}^+ + \boldsymbol{Y}(\boldsymbol{1} - \boldsymbol{A}\boldsymbol{A}^+)$ using (A.80).

A.3.2 Sylvester Equations

The Sylvester matrix equation

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{C} = \boldsymbol{0} \tag{A.86}$$

with square matrices $\boldsymbol{A}, \boldsymbol{B}$ has a unique solution $\boldsymbol{X}$ if and only if $\lambda_A + \lambda_B \neq 0$ for all eigenvalues λ_A, λ_B of $\boldsymbol{A}, \boldsymbol{B}$ respectively. For the important special case $\boldsymbol{B} = \boldsymbol{A}^H$ there is an additional property

$$\boldsymbol{B} = \boldsymbol{A}^H \text{ and } \boldsymbol{C} \geq 0 \Rightarrow \boldsymbol{X} \geq 0 \tag{A.87}$$

The same is true if $\boldsymbol{C} > 0$ then $\boldsymbol{X} > 0$.

A.3.3 Stein Equations

The Sylvester matrix equation

$$\boldsymbol{AXB} - \boldsymbol{X} + \boldsymbol{C} = \boldsymbol{0} \tag{A.88}$$

with square matrices $\boldsymbol{A}, \boldsymbol{B}$ has a unique solution $\boldsymbol{X}$ if and only if $\lambda_A \lambda_B \neq 1$ for all eigenvalues λ_A, λ_B of $\boldsymbol{A}, \boldsymbol{B}$ respectively. For the important special case $\boldsymbol{B} = \boldsymbol{A}^H$ there is an additional property

$$\boldsymbol{B} = \boldsymbol{A}^H \text{ and } \boldsymbol{C} \geq 0 \Rightarrow \boldsymbol{X} \geq 0 \tag{A.89}$$

The same is true if $\boldsymbol{C} > 0$ then $\boldsymbol{X} > 0$.

Appendix B

Analysis

B.1 Complex Analysis

This section provides some important theorems of complex analysis.

B.1.1 Analytic Functions

A function f is analytic on a domain $D \subset \mathbb{C}$ if it is equal to its unique Taylor series representation

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k \Leftrightarrow f \text{ analytic on } D \tag{B.1}$$

for any $z_0 \in D$.

B.1.2 Isolated Singularity

A function f has an isolated singularity at $z_0 \in \mathbb{C}$ if it is analytic on a domain $D \setminus z_0 \subset \mathbb{C}$. Then it is equal to its unique Laurent series representation

$$f(z) = \sum_{k=-\infty}^{\infty} c_k (z - z_0)^k \Leftrightarrow f \text{ analytic on } D \setminus z_0 \tag{B.2}$$

B.1.3 Residue

The residue of a function f with an isolated singularity at $z_0 \in \mathbb{C}$ is defined as

$$\operatorname{Res} f = \frac{1}{2\pi j} \oint_C f(z) \mathrm{d}z = c_{-1} \tag{B.3}$$

where C is a circle around z_0 and c_{-1} is the coefficient of $(z - z_0)^{-1}$ in the Laurent series representation of f as in (B.2).

Proof Parametrizing the circle C by $C = \{z : z = z_0 + re^{j2\pi t};\ \forall t \in [0,1], r > 0\}$ so that $\mathrm{d}z = 2\pi j(z - z_0)\mathrm{d}t$ and then integrating the Laurent series representation of f leads to

$$\frac{1}{2\pi j}\oint_C f(z)\mathrm{d}z = \frac{1}{2\pi j}\sum_{k=-\infty}^{\infty} c_k \oint_C (z - z_0)^k \mathrm{d}z = \sum_{k=-\infty}^{\infty} c_k \oint_C (re^{j2\pi t})^{k+1}\mathrm{d}t = c_{-1}$$

B.1.4 Identity Theorem

If a function f that is analytic on an open and connected domain $D \subset \mathbb{C}$ vanishes on a small region $R \subset D$, then it also vanishes on the entire domain D:

$$f(R) = 0 \Leftrightarrow f(D) = 0 \text{ , if } f \text{ analytic on } D \text{ and } R \subset D \tag{B.4}$$

and $f(D)$ is then the analytic continuation of $f(R)$.

Proof From $R \subset D$ the direction $f(D) = 0 \Rightarrow f(R) = 0$ follows immediately. To verify the reverse direction, assume $f(R) = 0$. Then due to (B.1) the Taylor series on $z \in R$ around an arbitrary $z_0 \in R$ is given by $f(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\partial^k f(z_0)}{\partial^k z}(z - z_0)^k = 0$ for all $z \in R$. This can be derived so that $\frac{\partial^l f(z)}{\partial^l z} = \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\partial^{k+l} f(z_0)}{\partial^{k+l} z}(z - z_0)^k = 0$ for all $z \in R$ and for all $l \in \mathbb{Z}_+$ so that we get $\frac{\partial^k f(z_0)}{\partial^k z} = 0$ for all $k \in \mathbb{Z}_+$. Since f is analytic on D the Taylor series on $z \in D$ around z_0 is the same, so that $f(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\partial^k f(z_0)}{\partial^k z}(z - z_0)^k = 0$ since the gradients are all zero in z_0 and thus (B.4) is verified.

B.1.5 Open Mapping Theorem

If a function f is analytic on an open and connected domain $D \subset \mathbb{C}$ then either $f(D)$ is also open or f is constant:

$$D \text{ open } \Rightarrow \text{ either } f(D) \text{ open or } f \text{ constant, if } f \text{ analytic on } D \tag{B.5}$$

Proof Define an infinitesimally small circular region $R(z_0) = \{z : z - z_0 = \epsilon e^{j\phi}\}$ around $z_0 \in D$. Since $R(z_0)$ is infinitesimally small and f is analytic, the Taylor series of f around z_0 can be reduced to $f(R(z_0)) = f(z_0) + \frac{\partial f(z_0)}{\partial z}(z - z_0) = f(z_0) + \frac{\partial f(z_0)}{\partial z}\epsilon e^{j\phi}$ and with $\frac{\partial f(z_0)}{\partial z} = ge^{j\phi_0}$ it can be seen that $f(R(z_0)) - f(z_0) = g\epsilon e^{j(\phi+\phi_0)}$. On the one hand if $g \neq 0$, $f(R(z_0))$ also defines an infinitesimally small circular region around $f(z_0)$ and therefore $f(z_0)$ cannot be a boundary point. This reasoning is valid for every $z_0 \in D$ so that $f(D)$ must be open. On the other hand if $g = 0$ then $f(R(z_0)) - f(z_0) = 0$ and using (B.4) we get $f(D) - f(z_0) = 0$ and thus f is constant.

B.1.6 Maximum Modulus Theorem

If a function f is analytic on an open and connected domain $D \subset \mathbb{C}$ then its modulus $|f|$ is maximal at the closure ∂D of the domain D:

$$\max_{z \in \bar{D}} |f(z)| = \max_{z \in \partial D} |f(z)| \text{ , if } f \text{ analytic on } D \tag{B.6}$$

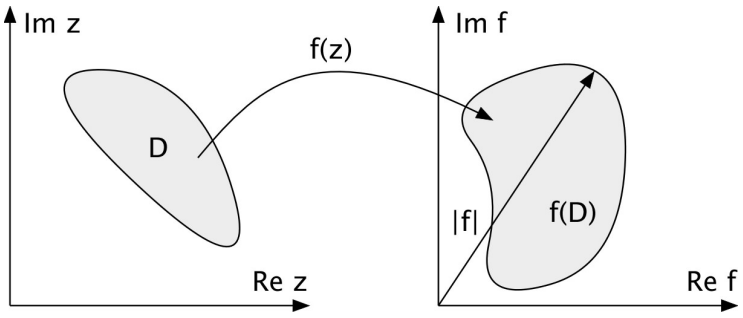


Figure B.1: The modulus of f has its extrema on the closure of $f(D)$.

Proof As can be seen in figure B.1, the modulus $|f|$ has its maximum on the boundary ∂D and (B.5) states that either $f(D)$ is open or f is constant. So either a $z \in \partial D$ provides a maximal $|f|$ and (B.6) is verified or f is constant and (B.6) is also true.

B.2 Norms

This section introduces the norms as a universal measure of length for vectors, matrices, signals and systems.

B.2.1 Norms for Constants

The matrix and induced matrix norms measure the length of constant matrices and transformation matrices.

The p Norm

For a constant matrix $\boldsymbol{A} \in \mathbb{C}^{m,n}$ the p norm is defined by

$$||\boldsymbol{A}||_p = \sqrt[p]{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p} \tag{B.7}$$

Usually $p \in \{1, 2, \infty\}$ and in the case $p = \infty$ the p norm is given by

$$||\boldsymbol{A}||_\infty = \max_{i\in\mathbb{N}_m,\ j\in\mathbb{N}_n} |a_{ij}| \tag{B.8}$$

The ip Norm

For a linear transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A} \in \mathbb{C}^{m,n}$ the ip norm is defined by

$$||\boldsymbol{A}||_{ip} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{||\boldsymbol{y}||_p}{||\boldsymbol{x}||_p} = \max_{||\boldsymbol{x}||_p=1} ||\boldsymbol{A}\boldsymbol{x}||_p \tag{B.9}$$

and the ip norm is induced by the matrix p norm. Usually $p \in \{1, 2, \infty\}$ and in these cases, the ip norm is given by

$$||\boldsymbol{A}||_{i1} = \max_{j \in \mathbb{N}_n} \sum_{i=1}^{m} |a_{ij}| \tag{B.10}$$

$$||\boldsymbol{A}||_{i2} = \sigma_{\max} \tag{B.11}$$

$$||\boldsymbol{A}||_{i\infty} = \max_{i \in \mathbb{N}_m} \sum_{j=1}^{n} |a_{ij}| \tag{B.12}$$

whereby $\sigma_{\max}$ is the largest singular value of $\boldsymbol{A}$.

B.2.2 Norms for Functions with Real Variables

The Lebesgue norm measures the length of a function that depends on a real or integer variable.

The Lp Norm

For continuous function f_c and discrete function f_d defined on domains $D_c \subset \mathbb{R}, D_d \subset \mathbb{Z}$ respectively, the Lp norm is defined by

$$\begin{aligned} ||f_c(D_c)||_{Lp} &= \sqrt[p]{\int_{D_c} ||f_c(t)||^p \mathrm{d}t} \\ ||f_d[D_d]||_{Lp} &= \sqrt[p]{\sum_{D_d} ||f_d[k]||^p} \end{aligned} \tag{B.13}$$

Any matrix norm of $f_c(t), f_d[k]$ can be used. Usually $p \in \{1, 2, \infty\}$ and in the case $p = \infty$ the Lp norm is given by

$$\begin{aligned} ||f_c(D_c)||_{L\infty} &= \sup_{t \in D_c} ||f_c(t)|| \\ ||f_d[D_d]||_{L\infty} &= \sup_{k \in D_d} ||f_d[k]|| \end{aligned} \tag{B.14}$$

B.2.3 Norms for Functions with Complex Variables

The Hardy norm measures the length of a function that depends on a complex variable.

The Hp Norm

For a function F analytic on a domain $D \subset \mathbb{C}$ the Hp norm is defined by

$$||F(D)||_{Hp} = ||F(\partial D)||_{Lp} \tag{B.15}$$

Proof This is due to the matrix version of (B.6) which states that $||F||$ is largest on the boundary ∂D and with the boundary parametrized by a $t \in \mathbb{R}$, the problem of finding the largest $||F||$ over a complex domain reduces to the Lp problem of finding the largest $||F||$ over a real domain.

Example In the case of F_c analytic on $D_c = \{s \in \mathbb{C} : \operatorname{Re} s \geq 0\}$ and F_d analytic on $D_d = \{z \in \mathbb{C} : |z| \geq 1\}$ the Hp norm is given by

$$\begin{aligned} ||F_c(D_c)||_{Hp} &= ||F_c(\partial D_c)||_{Lp} = ||F_c(j2\pi\mathbb{R})||_{Lp} \\ ||F_d(D_d)||_{Hp} &= ||F_d(\partial D_d)||_{Lp} = \left|\left|F_d(e^{j2\pi\mathbb{R}})\right|\right|_{Lp} \end{aligned} \tag{B.16}$$

The examples are important in the study of continuous (index c) and discrete (index d) stable systems, where the system transfer function is analytic in D_c and D_d respectively. The factor 2π appearing in these definitions stems from the normalization condition provided by the Parseval Identity (C.9).

B.3 Optimization

This section provides some helpful results in the domain of constrained optimization, with the constraint given in the form of a matrix inequality. Details can be found in [4].

B.3.1 Problem Formulation

An optimization problem of minimizing a scalar function f under matrix inequality constraint $\boldsymbol{G} = \boldsymbol{G}^H$ may be formulated in the following way

$$f(\boldsymbol{x}^*) = \inf_{\boldsymbol{x} \in F} f(\boldsymbol{x}) \text{ with } F = \{\boldsymbol{x} \in X : \boldsymbol{G}(\boldsymbol{x}) \leq 0\} \tag{B.17}$$

with $\boldsymbol{x}$ being the vector of variables over a space X. The set F is the feasible set, the set of all $\boldsymbol{x}$ that satisfy the constraint $\boldsymbol{G}$. A maximization of f can be formulated as a minimization of $-f$.

B.3.2 Lagrange Function

The constrained problem (B.17) can be formulated as an equivalent unconstrained problem

$$f(\boldsymbol{x}^*) = \inf_{\boldsymbol{x} \in X} \sup_{\boldsymbol{Y} \geq 0} L(\boldsymbol{x}, \boldsymbol{Y}) \tag{B.18}$$

with the Lagrange multiplier $\boldsymbol{Y} = \boldsymbol{Y}^H \geq 0$ and the Lagrange function defined as

$$L(\boldsymbol{x}, \boldsymbol{Y}) = f(\boldsymbol{x}) + \boldsymbol{Y} \bullet \boldsymbol{G}(\boldsymbol{x}) \tag{B.19}$$

Proof This stems from the fact that

$$\sup_{\boldsymbol{Y} \geq 0} L(\boldsymbol{x}, \boldsymbol{Y}) = f(\boldsymbol{x}) + \begin{cases} 0; & \boldsymbol{G}(\boldsymbol{x}) \leq 0 \\ \infty; & \text{else} \end{cases}$$

since in the case $\boldsymbol{G}(\boldsymbol{x}) \leq 0$ the upper limit follows by applying (A.75) on the hermitian matrices $\boldsymbol{Y} \geq 0, -\boldsymbol{G} \geq 0$ and in the other case there is no upper limit to the scalar product $\boldsymbol{Y} \bullet \boldsymbol{G}$. This means that in order to get $L(\boldsymbol{x}, \boldsymbol{Y})$ minimal, $\boldsymbol{x}$ must be in the feasible set F so that $\sup_{\boldsymbol{Y} \geq 0} L(\boldsymbol{x}, \boldsymbol{Y}) = f(\boldsymbol{x})$ and hence (B.18) is equivalent to (B.17) in the sense that they share the same minimizing $\boldsymbol{x}^*$ and the same minimal value $f(\boldsymbol{x}^*)$.

B.3.3 Primal and Dual Problem

To the primal optimization problem (B.17) and equivalently (B.18), a dual problem in the form of

$$g(\boldsymbol{Y}^*) = \sup_{\boldsymbol{Y}\geq 0} \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{Y}) \tag{B.20}$$

can be assigned due to the symmetry to (B.18). In general

$$f(\boldsymbol{x}^*) \geq g(\boldsymbol{Y}^*) \Leftrightarrow \text{weak duality} \tag{B.21}$$

which is called weak duality. In the case

$$f(\boldsymbol{x}^*) = g(\boldsymbol{Y}^*) \Leftrightarrow \text{strong duality} \tag{B.22}$$

which is called strong duality, the Lagrange function L has a saddle point at $L(\boldsymbol{x}^*, \boldsymbol{Y}^*) = f(\boldsymbol{x}^*) = g(\boldsymbol{Y}^*)$. This case is of high practical importance which is due to the large number of problems that can be found to exhibit strong duality and the performance of the solvers that specialize on these problems. It also allows the parameter perturbation analysis.

B.3.4 First Order Perturbation Analysis

The problem (B.17) may depend on parameters $\boldsymbol{z}$ that can be made explicit in the form of

$$f(\boldsymbol{x}^*, \boldsymbol{z}) = \inf_{\boldsymbol{x}\in F(\boldsymbol{z})} f(\boldsymbol{x}, \boldsymbol{z}) \text{ with } F(\boldsymbol{z}) = \{\boldsymbol{x} \in X : \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{z}) \leq 0\} \tag{B.23}$$

in order to allow the analysis of the effects of small parameter changes on the optimal value $f(\boldsymbol{x}^*, \boldsymbol{z})$. In the case of strong duality (B.22) an upper limit on the first order effects on the optimal value can be given by

$$f(\boldsymbol{x}^*, \boldsymbol{z}) - f(\boldsymbol{x}_0^*, \boldsymbol{z}_0) = \mathrm{d}f \leq \delta L = \frac{\partial L(\boldsymbol{x}_0^*, \boldsymbol{Y}_0^*, \boldsymbol{z}_0)}{\partial \boldsymbol{z}} \delta \boldsymbol{z} \tag{B.24}$$

with $\boldsymbol{x}_0^*$ and $\boldsymbol{Y}_0^*$ being the solution of the primal and dual problems for a nominal parameter vector $\boldsymbol{z}_0$ and $\boldsymbol{x}^*$ and $\boldsymbol{Y}^*$ being the solution of the primal and dual problems for a perturbed parameter vector $\boldsymbol{z} = \boldsymbol{z}_0 + \delta\boldsymbol{z}$. The parameter dependent Lagrange function is defined as

$$L(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{z}) = f(\boldsymbol{x}, \boldsymbol{z}) + \boldsymbol{Y} \bullet \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{z}) \tag{B.25}$$

in analogy to (B.19).

Proof This upper limit can be verified by using strong duality (B.22) and (B.20) to get

$$f(\boldsymbol{x}^*, \boldsymbol{z}) = L(\boldsymbol{x}^*, \boldsymbol{Y}^*, \boldsymbol{z}) = \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{Y}^*, \boldsymbol{z}) \leq L(X, \boldsymbol{Y}^*, \boldsymbol{z})$$

For $\boldsymbol{x}_0^* \in X$ this leads to

$$f(\boldsymbol{x}^*, \boldsymbol{z}) \leq L(\boldsymbol{x}_0^*, \boldsymbol{Y}^*, \boldsymbol{z}) = \\ = \underbrace{L(\boldsymbol{x}_0^*, \boldsymbol{Y}_0^*, \boldsymbol{z}_0)}_{f(\boldsymbol{x}_0^*, \boldsymbol{z}_0)} + \underbrace{\frac{\partial L(\boldsymbol{x}_0^*, \boldsymbol{Y}_0^*, \boldsymbol{z}_0)}{\partial \boldsymbol{Y}} \delta \boldsymbol{Y}}_{\delta \boldsymbol{Y} \bullet \boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0)} + \frac{\partial L(\boldsymbol{x}_0^*, \boldsymbol{Y}_0^*, \boldsymbol{z}_0)}{\partial \boldsymbol{z}} \delta \boldsymbol{z}$$

with $\boldsymbol{Y}^* = \boldsymbol{Y}_0^* + \delta\boldsymbol{Y}$. An upper limit for the scalar product can be given by

$$\delta\boldsymbol{Y} \bullet \boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0) = \underbrace{\overbrace{\boldsymbol{Y}^*}^{\geq 0} \bullet \overbrace{\boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0)}^{\leq 0}}_{\leq 0} - \underbrace{\boldsymbol{Y}_0^* \bullet \boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0)}_{=0}$$

The first term is non positive due to (A.75) applied on the hermitian matrices $\boldsymbol{Y}^* \geq 0, -\boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0) \geq 0$. The second term is zero because with strong duality we have

$$f(\boldsymbol{x}_0^*, \boldsymbol{z}_0) = L(\boldsymbol{x}_0^*, \boldsymbol{Y}_0^*, \boldsymbol{z}_0) = f(\boldsymbol{x}_0^*, \boldsymbol{z}_0) + \boldsymbol{Y}_0^* \bullet \boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0)$$

so that $\delta\boldsymbol{Y} \bullet \boldsymbol{G}(\boldsymbol{x}_0^*, \boldsymbol{z}_0) \leq 0$ and (B.24) follows. A more rigorous proof and further results can be found in [4].

Appendix C

Systems

C.1 Linear Systems

This section provides some important characteristics of linear time invariant systems.

C.1.1 System Representations

A linear time invariant system with input $\boldsymbol{u} \in \mathbb{R}^m$, output $\boldsymbol{y} \in \mathbb{R}^n$ and state $\boldsymbol{x} \in \mathbb{C}^p$ may be time continuous or discrete. Continuous (discrete) systems can be represented by a set of linear first order differential (difference) equations in the following standard form

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) \\ \boldsymbol{y}(t) &= \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t) \end{aligned} \tag{C.1}$$

for continuous systems and

$$\begin{aligned} \boldsymbol{x}[k+1] &= \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{B}\boldsymbol{u}[k] \\ \boldsymbol{y}[k] &= \boldsymbol{C}\boldsymbol{x}[k] + \boldsymbol{D}\boldsymbol{u}[k] \end{aligned} \tag{C.2}$$

for discrete systems. These are the state space representations in time domain of these systems. The matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are often called *dynamic*, *input*, *output* and *throughput* matrices respectively. These systems being linear, there exist integral transforms

$$\boldsymbol{x}(s) = \int_{\mathbb{R}_+} \boldsymbol{x}(t) e^{-st} \mathrm{d}t \tag{C.3}$$

for continuous systems and

$$\boldsymbol{x}(z) = \sum_{\mathbb{Z}_+} \boldsymbol{x}[k] z^{-k} \tag{C.4}$$

for discrete systems, that transform these systems to a set of linear algebraic equations

$$\boldsymbol{y}(s) = \boldsymbol{P}(s)\boldsymbol{u}(s) \tag{C.5}$$

for continuous systems and

$$\boldsymbol{y}(z) = \boldsymbol{P}(z)\boldsymbol{u}(z) \tag{C.6}$$

for discrete systems. These are the transfer function representations in frequency domain of these systems. The matrix of transfer functions $\boldsymbol{P}$ is called the transfer matrix. The link between state space and transfer function representation can be given by

$$\boldsymbol{P}(x) = \boldsymbol{C}(x\mathbf{1} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D} =: \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right] \tag{C.7}$$

with $x \in \{s, z\}$ for continuous and discrete systems respectively. There is also a link between a continuous and a discrete system given by

$$\boldsymbol{x}[k] = \boldsymbol{x}(kT) \text{ and } z = e^{sT} \tag{C.8}$$

with T being the sample time of the discrete system.

Proof This can be seen by approximating the integral in (C.3) as a sum

$$\boldsymbol{x}(s) \approx \sum_{\mathbb{Z}_+} \boldsymbol{x}(kT)e^{-skT}$$

and comparing this to (C.4).

C.1.2 Parseval Theorem

The transformations (C.3) and (C.4) relate the L2 norm of continuous $\boldsymbol{x}_c(t)$ and discrete $\boldsymbol{x}_d[k]$ to the H2 norm of their respective spectra $\boldsymbol{x}_c(s)$ and $\boldsymbol{x}_d(z)$ in the form of

$$\begin{aligned} ||\boldsymbol{x}_c(t)||_{L2} &= ||\boldsymbol{x}_c(s)||_{H2} \\ ||\boldsymbol{x}_d[k]||_{L2} &= ||\boldsymbol{x}_d(z)||_{H2} \end{aligned} \tag{C.9}$$

which is called the *Parseval theorem*.

Proof This can be verified for a discrete signal $\boldsymbol{x}[k]$. According to (B.16) the square of the H2 norm of $\boldsymbol{x}(z)$ is given by

$$||\boldsymbol{x}(z)||_{H2}^2 = \left|\left|\boldsymbol{x}(e^{j2\pi t})\right|\right|_{L2}^2 = \int_0^1 \left|\left|\boldsymbol{x}(e^{j2\pi t})\right|\right|^2 \mathrm{d}t$$

The 2 norm (B.7) is used so that $\left|\left|\boldsymbol{x}(e^{j2\pi t})\right|\right|^2 = \boldsymbol{x}^H(e^{j2\pi t})\boldsymbol{x}(e^{j2\pi t})$. With the transformation (C.4) this leads to

$$||\boldsymbol{x}(z)||_{H2}^2 = \int_0^1 \left(\sum_{k=0}^{\infty} \boldsymbol{x}^H[k]e^{j2\pi kt}\right)\left(\sum_{l=0}^{\infty} \boldsymbol{x}[k]e^{-j2\pi kt}\right)\mathrm{d}t =$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \boldsymbol{x}^H[k]\boldsymbol{x}[l]\int_0^1 e^{j2\pi(k-l)t}\mathrm{d}t$$

Note that

$$\int_0^1 e^{j2\pi kt}\mathrm{d}t = \begin{cases} 1; & k = 0 \\ 0; & k \in \mathbb{Z} \setminus \{0\} \end{cases}$$

therefore

$$||\boldsymbol{x}(z)||_{H2}^2 = \sum_{k=0}^{\infty} \boldsymbol{x}^H[k]\boldsymbol{x}[k] = ||\boldsymbol{x}[k]||_{L2}^2$$

and (C.9) follows for $\boldsymbol{x}_d = \boldsymbol{x}$. The proof for continuous systems is similar.

C.1.3 L2/H2 System Norm

The L2/H2 norm of a discrete system $\boldsymbol{P}$ as in (C.2) can be given explicitly by

$$||\boldsymbol{P}[k]||_{L2} = ||\boldsymbol{P}(z)||_{H2} = \sqrt{\mathrm{tr}(\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D})} \tag{C.10}$$

with $\boldsymbol{G}$ being the Gram observability matrix that according to (C.34) solves

$$\boldsymbol{A}^H\boldsymbol{G}\boldsymbol{A} - \boldsymbol{G} + \boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{0} \tag{C.11}$$

using the 2 norm as in (B.7). This can also be stated in inequality form by

$$||\boldsymbol{P}[k]||_{L2} = ||\boldsymbol{P}(z)||_{H2} < \gamma \tag{C.12}$$

if and only if there exists a $\boldsymbol{Z} > 0$ so that

$$\begin{aligned} &\boldsymbol{A}^H\boldsymbol{Z}\boldsymbol{A} - \boldsymbol{Z} + \boldsymbol{C}^H\boldsymbol{C} < 0 \\ &\mathrm{tr}(\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D}) < \gamma^2 \end{aligned} \tag{C.13}$$

Proof Due to (C.9) we have $||\boldsymbol{P}[k]||_{L2} = ||\boldsymbol{P}(z)||_{H2}$. The impulse response $\boldsymbol{P}[k]$ of the discrete system (C.2) is given by

$$\boldsymbol{P}[k] = \begin{cases} \boldsymbol{D}; & k = 0 \\ \boldsymbol{C}\boldsymbol{A}^{k-1}\boldsymbol{B}; & k > 0 \end{cases}$$

so that the L2 norm can be calculated directly

$$\begin{aligned} ||\boldsymbol{P}[k]||_{L2}^2 = \sum_{k=0}^{\infty} ||\boldsymbol{P}[k]||^2 = \mathrm{tr}(\boldsymbol{D}^H\boldsymbol{D}) + \sum_{k=0}^{\infty} \mathrm{tr}(\boldsymbol{B}^H{\boldsymbol{A}^H}^k\boldsymbol{C}^H\boldsymbol{C}\boldsymbol{A}^k\boldsymbol{B}) = \\ = \mathrm{tr}(\boldsymbol{D}^H\boldsymbol{D} + \boldsymbol{B}^H\left(\sum_{k=0}^{\infty} {\boldsymbol{A}^H}^k\boldsymbol{C}^H\boldsymbol{C}\boldsymbol{A}^k\right)\boldsymbol{B}) \end{aligned}$$

according to (B.13) while using the 2 norm (B.7). It can be easily verified that the series $\boldsymbol{G} = \sum_{k=0}^{\infty} {\boldsymbol{A}^H}^k\boldsymbol{C}^H\boldsymbol{C}\boldsymbol{A}^k$ solves the Stein equation (C.11) so that the result (C.10) is obtained.

Now assume $\boldsymbol{A}$ stable and (C.12). Let $\boldsymbol{Z}(\epsilon) = \sum_{k=0}^{\infty} {\boldsymbol{A}^H}^k(\boldsymbol{C}^H\boldsymbol{C} + \epsilon\boldsymbol{1})\boldsymbol{A}^k$ so that $\boldsymbol{Z}(0) = \boldsymbol{G}$. Since $\boldsymbol{Z}(\epsilon)$ is a continuous function there exists a small enough $\epsilon_0 > 0$ so that $\mathrm{tr}(\boldsymbol{B}^H\boldsymbol{Z}(\epsilon_0)\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D}) < \gamma^2$. It can be easily verified that $\boldsymbol{Z}(\epsilon_0)$ solves the Stein equation $\boldsymbol{A}^H\boldsymbol{Z}(\epsilon_0)\boldsymbol{A} - \boldsymbol{Z}(\epsilon_0) + \boldsymbol{C}^H\boldsymbol{C} + \epsilon_0\boldsymbol{1} = \boldsymbol{0}$. Due to $\boldsymbol{A}$ stable and $\boldsymbol{C}^H\boldsymbol{C} + \epsilon_0\boldsymbol{1} > 0$ we get $\boldsymbol{Z}(\epsilon_0) > 0$ according to (A.89).

Thus $\boldsymbol{A}^H\boldsymbol{Z}(\epsilon_0)\boldsymbol{A} - \boldsymbol{Z}(\epsilon_0) + \boldsymbol{C}^H\boldsymbol{C} < 0$ so that (C.13) follows. In order to verify the reverse direction assume (C.13). Then there exists an $\boldsymbol{E} > 0$ so that $\boldsymbol{A}^H\boldsymbol{Z}\boldsymbol{A} - \boldsymbol{Z} + \boldsymbol{C}^H\boldsymbol{C} + \boldsymbol{E} = \boldsymbol{0}$. Since $\boldsymbol{C}^H\boldsymbol{C} + \boldsymbol{E} > 0$ and $\boldsymbol{Z} > 0$ the dynamic matrix $\boldsymbol{A}$ is stable and the L2/H2 norm is given by (C.10) with Gram observability matrix $\boldsymbol{G}$ being the unique positive definite solution of (C.11). Thus $\boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{G} - \boldsymbol{A}^H\boldsymbol{G}\boldsymbol{A}$ so that $\boldsymbol{A}^H(\boldsymbol{Z}-\boldsymbol{G})\boldsymbol{A} - (\boldsymbol{Z}-\boldsymbol{G}) + \boldsymbol{E} = \boldsymbol{0}$. Since $\boldsymbol{A}$ is stable and $\boldsymbol{E} > 0$ this Stein equation has a unique positive definite solution $\boldsymbol{Z}-\boldsymbol{G} > 0$ so that $\gamma^2 > \mathrm{tr}(\boldsymbol{B}^H\boldsymbol{Z}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D}) \geq \mathrm{tr}(\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D})$ and (C.12) follows.

Remark A dual result to (C.10), (C.11) can be obtained using the Gram controllability matrix. The L2/H2 norm of $\boldsymbol{P}$ as in (C.2) can be given by

$$||\boldsymbol{P}[k]||_{L2} = ||\boldsymbol{P}(z)||_{H2} = \sqrt{\mathrm{tr}(\boldsymbol{B}^H\boldsymbol{G}\boldsymbol{B} + \boldsymbol{D}^H\boldsymbol{D})}$$

with $\boldsymbol{G}$ being the Gram controllability matrix that according to (C.32) solves

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^H - \boldsymbol{G} + \boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{0}$$

using the 2 norm as in (B.7).

C.1.4 Star Norm

The concept of the star norm used in section 4.1.2 as an upper bound on the L1 norm was developed in [10] and thoroughly treated in [1] for the continuous case. The discrete case was treated in [9], [12]. An alternative proof of the discrete case matrix inequalities based on the concept given in [9] is presented in the following.

Proposition Consider $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}$ and let X, Y, Z be the spaces of $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ respectively and define

$$\begin{aligned} X_R &= \{\boldsymbol{x} : \ \boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} \leq 1\} \\ Y_S &= \{\boldsymbol{y} : \ \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \leq 1\} \\ Z_T &= \{\boldsymbol{z} : \ \boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} \leq 1\} \end{aligned} \tag{C.14}$$

with $\boldsymbol{R} \geq 0, \boldsymbol{S} \geq 0, \boldsymbol{T} \geq 0$ but $\boldsymbol{R} \neq 0, \boldsymbol{S} \neq 0$. Then

$$\forall \boldsymbol{x} \in X_R, \forall \boldsymbol{y} \in Y_S : \ \boldsymbol{z} \in Z_T \Leftrightarrow$$

$$\Leftrightarrow \forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y : \ \boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} \leq \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \Rightarrow \boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} \leq \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \tag{C.15}$$

$$\Leftrightarrow \forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y : \ \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \leq \boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} \Rightarrow \boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} \leq \boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} \tag{C.16}$$

$$\Leftrightarrow \forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y, \exists \alpha \in [0,1] : \ \boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} \leq \alpha\boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} + (1-\alpha)\boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \tag{C.17}$$

$$\Leftrightarrow \exists \alpha \in [0,1] : \ \begin{bmatrix} \boldsymbol{A}^H\boldsymbol{T}\boldsymbol{A} - \alpha\boldsymbol{R} & \boldsymbol{A}^H\boldsymbol{T}\boldsymbol{B} \\ \boldsymbol{B}^H\boldsymbol{T}\boldsymbol{A} & \boldsymbol{B}^H\boldsymbol{T}\boldsymbol{B} - (1-\alpha)\boldsymbol{S} \end{bmatrix} \leq 0 \tag{C.18}$$

Proof On the one hand suppose $\exists \boldsymbol{x} \in X, \exists \boldsymbol{y} \in Y : \boldsymbol{x}^H\boldsymbol{R}\boldsymbol{x} \leq \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y} \Rightarrow \boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} > \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y}$. Hence there exists a $\beta > 0$ so that $\boldsymbol{z}^H\boldsymbol{T}\boldsymbol{z} > \beta^2 > \boldsymbol{y}^H\boldsymbol{S}\boldsymbol{y}$. Now define $\bar{\boldsymbol{x}} = \boldsymbol{x}/\beta, \bar{\boldsymbol{y}} = \boldsymbol{y}/\beta, \bar{\boldsymbol{z}} = \boldsymbol{z}/\beta$ then $\bar{\boldsymbol{z}} = \boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{z}}^H\boldsymbol{T}\bar{\boldsymbol{z}} > 1$ as well

as $\bar{\boldsymbol{x}}^H \boldsymbol{R} \bar{\boldsymbol{x}} < 1$ and $\bar{\boldsymbol{y}}^H \boldsymbol{S} \bar{\boldsymbol{y}} < 1$. Thus $\exists \bar{\boldsymbol{x}} \in X_R, \exists \bar{\boldsymbol{y}} \in Y_S : \ \bar{\boldsymbol{z}} \notin Z_T$ which verifies sufficiency in (C.15).

On the other hand suppose $\forall \boldsymbol{x} \in X, \forall \boldsymbol{y} \in Y : \boldsymbol{x}^H \boldsymbol{R} \boldsymbol{x} \leq \boldsymbol{y}^H \boldsymbol{S} \boldsymbol{y} \Rightarrow \boldsymbol{z}^H \boldsymbol{T} \boldsymbol{z} \leq \boldsymbol{y}^H \boldsymbol{S} \boldsymbol{y}$. Using $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}$ this translates into

$$\begin{aligned} &\forall \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \in X \times Y : \\ &\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^H \begin{bmatrix} -\boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \geq 0 \Rightarrow \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^H \begin{bmatrix} -\boldsymbol{A}^H \boldsymbol{T} \boldsymbol{A} & -\boldsymbol{A}^H \boldsymbol{T} \boldsymbol{B} \\ \boldsymbol{B}^H \boldsymbol{T} \boldsymbol{A} & \boldsymbol{S} - \boldsymbol{B}^H \boldsymbol{T} \boldsymbol{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \geq 0 \end{aligned} \tag{C.19}$$

Since $\boldsymbol{R} \geq 0, \boldsymbol{S} \geq 0$ but $\boldsymbol{S} \neq 0$ there

$$\exists \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \in X \times Y : \ \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^H \begin{bmatrix} -\boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} > 0 \tag{C.20}$$

Due to (C.20) the S-lemma (A.66) can be applied on (C.19) which then becomes equivalent to

$$\exists \alpha \geq 0 : \ \begin{bmatrix} \boldsymbol{A}^H \boldsymbol{T} \boldsymbol{A} - \alpha \boldsymbol{R} & \boldsymbol{A}^H \boldsymbol{T} \boldsymbol{B} \\ \boldsymbol{B}^H \boldsymbol{T} \boldsymbol{A} & \boldsymbol{B}^H \boldsymbol{T} \boldsymbol{B} - (1-\alpha) \boldsymbol{S} \end{bmatrix} \leq 0$$

With $\boldsymbol{B}^H \boldsymbol{T} \boldsymbol{B} \geq 0$ and $\boldsymbol{S} \geq 0$ and by applying (A.63) it can be deduced from the lower right block that $\alpha \leq 1$ so that the right hand side of (C.18) follows. This matrix inequality form is equivalent to the right hand side of (C.17) from where it can be seen that $\forall \boldsymbol{x} \in X_R, \forall \boldsymbol{y} \in Y_S : \ \boldsymbol{z} \in Z_T$ which verifies necessity in (C.15). Condition (C.16) then follows from the symmetry in (C.17) with respect to $\boldsymbol{x}, \boldsymbol{R}$ and $\boldsymbol{y}, \boldsymbol{S}$.

Definition Consider the discrete system (C.2) written shortly as

$$\begin{aligned} \boldsymbol{x}_+ &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \\ \boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u} \end{aligned} \tag{C.21}$$

and let U, X, Y be the input, state and output spaces. Define

$$\begin{aligned} U_1 &= \{\boldsymbol{u} : \ \boldsymbol{u}^H \boldsymbol{u} \leq 1\} \\ X_Z &= \{\boldsymbol{x} : \ \boldsymbol{x}^H \boldsymbol{Z} \boldsymbol{x} \leq 1\} \\ Y_Z &= \{\boldsymbol{y} : \ \boldsymbol{y}^H \frac{1}{\gamma^2} \boldsymbol{y} \leq 1\} \end{aligned} \tag{C.22}$$

with $\boldsymbol{Z} \geq 0, \gamma > 0$ but $\boldsymbol{Z} \neq 0$. Then the ellipsoid

$$\begin{aligned} X_Z \text{ is inescapable} &\Leftrightarrow \forall \boldsymbol{u} \in U_1, \forall \boldsymbol{x} \in X_Z : \ \boldsymbol{x}_+ \in X_Z \\ Y_Z \text{ is inescapable} &\Leftrightarrow \forall \boldsymbol{u} \in U_1, \forall \boldsymbol{x} \in X_Z : \ \boldsymbol{y} \in Y_Z \end{aligned} \tag{C.23}$$

Proposition Consider the discrete system (C.21) and the ellipsoids as defined in (C.22). Then the ellipsoids X_Z, Y_Z are inescapable if and only if

$$\begin{aligned} &\exists \alpha \in [0,1] : \begin{bmatrix} \boldsymbol{A}^H \boldsymbol{Z} \boldsymbol{A} - (1-\alpha) \boldsymbol{Z} & \boldsymbol{A}^H \boldsymbol{Z} \boldsymbol{B} \\ \boldsymbol{B}^H \boldsymbol{Z} \boldsymbol{A} & \boldsymbol{B}^H \boldsymbol{Z} \boldsymbol{B} - \alpha \boldsymbol{1} \end{bmatrix} \leq 0 \\ &\exists \beta \in [0, \gamma^2] : \begin{bmatrix} \boldsymbol{C}^H \boldsymbol{C} - \beta \boldsymbol{Z} & \boldsymbol{C}^H \boldsymbol{D} \\ \boldsymbol{D}^H \boldsymbol{C} & \boldsymbol{D}^H \boldsymbol{D} - (\gamma^2 - \beta) \boldsymbol{1} \end{bmatrix} \leq 0 \end{aligned} \tag{C.24}$$

Proof From the definition (C.23) and the definitions (C.22) the ellipsoids X_Z, Y_Z are inescapable if and only if

$$\forall \boldsymbol{x} \in X, \forall \boldsymbol{u} \in U, \exists \alpha' \in [0,1] : \boldsymbol{x}_+^H \boldsymbol{Z} \boldsymbol{x}_+ \leq \alpha' \boldsymbol{x}^H \boldsymbol{Z} \boldsymbol{x} + (1-\alpha') \boldsymbol{u}^H \boldsymbol{u}$$
$$\forall \boldsymbol{x} \in X, \forall \boldsymbol{u} \in U, \exists \beta' \in [0,1] : \boldsymbol{y}^H \frac{1}{\gamma^2} \boldsymbol{y} \leq \beta' \boldsymbol{x}^H \boldsymbol{Z} \boldsymbol{x} + (1-\beta') \boldsymbol{u}^H \boldsymbol{u}$$

according to (C.17). With (C.18) and the substitutions $\alpha = 1 - \alpha', \beta = \gamma^2 \beta'$ the matrix inequalities in (C.24) follow.

C.1.5 Realness

A physical signal, continuous $\boldsymbol{x}_c(t)$ or discrete $\boldsymbol{x}_d[k]$, in time domain is a measurable real quantity, hence $\boldsymbol{x}_c(t), \boldsymbol{x}_d[k] \in \mathbb{R}$. In frequency domain, this implies

$$\begin{aligned} \overline{\boldsymbol{x}_c(s)} &= \boldsymbol{x}_c(\overline{s}) \\ \overline{\boldsymbol{x}_d(z)} &= \boldsymbol{x}_d(\overline{z}) \end{aligned} \tag{C.25}$$

A linear time invariant continuous or discrete system $\boldsymbol{P}(x)$ with $x \in \{s, z\}$ and with realization $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ is real if it maps real valued input signals onto real valued output signals. This is true if and only if

$$\overline{\boldsymbol{P}(x)} = \boldsymbol{P}(\overline{x}) \tag{C.26}$$

or equivalently if and only if

$$\left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right] = \left[\begin{array}{c|c} \overline{\boldsymbol{A}} & \overline{\boldsymbol{B}} \\ \hline \overline{\boldsymbol{C}} & \overline{\boldsymbol{D}} \end{array}\right] \tag{C.27}$$

which means that these realizations are equivalent, see (C.35) and hence

$$\exists \boldsymbol{T} : \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \overline{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}} \begin{bmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}^{-1} \tag{C.28}$$

according to (C.36).

Proof The condition (C.25) for signals can be verified with the transformations (C.3) and (C.4) so that

$$\begin{aligned} \overline{\boldsymbol{x}_c(s)} &= \int_{\mathbb{R}_+} \overline{\boldsymbol{x}_c(t)} e^{-\overline{s}t} \mathrm{d}t = \int_{\mathbb{R}_+} \boldsymbol{x}_c(t) e^{-\overline{s}t} \mathrm{d}t = \boldsymbol{x}_c(\overline{s}) \\ \overline{\boldsymbol{x}_d(z)} &= \sum_{\mathbb{Z}_+} \overline{\boldsymbol{x}_d[k]} \overline{z}^{-k} = \sum_{\mathbb{Z}_+} \boldsymbol{x}_d[k] \overline{z}^{-k} = \boldsymbol{x}_d(\overline{z}) \end{aligned}$$

because $\boldsymbol{x}_c(t) = \overline{\boldsymbol{x}_c(t)}, \boldsymbol{x}_d[k] = \overline{\boldsymbol{x}_d[k]}$ since they are real. With the same reasoning a system $\boldsymbol{P}(x)$ continuous or discrete with $x \in \{s, z\}$, is real if and only if for all real inputs $\overline{\boldsymbol{u}(x)} = \boldsymbol{u}(\overline{x})$ it produces real outputs $\overline{\boldsymbol{y}(x)} = \overline{\boldsymbol{P}(x)\boldsymbol{u}(x)} = \overline{\boldsymbol{P}(x)}\boldsymbol{u}(\overline{x}) = \boldsymbol{P}(\overline{x})\boldsymbol{u}(\overline{x}) = \boldsymbol{y}(\overline{x})$. Hence $\overline{\boldsymbol{P}(x)} = \boldsymbol{P}(\overline{x})$ with $x \in \{s, z\}$. Using the link (C.7) this then leads to (C.27).

C.1.6 Stability

A linear time invariant system with dynamic matrix $\boldsymbol{A}$ is said to be stable if and only if

$$\begin{aligned} \operatorname{Re} s < 0 \; \forall s : \det(s\mathbf{1} - \boldsymbol{A}) = 0 \quad & \text{for continuous systems} \\ |z| < 1 \; \forall z : \det(z\mathbf{1} - \boldsymbol{A}) = 0 \quad & \text{for discrete systems} \end{aligned} \tag{C.29}$$

Furthermore is the system stable if and only if for any $\boldsymbol{Q} > 0$ there exists a unique $\boldsymbol{P} > 0$ which is the solution to

$$\begin{aligned} \boldsymbol{A}^H \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{Q} = \mathbf{0} \quad & \text{for continuous systems} \\ \boldsymbol{A}^H \boldsymbol{P}\boldsymbol{A} - \boldsymbol{P} + \boldsymbol{Q} = \mathbf{0} \quad & \text{for discrete systems} \end{aligned} \tag{C.30}$$

which are called Lyapunov matrix equations.

Proof This is due to the solutions of (C.1), (C.2) being respectively given by

$$\boldsymbol{y}(t) = \boldsymbol{C} e^{\boldsymbol{A}t} \left(\boldsymbol{x}(0) + \int_0^t e^{-\boldsymbol{A}\tau} \boldsymbol{B}\boldsymbol{u}(\tau) \mathrm{d}\tau \right) + \boldsymbol{D}\boldsymbol{u}(t) \quad \text{for continuous systems}$$

$$\boldsymbol{y}[k] = \boldsymbol{C}\boldsymbol{A}^k \left(\boldsymbol{x}[0] + \sum_{\kappa=0}^{k-1} \boldsymbol{A}^{-(\kappa+1)} \boldsymbol{B}\boldsymbol{u}[\kappa] \right) + \boldsymbol{D}\boldsymbol{u}[k] \quad \text{for discrete systems}$$

so that the solution only converges and thus the output only remains bounded for bounded inputs if (C.29) is satisfied. Now on the one hand assume $\boldsymbol{A}$ stable then according to (A.87), (A.89) the Sylvester and Stein equations in (C.30) have a unique solution $\boldsymbol{P} > 0$ for any $\boldsymbol{Q} > 0$. On the other hand assume (C.30) and define $V(\boldsymbol{x}) = \boldsymbol{x}^H \boldsymbol{P}\boldsymbol{x}$ which is a Lyapunov function candidate due to $\boldsymbol{P} > 0$ then

$$\begin{aligned} \dot{V}(\boldsymbol{x}) = \boldsymbol{x}^H (\boldsymbol{A}^H \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}) \boldsymbol{x} = -\boldsymbol{x}^H \boldsymbol{Q}\boldsymbol{x} \quad & \text{for continuous systems} \\ \Delta V(\boldsymbol{x}) = \boldsymbol{x}^H (\boldsymbol{A}^H \boldsymbol{P}\boldsymbol{A} - \boldsymbol{P}) \boldsymbol{x} = -\boldsymbol{x}^H \boldsymbol{Q}\boldsymbol{x} \quad & \text{for discrete systems} \end{aligned}$$

so that $V(\boldsymbol{x})$ is a Lyapunov function since $\boldsymbol{Q} > 0$ and $\boldsymbol{A}$ is stable according to Lyapunov stability theory.

C.1.7 Controllability

A linear time invariant system with dynamic matrix $\boldsymbol{A} \in \mathbb{C}^{p,p}$ and input matrix $\boldsymbol{B} \in \mathbb{C}^{p,m}$ is said to be controllable if the future state $\boldsymbol{x}$ can be controlled by system input $\boldsymbol{u}$ which is true if and only if

$$\boldsymbol{A}, \boldsymbol{B} \text{ controllable} \Leftrightarrow \begin{cases} \operatorname{rnk} \begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \cdots & \boldsymbol{A}^{p-1}\boldsymbol{B} \end{bmatrix} = p \\ \boldsymbol{B}^T \boldsymbol{x} \neq \mathbf{0} \; \forall \boldsymbol{x} : \boldsymbol{A}^T \boldsymbol{x} = \lambda \boldsymbol{x} \end{cases} \tag{C.31}$$

Furthermore is the system controllable if and only if the Gram controllability matrix $\boldsymbol{G}$ which is the unique solution to

$$\begin{aligned} \boldsymbol{A}\boldsymbol{G} + \boldsymbol{G}\boldsymbol{A}^H + \boldsymbol{B}\boldsymbol{B}^H = \mathbf{0} \quad & \text{for continuous systems} \\ \boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^H - \boldsymbol{G} + \boldsymbol{B}\boldsymbol{B}^H = \mathbf{0} \quad & \text{for discrete systems} \end{aligned} \tag{C.32}$$

exists and $\boldsymbol{G} > 0$.

C.1.8 Observability

A linear time invariant system with dynamic matrix $\boldsymbol{A} \in \mathbb{C}^{p,p}$ and output matrix $\boldsymbol{C} \in \mathbb{C}^{n,p}$ is said to be observable if the past state $\boldsymbol{x}$ can be observed by system output $\boldsymbol{y}$ which is true if and only if

$$\boldsymbol{A}, \boldsymbol{C} \text{ observable } \Leftrightarrow \begin{cases} \operatorname{rnk}\begin{bmatrix} \boldsymbol{C}^T & \boldsymbol{A}^T\boldsymbol{C}^T & \cdots & \boldsymbol{A}^{T^{p-1}}\boldsymbol{C}^T \end{bmatrix} = p \\ \boldsymbol{C}\boldsymbol{x} \neq \boldsymbol{0}\ \forall \boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \end{cases} \tag{C.33}$$

Furthermore is the system observable if and only if the Gram observability matrix $\boldsymbol{G}$ which is the unique solution to

$$\begin{aligned} \boldsymbol{A}^H\boldsymbol{G} + \boldsymbol{G}\boldsymbol{A} + \boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{0} \quad &\text{for continuous systems} \\ \boldsymbol{A}^H\boldsymbol{G}\boldsymbol{A} - \boldsymbol{G} + \boldsymbol{C}^H\boldsymbol{C} = \boldsymbol{0} \quad &\text{for discrete systems} \end{aligned} \tag{C.34}$$

exists and $\boldsymbol{G} > 0$.

C.1.9 Equivalence

Two state space representations are equivalent if their transfer functions are identical:

$$(\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{C}_1, \boldsymbol{D}_1) \sim (\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{C}_2, \boldsymbol{D}_2) \Leftrightarrow \left[\begin{array}{c|c} \boldsymbol{A}_1 & \boldsymbol{B}_1 \\ \hline \boldsymbol{C}_1 & \boldsymbol{D}_1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{A}_2 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_2 & \boldsymbol{D}_2 \end{array}\right] \tag{C.35}$$

with $x \in \{s, z\}$ for continuous and discrete systems respectively. Equivalent state space representations therefore describe the same system.

C.1.10 Change of Basis

The basis in which the states of a system in state space representation are described can be changed. However, stability, controllability, observability and the transfer matrix $\boldsymbol{P}$ are preserved under basis transformation and the transformed state space representation is equivalent to the original one.

$$\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} & \boldsymbol{T}\boldsymbol{B} \\ \hline \boldsymbol{C}\boldsymbol{T}^{-1} & \boldsymbol{D} \end{array}\right] \tag{C.36}$$

This also means that two realizations are equivalent according to (C.35) if and only if there exists an invertible $\boldsymbol{T}$ so that (C.36) is true.

Proof The equivalence follows from (C.1) or (C.2) if the original state $\boldsymbol{x}$ is replaced by the transformed state $\boldsymbol{\xi} = \boldsymbol{T}\boldsymbol{x}$. The invariances follow by reexamination of (C.29), (C.31), (C.33) and (C.7) with the new state space representation.

C.1.11 Kalman Decomposition

A change of basis can be used to recover the controllable and observable eigenmodes of the system. This special basis transformation is called the Kalman

decomposition that transforms any system $\boldsymbol{P}$ into the form

$$\boldsymbol{P} = \left[\begin{array}{cccc|c} \boldsymbol{A}_{c\bar{o}} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} & \boldsymbol{A}_{14} & \boldsymbol{B}_{c\bar{o}} \\ \boldsymbol{0} & \boldsymbol{A}_{co} & \boldsymbol{0} & \boldsymbol{A}_{24} & \boldsymbol{B}_{co} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\bar{c}\bar{o}} & \boldsymbol{A}_{34} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\bar{c}o} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{C}_{co} & \boldsymbol{0} & \boldsymbol{C}_{\bar{c}o} & \boldsymbol{D} \end{array}\right] \text{ with}$$

$$\begin{aligned} &\text{controllable and observable } \left[\begin{array}{c|c} \boldsymbol{A}_{co} & \boldsymbol{B}_{co} \\ \hline \boldsymbol{C}_{co} & \boldsymbol{D} \end{array}\right] \\ &\text{controllable } \left[\begin{array}{cc|c} \boldsymbol{A}_{c\bar{o}} & \boldsymbol{A}_{12} & \boldsymbol{B}_{c\bar{o}} \\ \boldsymbol{0} & \boldsymbol{A}_{co} & \boldsymbol{B}_{co} \\ \hline \boldsymbol{0} & \boldsymbol{C}_{co} & \boldsymbol{D} \end{array}\right] \\ &\text{observable } \left[\begin{array}{cc|c} \boldsymbol{A}_{co} & \boldsymbol{A}_{24} & \boldsymbol{B}_{co} \\ \boldsymbol{0} & \boldsymbol{A}_{\bar{c}o} & \boldsymbol{0} \\ \hline \boldsymbol{C}_{co} & \boldsymbol{C}_{\bar{c}o} & \boldsymbol{D} \end{array}\right] \end{aligned} \tag{C.37}$$

This decomposes the system's total number of states p into

$$p = p_{c\bar{o}} + p_{co} + p_{\bar{c}\bar{o}} + p_{\bar{c}o} \tag{C.38}$$

where $p_{c\bar{o}}, p_{co}, p_{\bar{c}\bar{o}}, p_{\bar{c}o}$ are the respective dimensions of $\boldsymbol{A}_{c\bar{o}}, \boldsymbol{A}_{co}, \boldsymbol{A}_{\bar{c}\bar{o}}, \boldsymbol{A}_{\bar{c}o}$. Regarding input output behavior, the system $\boldsymbol{P}$ can be reduced in state dimension to its controllable and observable subsystem by cutting off the uncontrollable and unobservable parts so that

$$\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{A}_{co} & \boldsymbol{B}_{co} \\ \hline \boldsymbol{C}_{co} & \boldsymbol{D} \end{array}\right] \tag{C.39}$$

since the input controls only controllable parts and the output observes only observable parts. The resulting state space representation can not be further reduced without changing the system $\boldsymbol{P}$. Therefore a controllable and observable state space representation is called *minimal*.

C.1.12 McMillan Degree

The state dimension p_{co} of this minimal state space representation of the system $\boldsymbol{P}$ is the number of controllable and observable eigenmodes. Since controllability and observability is invariant under change of basis (C.36), this number is also a characteristic property of $\boldsymbol{P}$ regardless of the state space representation chosen and is called the McMillan degree δ of the system

$$\delta(\boldsymbol{P}) = p_{co} \leq p \tag{C.40}$$

where p is the state dimension of any state space representation of $\boldsymbol{P}$. This inequality reflects the fact that a controllable and observable state space representation has state dimension $p = p_{co}$ and is therefore minimal.

The combined McMillan degree $\delta(\boldsymbol{P}_1, \boldsymbol{P}_2)$ of two systems $\boldsymbol{P}_1, \boldsymbol{P}_2$ can not exceed

$$\delta(\boldsymbol{P}_1, \boldsymbol{P}_2) \leq \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2) \tag{C.41}$$

but may be lower, depending on the actual state space representations and the connection between them.

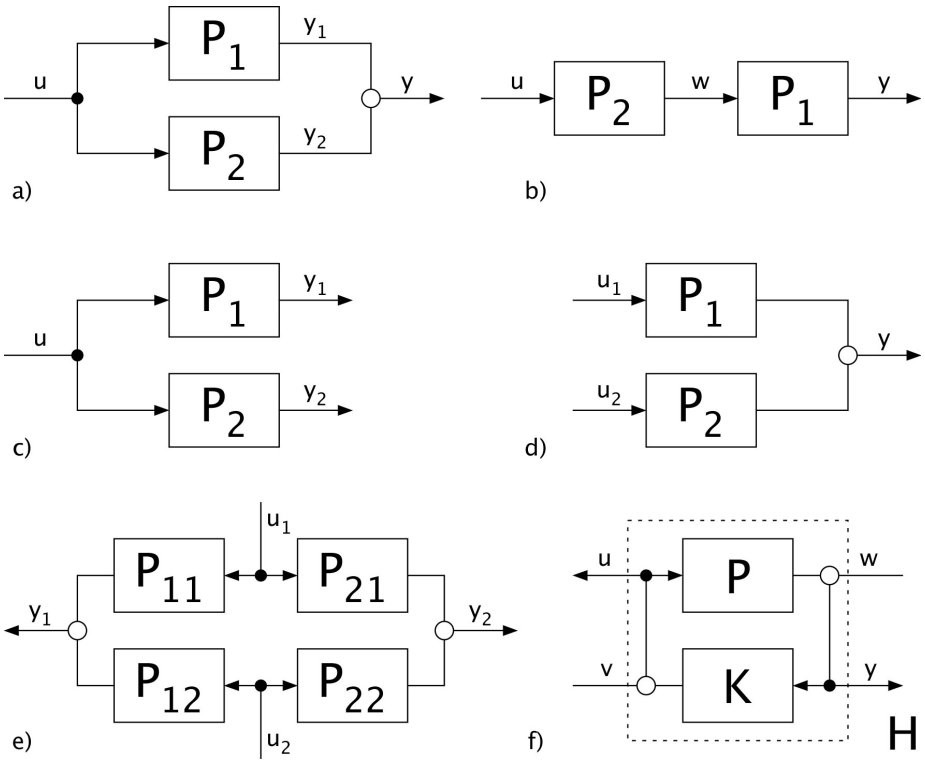


Figure C.1: The different system connections: a) parallel b) serial c) common input d) common output e) bridge f) feedback loop.

Proof This stems from the fact that using (C.37) we can reduce the state space representations of $\boldsymbol{P}_1, \boldsymbol{P}_2$ to its minimal form so that the combined state vector has the dimension $p = \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2)$. From (C.40) we have $\delta(\boldsymbol{P}_1, \boldsymbol{P}_2) \leq p$ so that (C.41) is obtained.

C.2 Connections

This section provides some formulae for operations on systems and the various possible connections between systems to form super systems as illustrated in figure C.1.

C.2.1 Flow Inversion

The signal flow through a system

$$\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right]$$

may be inverted if $\exists \boldsymbol{D}^{-1}$ and the inverted system is then given by

$$\boldsymbol{P}^{-1} = \left[\begin{array}{c|c} \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{B}\boldsymbol{D}^{-1} \\ \hline -\boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{D}^{-1} \end{array}\right] \tag{C.42}$$

with McMillan degree

$$\delta(\boldsymbol{P}) = \delta(\boldsymbol{P}^{-1}) \tag{C.43}$$

Proof This follows from the application of (A.55) on (C.7). To verify the preservation of the McMillan degree let $\boldsymbol{P}^{-1}$ be unobservable so that $\forall \boldsymbol{x} : (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})\boldsymbol{x} = \lambda\boldsymbol{x}$ we have $\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{x} = \boldsymbol{0}$ using (C.33). This leads to $(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{0}$ so that if $\boldsymbol{P}^{-1}$ is unobservable then $\boldsymbol{P}$ is unobservable or alternatively if $\boldsymbol{P}$ is observable then $\boldsymbol{P}^{-1}$ is observable. The same is true for controllability, so that if $\boldsymbol{P}$ is minimal then $\boldsymbol{P}^{-1}$ is minimal. This means there may be cases where $\boldsymbol{P}$ is not minimal with $\boldsymbol{P}^{-1}$ minimal. Since $\boldsymbol{P}^{-1}$ has the same state dimension than $\boldsymbol{P}$ following (C.42) this leads to $\delta(\boldsymbol{P}) \leq \delta(\boldsymbol{P}^{-1})$ and therefore $\delta(\boldsymbol{P}) \leq \delta(\boldsymbol{P}^{-1}) \leq \delta((\boldsymbol{P}^{-1})^{-1}) = \delta(\boldsymbol{P})$ so that (C.43) follows.

C.2.2 Parallel Connection

Two systems

$$\boldsymbol{P}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right];\ i \in \{1, 2\}$$

in parallel connection give

$$\boldsymbol{P}_1 + \boldsymbol{P}_2 = \left[\begin{array}{cc|c} \boldsymbol{A}_1 & \boldsymbol{0} & \boldsymbol{B}_1 \\ \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_1 & \boldsymbol{C}_2 & \boldsymbol{D}_1 + \boldsymbol{D}_2 \end{array}\right] \tag{C.44}$$

with McMillan degree

$$\delta(\boldsymbol{P}_1 + \boldsymbol{P}_2) \leq \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2) \tag{C.45}$$

Proof This is due to $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{C}_1\boldsymbol{x}_1 + \boldsymbol{C}_2\boldsymbol{x}_2 + (\boldsymbol{D}_1 + \boldsymbol{D}_2)\boldsymbol{u}$. The McMillan degree can have maximal range $\delta(\boldsymbol{P}_1 + \boldsymbol{P}_2) \in [0, \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2)]$. The lower bound is attained if for example $\boldsymbol{P}_2 = -\boldsymbol{P}_1$ so that $\delta(\boldsymbol{P}_1 + \boldsymbol{P}_2) = \delta(\boldsymbol{0}) = 0$.

C.2.3 Serial Connection

Two systems

$$\boldsymbol{P}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right];\ i \in \{1, 2\}$$

in serial connection give

$$\boldsymbol{P}_1\boldsymbol{P}_2 = \left[\begin{array}{cc|c} \boldsymbol{A}_1 & \boldsymbol{B}_1\boldsymbol{C}_2 & \boldsymbol{B}_1\boldsymbol{D}_2 \\ \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_1 & \boldsymbol{D}_1\boldsymbol{C}_2 & \boldsymbol{D}_1\boldsymbol{D}_2 \end{array}\right] \tag{C.46}$$

with McMillan degree

$$\delta(\boldsymbol{P}_1\boldsymbol{P}_2) \leq \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2) \tag{C.47}$$

Proof This is due to $\boldsymbol{w} = \boldsymbol{C}_2\boldsymbol{x}_2 + \boldsymbol{D}_2\boldsymbol{u}$ so that $\dot{\boldsymbol{x}}_1 = \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{B}_1(\boldsymbol{C}_2\boldsymbol{x}_2 + \boldsymbol{D}_2\boldsymbol{u})$ and $\boldsymbol{y} = \boldsymbol{C}_1\boldsymbol{x}_1 + \boldsymbol{D}_1(\boldsymbol{C}_2\boldsymbol{x}_2 + \boldsymbol{D}_2\boldsymbol{u})$. The McMillan degree can have maximal range $\delta(\boldsymbol{P}_1\boldsymbol{P}_2) \in [0, \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2)]$. The lower bound is attained if for example $\boldsymbol{P}_2 = \boldsymbol{P}_1^{-1}$ so that $\delta(\boldsymbol{P}_1\boldsymbol{P}_2) = \delta(\boldsymbol{1}) = 0$.

C.2.4 Feedback Loop

A plant $\boldsymbol{P}$ and controller $\boldsymbol{K}$

$$\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{A}_P & \boldsymbol{B}_P \\ \hline \boldsymbol{C}_P & \boldsymbol{D}_P \end{array}\right];\ \boldsymbol{K} = \left[\begin{array}{c|c} \boldsymbol{A}_K & \boldsymbol{B}_K \\ \hline \boldsymbol{C}_K & \boldsymbol{D}_K \end{array}\right]$$

may be put in feedback loop if $\exists \boldsymbol{E} = (\mathbf{1} - \boldsymbol{D}_P\boldsymbol{D}_K)^{-1}, \exists \boldsymbol{F} = (\mathbf{1} - \boldsymbol{D}_K\boldsymbol{D}_P)^{-1}$ then the closed loop system is given by

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \text{ with } \boldsymbol{H} = \begin{bmatrix} -\boldsymbol{K} & \mathbf{1} \\ \mathbf{1} & -\boldsymbol{P} \end{bmatrix}^{-1} =$$
$$= \left[\begin{array}{cc|cc} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}\boldsymbol{D}_K\boldsymbol{C}_P & \boldsymbol{B}_P\boldsymbol{F}\boldsymbol{C}_K & \boldsymbol{B}_P\boldsymbol{F} & \boldsymbol{B}_P\boldsymbol{F}\boldsymbol{D}_K \\ \boldsymbol{B}_K\boldsymbol{E}\boldsymbol{C}_P & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{E}\boldsymbol{D}_P\boldsymbol{C}_K & \boldsymbol{B}_K\boldsymbol{E}\boldsymbol{D}_P & \boldsymbol{B}_K\boldsymbol{E} \\ \hline \boldsymbol{E}\boldsymbol{C}_P & \boldsymbol{E}\boldsymbol{D}_P\boldsymbol{C}_K & \boldsymbol{E}\boldsymbol{D}_P & \boldsymbol{E} \\ \boldsymbol{F}\boldsymbol{D}_K\boldsymbol{C}_P & \boldsymbol{F}\boldsymbol{C}_K & \boldsymbol{F} & \boldsymbol{F}\boldsymbol{D}_K \end{array}\right] \tag{C.48}$$

with McMillan degree

$$\delta(\boldsymbol{H}) = \delta(\boldsymbol{P}) + \delta(\boldsymbol{K}) \tag{C.49}$$

Proof The result follows from $\boldsymbol{y} = \boldsymbol{C}_P\boldsymbol{x}_P + \boldsymbol{D}_P\boldsymbol{u} + \boldsymbol{w}$ and $\boldsymbol{u} = \boldsymbol{C}_K\boldsymbol{x}_K + \boldsymbol{D}_K\boldsymbol{y} + \boldsymbol{v}$ so that $\boldsymbol{y} = \boldsymbol{E}(\boldsymbol{C}_P\boldsymbol{x}_P + \boldsymbol{D}_P\boldsymbol{C}_K\boldsymbol{x}_K + \boldsymbol{w} + \boldsymbol{D}_P\boldsymbol{v})$ and $\boldsymbol{u} = \boldsymbol{F}(\boldsymbol{C}_K\boldsymbol{x}_K + \boldsymbol{D}_K\boldsymbol{C}_P\boldsymbol{x}_P + \boldsymbol{v} + \boldsymbol{D}_K\boldsymbol{w})$ leads to the above result by making use of (A.56). To verify the McMillan degree let the state space representation of $\boldsymbol{P}, \boldsymbol{K}$ be minimal and assume $\boldsymbol{H}$ unobservable. Using inobservability from (C.33) on (C.48) there exist $\boldsymbol{x}_P, \boldsymbol{x}_K$ so that

$$\begin{bmatrix} \boldsymbol{A}_P + \boldsymbol{B}_P\boldsymbol{F}\boldsymbol{D}_K\boldsymbol{C}_P & \boldsymbol{B}_P\boldsymbol{F}\boldsymbol{C}_K \\ \boldsymbol{B}_K\boldsymbol{E}\boldsymbol{C}_P & \boldsymbol{A}_K + \boldsymbol{B}_K\boldsymbol{E}\boldsymbol{D}_P\boldsymbol{C}_K \\ \boldsymbol{E}\boldsymbol{C}_P & \boldsymbol{E}\boldsymbol{D}_P\boldsymbol{C}_K \\ \boldsymbol{F}\boldsymbol{D}_K\boldsymbol{C}_P & \boldsymbol{F}\boldsymbol{C}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_P \\ \boldsymbol{x}_K \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{x}_P \\ \boldsymbol{x}_K \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

This simplifies to

$$\begin{bmatrix} \boldsymbol{A}_P & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_K \\ \boldsymbol{C}_P & \boldsymbol{D}_P\boldsymbol{C}_K \\ \boldsymbol{D}_K\boldsymbol{C}_P & \boldsymbol{C}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_P \\ \boldsymbol{x}_K \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{x}_P \\ \boldsymbol{x}_K \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

But $\boldsymbol{P}, \boldsymbol{K}$ are supposed to have minimal state space representations thus in addition to that we have $\boldsymbol{P}, \boldsymbol{K}$ observable

$$\boldsymbol{C}_P\boldsymbol{x}_P \neq \mathbf{0}\ \forall \boldsymbol{x}_P : \boldsymbol{A}_P\boldsymbol{x}_P = \lambda\boldsymbol{x}_P$$
$$\boldsymbol{C}_K\boldsymbol{x}_K \neq \mathbf{0}\ \forall \boldsymbol{x}_K : \boldsymbol{A}_K\boldsymbol{x}_K = \lambda\boldsymbol{x}_K$$

If λ is an eigenvalue of $\boldsymbol{A}_P$ but not $\boldsymbol{A}_K$ then $\boldsymbol{x}_K = \mathbf{0}$. This further simplifies the inobservability equation

$$\begin{bmatrix} \boldsymbol{A}_P \\ \boldsymbol{C}_P \\ \boldsymbol{D}_K\boldsymbol{C}_P \end{bmatrix} \boldsymbol{x}_P = \lambda \begin{bmatrix} \boldsymbol{x}_P \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

But $\boldsymbol{C}_P\boldsymbol{x}_P = \boldsymbol{0}$ contradicts the observability of $\boldsymbol{P}$ so $\boldsymbol{H}$ is observable. The same is true if λ is an eigenvalue of $\boldsymbol{A}_K$ but not $\boldsymbol{A}_P$. If λ is a common eigenvalue of $\boldsymbol{A}_P, \boldsymbol{A}_K$ then the inobservability equation gives

$$\begin{aligned}\boldsymbol{C}_P\boldsymbol{x}_P + \boldsymbol{D}_P\boldsymbol{C}_K\boldsymbol{x}_K &= \boldsymbol{0}\\ \boldsymbol{D}_K\boldsymbol{C}_P\boldsymbol{x}_P + \boldsymbol{C}_K\boldsymbol{x}_K &= \boldsymbol{0}\end{aligned}$$

This leads to

$$\begin{aligned}(\boldsymbol{1} - \boldsymbol{D}_P\boldsymbol{D}_K)\boldsymbol{C}_P\boldsymbol{x}_P = \boldsymbol{E}\boldsymbol{C}_P\boldsymbol{x}_P = \boldsymbol{0} \Leftrightarrow \boldsymbol{C}_P\boldsymbol{x}_P = \boldsymbol{0}\\ (\boldsymbol{1} - \boldsymbol{D}_K\boldsymbol{D}_P)\boldsymbol{C}_K\boldsymbol{x}_K = \boldsymbol{F}\boldsymbol{C}_K\boldsymbol{x}_K = \boldsymbol{0} \Leftrightarrow \boldsymbol{C}_K\boldsymbol{x}_K = \boldsymbol{0}\end{aligned}$$

since $\boldsymbol{E}, \boldsymbol{F}$ are required to be invertible to form the feedback loop. This contradicts the observability of $\boldsymbol{P}$ and $\boldsymbol{K}$ so $\boldsymbol{H}$ is observable. So $\boldsymbol{P}, \boldsymbol{K}$ minimal leads to $\boldsymbol{H}$ observable. The same is true for controllability so that if $\boldsymbol{P}, \boldsymbol{K}$ are minimal then $\boldsymbol{H}$ is minimal. This means there may be cases where one of $\boldsymbol{P}, \boldsymbol{K}$ is not minimal with $\boldsymbol{H}$ minimal. Since the state dimension of $\boldsymbol{H}$ is the sum of the state dimensions of $\boldsymbol{P}, \boldsymbol{K}$ following (C.48) this leads to $\delta(\boldsymbol{H}) \geq \delta(\boldsymbol{P}) + \delta(\boldsymbol{K})$. But from (C.41) we see that $\delta(\boldsymbol{H}) \leq \delta(\boldsymbol{P}) + \delta(\boldsymbol{K})$ so that (C.49) follows.

C.2.5 Common Input

Two systems

$$\boldsymbol{P}_i = \left[\begin{array}{c|c} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i \end{array}\right];\ i \in \{1, 2\}$$

with common input give

$$\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix} = \left[\begin{array}{cc|c} \boldsymbol{A}_1 & \boldsymbol{0} & \boldsymbol{B}_1 \\ \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_1 & \boldsymbol{0} & \boldsymbol{D}_1 \\ \boldsymbol{0} & \boldsymbol{C}_2 & \boldsymbol{D}_2 \end{array}\right] \tag{C.50}$$

with McMillan degree

$$\max(\delta(\boldsymbol{P}_1), \delta(\boldsymbol{P}_2)) \leq \delta\left(\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix}\right) \leq \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2) \tag{C.51}$$

Proof To verify the lower bound of the McMillan degree, let $\boldsymbol{P}_1, \boldsymbol{P}_2$ be minimal or reduce them to their minimal subsystems. Then the connected system is observable:

$$\begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}_1\boldsymbol{x}_1\\ \boldsymbol{C}_2\boldsymbol{x}_2\end{bmatrix} \neq \boldsymbol{0}$$

$$\forall \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} : \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_1\boldsymbol{x}_1\\ \boldsymbol{A}_2\boldsymbol{x}_2\end{bmatrix} = \lambda\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}$$

If λ is an eigenvalue of $\boldsymbol{A}_1$ but not $\boldsymbol{A}_2$ then $\boldsymbol{x}_2 = \boldsymbol{0}$ however $\boldsymbol{C}_1\boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being observable suffices to assure observability. The same is true if λ is an eigenvalue of $\boldsymbol{A}_2$ but not $\boldsymbol{A}_1$. If λ is a common eigenvalue of $\boldsymbol{A}_1, \boldsymbol{A}_2$ then

$\boldsymbol{C}_1\boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being observable and $\boldsymbol{C}_2\boldsymbol{x}_2 \neq \boldsymbol{0}$ from $\boldsymbol{P}_2$ being observable. But the connected system is not always controllable:

$$\begin{bmatrix}\boldsymbol{B}_1\\ \boldsymbol{B}_2\end{bmatrix}^T \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \boldsymbol{B}_1^T\boldsymbol{x}_1 + \boldsymbol{B}_2^T\boldsymbol{x}_2 \overset{?}{\neq} \boldsymbol{0}$$

$$\forall \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} : \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_2\end{bmatrix}^T \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_1^T\boldsymbol{x}_1\\ \boldsymbol{A}_2^T\boldsymbol{x}_2\end{bmatrix} = \lambda \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}$$

If λ is an eigenvalue of $\boldsymbol{A}_1$ but not $\boldsymbol{A}_2$ then $\boldsymbol{x}_2 = \boldsymbol{0}$ and $\boldsymbol{B}_1^T\boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being controllable suffices to assure controllability. The same is true if λ is an eigenvalue of $\boldsymbol{A}_2$ but not $\boldsymbol{A}_1$. If λ is a common eigenvalue of $\boldsymbol{A}_1, \boldsymbol{A}_2$ then $\boldsymbol{B}_1^T\boldsymbol{x}_1 + \boldsymbol{B}_2^T\boldsymbol{x}_2 = \boldsymbol{0}$ is possible. In this case, the common λ is an uncontrollable eigenmode of the connected system and reduces the McMillan degree by one. If it is a multiple eigenvalue of both matrices then it reduces the McMillan degree by its multiplicity. Now let $\delta(\boldsymbol{P}_1) \leq \delta(\boldsymbol{P}_2)$ then the lower bound of the McMillan degree of the connected system can be attained if all eigenvalues of $\boldsymbol{A}_1$ are also eigenvalues of $\boldsymbol{A}_2$. Then the McMillan degree of the unconnected system $\delta(\boldsymbol{P}_1)+\delta(\boldsymbol{P}_2)$ can be reduced by the degree of the smaller system $\delta(\boldsymbol{P}_1)$ through connection. The result is the McMillan degree of the larger system $\delta(\boldsymbol{P}_2)$ of the two or in general $\max(\delta(\boldsymbol{P}_1), \delta(\boldsymbol{P}_2))$.

Remark An important special case is $\boldsymbol{A}_i = \boldsymbol{A}, \boldsymbol{B}_i = \boldsymbol{B};\ i \in \{1, 2\}$ then

$$\begin{bmatrix}\boldsymbol{P}_1\\ \boldsymbol{P}_2\end{bmatrix} = \left[\begin{array}{c|c}\boldsymbol{A} & \boldsymbol{B}\\ \hline \boldsymbol{C}_1 & \boldsymbol{D}_1\\ \boldsymbol{C}_2 & \boldsymbol{D}_2\end{array}\right] \tag{C.52}$$

C.2.6 Common Output

Two systems

$$\boldsymbol{P}_i = \left[\begin{array}{c|c}\boldsymbol{A}_i & \boldsymbol{B}_i\\ \hline \boldsymbol{C}_i & \boldsymbol{D}_i\end{array}\right];\ i \in \{1, 2\}$$

with common output give

$$\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2\end{bmatrix} = \left[\begin{array}{cc|cc}\boldsymbol{A}_1 & \boldsymbol{0} & \boldsymbol{B}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{0} & \boldsymbol{B}_2\\ \hline \boldsymbol{C}_1 & \boldsymbol{C}_2 & \boldsymbol{D}_1 & \boldsymbol{D}_2\end{array}\right] \tag{C.53}$$

with McMillan degree

$$\max(\delta(\boldsymbol{P}_1), \delta(\boldsymbol{P}_2)) \leq \delta\left(\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2\end{bmatrix}\right) \leq \delta(\boldsymbol{P}_1) + \delta(\boldsymbol{P}_2) \tag{C.54}$$

Proof To verify the lower bound of the McMillan degree, let $\boldsymbol{P}_1, \boldsymbol{P}_2$ be minimal or reduce them to their minimal subsystems. Then the connected system is controllable:

$$\begin{bmatrix}\boldsymbol{B}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{B}_2\end{bmatrix}^T \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{B}_1^T\boldsymbol{x}_1\\ \boldsymbol{B}_2^T\boldsymbol{x}_2\end{bmatrix} \neq \boldsymbol{0}$$

$$\forall \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} : \begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_2\end{bmatrix}^T \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_1^T\boldsymbol{x}_1\\ \boldsymbol{A}_2^T\boldsymbol{x}_2\end{bmatrix} = \lambda \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}$$

If λ is an eigenvalue of $\boldsymbol{A}_1$ but not $\boldsymbol{A}_2$ then $\boldsymbol{x}_2 = \boldsymbol{0}$ however $\boldsymbol{B}_1^T \boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being controllable suffices to assure controllability. The same is true if λ is an eigenvalue of $\boldsymbol{A}_2$ but not $\boldsymbol{A}_1$. If λ is a common eigenvalue of $\boldsymbol{A}_1, \boldsymbol{A}_2$ then $\boldsymbol{B}_1^T \boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being controllable and $\boldsymbol{B}_2^T \boldsymbol{x}_2 \neq \boldsymbol{0}$ from $\boldsymbol{P}_2$ being controllable. But the connected system is not always observable:

$$\begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \boldsymbol{C}_1 \boldsymbol{x}_1 + \boldsymbol{C}_2 \boldsymbol{x}_2 \stackrel{?}{\neq} \boldsymbol{0}$$

$$\forall \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} : \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \boldsymbol{x}_1 \\ \boldsymbol{A}_2 \boldsymbol{x}_2 \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}$$

If λ is an eigenvalue of $\boldsymbol{A}_1$ but not $\boldsymbol{A}_2$ then $\boldsymbol{x}_2 = \boldsymbol{0}$ and $\boldsymbol{C}_1 \boldsymbol{x}_1 \neq \boldsymbol{0}$ from $\boldsymbol{P}_1$ being observable suffices to assure observability. The same is true if λ is an eigenvalue of $\boldsymbol{A}_2$ but not $\boldsymbol{A}_1$. If λ is a common eigenvalue of $\boldsymbol{A}_1, \boldsymbol{A}_2$ then $\boldsymbol{C}_1 \boldsymbol{x}_1 + \boldsymbol{C}_2 \boldsymbol{x}_2 = \boldsymbol{0}$ is possible. In this case, the common λ is an unobservable eigenmode of the connected system and reduces the McMillan degree by one. If it is a multiple eigenvalue of both matrices then it reduces the McMillan degree by its multiplicity. Now let $\delta(\boldsymbol{P}_1) \leq \delta(\boldsymbol{P}_2)$ then the lower bound of the McMillan degree of the connected system can be attained if all eigenvalues of $\boldsymbol{A}_1$ are also eigenvalues of $\boldsymbol{A}_2$. Then the McMillan degree of the unconnected system $\delta(\boldsymbol{P}_1)+\delta(\boldsymbol{P}_2)$ can be reduced by the degree of the smaller system $\delta(\boldsymbol{P}_1)$ through connection. The result is the McMillan degree of the larger system $\delta(\boldsymbol{P}_2)$ of the two or in general $\max(\delta(\boldsymbol{P}_1), \delta(\boldsymbol{P}_2))$.

Remark An important special case is $\boldsymbol{A}_i = \boldsymbol{A}, \boldsymbol{C}_i = \boldsymbol{C};\ i \in \{1, 2\}$ then

$$\begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 \end{bmatrix} = \left[\begin{array}{c|cc} \boldsymbol{A} & \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C} & \boldsymbol{D}_1 & \boldsymbol{D}_2 \end{array}\right] \tag{C.55}$$

C.2.7 Bridge

Four systems

$$\boldsymbol{P}_{ij} = \left[\begin{array}{c|c} \boldsymbol{A}_{ij} & \boldsymbol{B}_{ij} \\ \hline \boldsymbol{C}_{ij} & \boldsymbol{D}_{ij} \end{array}\right];\ i, j \in \{1, 2\}$$

in bridge connection give

$$\begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{bmatrix} = \left[\begin{array}{cccc|cc} \boldsymbol{A}_{11} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{12} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_{12} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{21} & \boldsymbol{0} & \boldsymbol{B}_{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{22} & \boldsymbol{0} & \boldsymbol{B}_{22} \\ \hline \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{array}\right] \tag{C.56}$$

with McMillan degree

$$\begin{aligned} &\max(\delta(\boldsymbol{P}_{11}), \delta(\boldsymbol{P}_{12}), \delta(\boldsymbol{P}_{21}), \delta(\boldsymbol{P}_{22})) \leq \\ &\qquad \leq \delta\left(\begin{bmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{bmatrix}\right) \leq \\ &\qquad \leq \delta(\boldsymbol{P}_{11}) + \delta(\boldsymbol{P}_{12}) + \delta(\boldsymbol{P}_{21}) + \delta(\boldsymbol{P}_{22}) \end{aligned} \tag{C.57}$$

Proof The bridge can obtained by first connecting $\boldsymbol{P}_{11}, \boldsymbol{P}_{12}$ and $\boldsymbol{P}_{21}, \boldsymbol{P}_{22}$ each to common outputs using (C.53) and (C.54). And then connecting the two resulting $\begin{bmatrix}\boldsymbol{P}_{11} & \boldsymbol{P}_{12}\end{bmatrix}$ and $\begin{bmatrix}\boldsymbol{P}_{21} & \boldsymbol{P}_{22}\end{bmatrix}$ to a common input using (C.50) and (C.51).

Remark An important special case is $\boldsymbol{A}_{ij} = \boldsymbol{A}, \boldsymbol{B}_{ij} = \boldsymbol{B}_j, \boldsymbol{C}_{ij} = \boldsymbol{C}_i;\ i, j \in \{1, 2\}$ then

$$\begin{bmatrix}\boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22}\end{bmatrix} = \left[\begin{array}{c|cc} \boldsymbol{A} & \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_1 & \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{C}_2 & \boldsymbol{D}_{21} & \boldsymbol{D}_{22}\end{array}\right] \tag{C.58}$$

C.3 Factorizations

This section provides formulae for the important right and left factorization of systems.

C.3.1 Right Factorization

A system

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{Y}^{-1} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D}\end{array}\right]$$

has a right factorization in the form of

$$\begin{bmatrix}\boldsymbol{X} \\ \boldsymbol{Y}\end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{A}+\boldsymbol{B}\boldsymbol{F} & \boldsymbol{B}\boldsymbol{U} \\ \hline \boldsymbol{C}+\boldsymbol{D}\boldsymbol{F} & \boldsymbol{D}\boldsymbol{U} \\ \boldsymbol{F} & \boldsymbol{U}\end{array}\right] \tag{C.59}$$

with free parameters $\boldsymbol{F}, \boldsymbol{U}$ and $\boldsymbol{U}$ invertible. Its McMillan degree is bounded by

$$\delta\left(\begin{bmatrix}\boldsymbol{X} \\ \boldsymbol{Y}\end{bmatrix}\right) \geq \delta(\boldsymbol{P}) \tag{C.60}$$

see [7] for details.

Proof Start with two systems having common input

$$\begin{bmatrix}\boldsymbol{X} \\ \boldsymbol{Y}\end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{A}' & \boldsymbol{B}' \\ \hline \boldsymbol{C}_X & \boldsymbol{D}_X \\ \boldsymbol{C}_Y & \boldsymbol{D}_Y\end{array}\right]$$

with $\boldsymbol{D}_Y$ invertible in order to allow the inversion of $\boldsymbol{Y}$. Using flow inversion (C.42) and serial connection (C.46) this leads to

$$\boldsymbol{X}\boldsymbol{Y}^{-1} = \left[\begin{array}{cc|c} \boldsymbol{A}' & -\boldsymbol{B}'\boldsymbol{D}_Y^{-1}\boldsymbol{C}_Y & \boldsymbol{B}'\boldsymbol{D}_Y^{-1} \\ \boldsymbol{0} & \boldsymbol{A}'-\boldsymbol{B}'\boldsymbol{D}_Y^{-1}\boldsymbol{C}_Y & \boldsymbol{B}'\boldsymbol{D}_Y^{-1} \\ \hline \boldsymbol{C}_X & -\boldsymbol{D}_X\boldsymbol{D}_Y^{-1}\boldsymbol{C}_Y & \boldsymbol{D}_X\boldsymbol{D}_Y^{-1}\end{array}\right]$$

Applying the transformation $\boldsymbol{T}$ as in (C.36) and cutting off the uncontrollable part leads to

$$\boldsymbol{X}\boldsymbol{Y}^{-1} = \left[\begin{array}{c|c} \boldsymbol{A}'-\boldsymbol{B}'\boldsymbol{D}_Y^{-1}\boldsymbol{C}_Y & \boldsymbol{B}'\boldsymbol{D}_Y^{-1} \\ \hline \boldsymbol{C}_X-\boldsymbol{D}_X\boldsymbol{D}_Y^{-1}\boldsymbol{C}_Y & \boldsymbol{D}_X\boldsymbol{D}_Y^{-1}\end{array}\right] \text{ with } \boldsymbol{T} = \begin{bmatrix}\boldsymbol{1} & -\boldsymbol{1} \\ \boldsymbol{0} & \boldsymbol{1}\end{bmatrix}$$

Comparing this to

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{Y}^{-1} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right]$$

provides a solution:

$$\boldsymbol{B} = \boldsymbol{B}'\boldsymbol{D}_Y^{-1} \Rightarrow \boldsymbol{A} = \boldsymbol{A}' - \boldsymbol{B}\boldsymbol{C}_Y,\ \boldsymbol{D} = \boldsymbol{D}_X\boldsymbol{D}_Y^{-1} \Rightarrow \boldsymbol{C} = \boldsymbol{C}_X - \boldsymbol{D}\boldsymbol{C}_Y$$

with free parameters $\boldsymbol{F} = \boldsymbol{C}_Y, \boldsymbol{U} = \boldsymbol{D}_Y^{-1}$ and $\boldsymbol{U}$ invertible since $\boldsymbol{D}_Y$ is supposed to be invertible, this immediately gives (C.59). To verify the McMillan degree let (C.59) be uncontrollable so that by using incontrollability from (C.31) on (C.59) there exists $\boldsymbol{x}$ so that

$$\begin{bmatrix}\boldsymbol{A}+\boldsymbol{B}\boldsymbol{F} & \boldsymbol{B}\boldsymbol{U}\end{bmatrix}^T \boldsymbol{x} = \lambda \begin{bmatrix}\boldsymbol{x} & \boldsymbol{0}\end{bmatrix}$$

This simplifies to

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\end{bmatrix}^T \boldsymbol{x} = \lambda \begin{bmatrix}\boldsymbol{x} & \boldsymbol{0}\end{bmatrix}$$

since $\boldsymbol{U}$ is invertible. This means that if (C.59) is uncontrollable then $\boldsymbol{P}$ is uncontrollable or alternatively if $\boldsymbol{P}$ is controllable then (C.59) is controllable. Now let (C.59) be unobservable so that by using inobservability from (C.33) on (C.59) there exists $\boldsymbol{x}$ so that

$$\begin{bmatrix}\boldsymbol{A}+\boldsymbol{B}\boldsymbol{F} \\ \boldsymbol{C}+\boldsymbol{D}\boldsymbol{F} \\ \boldsymbol{F}\end{bmatrix} \boldsymbol{x} = \lambda \begin{bmatrix}\boldsymbol{x} \\ \boldsymbol{0} \\ \boldsymbol{0}\end{bmatrix}$$

This simplifies to

$$\begin{bmatrix}\boldsymbol{A} \\ \boldsymbol{C}\end{bmatrix} \boldsymbol{x} = \lambda \begin{bmatrix}\boldsymbol{x} \\ \boldsymbol{0}\end{bmatrix}$$

This means that if (C.59) is unobservable then $\boldsymbol{P}$ is unobservable or alternatively if $\boldsymbol{P}$ is observable then (C.59) is observable. Therefore if $\boldsymbol{P}$ is minimal then (C.59) is minimal. This means there may be cases where $\boldsymbol{P}$ is not minimal with (C.59) minimal. Since (C.59) has the same state dimensions than $\boldsymbol{P}$, (C.60) follows.

C.3.2 Left Factorization

A system

$$\boldsymbol{P} = \boldsymbol{Y}^{-1}\boldsymbol{X} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right]$$

has a left factorization in the form of

$$\begin{bmatrix}\boldsymbol{X} & \boldsymbol{Y}\end{bmatrix} = \left[\begin{array}{c|cc} \boldsymbol{A}+\boldsymbol{L}\boldsymbol{C} & \boldsymbol{B}+\boldsymbol{L}\boldsymbol{D} & \boldsymbol{L} \\ \hline \boldsymbol{V}\boldsymbol{C} & \boldsymbol{V}\boldsymbol{D} & \boldsymbol{V} \end{array}\right] \tag{C.61}$$

with free parameters $\boldsymbol{L}, \boldsymbol{V}$ and $\boldsymbol{V}$ invertible. Its McMillan degree is bounded by

$$\delta\left(\begin{bmatrix}\boldsymbol{X} & \boldsymbol{Y}\end{bmatrix}\right) \geq \delta(\boldsymbol{P}) \tag{C.62}$$

see [7] for details.

Proof Start with two systems having common output

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} = \left[\begin{array}{c|cc} \boldsymbol{A}' & \boldsymbol{B}_X' & \boldsymbol{B}_Y \\ \hline \boldsymbol{C}' & \boldsymbol{D}_X & \boldsymbol{D}_Y \end{array}\right]$$

with $\boldsymbol{D}_Y$ invertible in order to allow the inversion of $\boldsymbol{Y}$. Using flow inversion (C.42) and serial connection (C.46) this leads to

$$\boldsymbol{Y}^{-1}\boldsymbol{X} = \left[\begin{array}{cc|c} \boldsymbol{A}' - \boldsymbol{B}_Y\boldsymbol{D}_Y^{-1}\boldsymbol{C}' & -\boldsymbol{B}_Y\boldsymbol{D}_Y^{-1}\boldsymbol{C}' & -\boldsymbol{B}_Y\boldsymbol{D}_Y^{-1}\boldsymbol{D}_X \\ \boldsymbol{0} & \boldsymbol{A}' & \boldsymbol{B}_X \\ \hline \boldsymbol{D}_Y^{-1}\boldsymbol{C}' & \boldsymbol{D}_Y^{-1}\boldsymbol{C}' & \boldsymbol{D}_Y^{-1}\boldsymbol{D}_X \end{array}\right]$$

Applying the transformation $\boldsymbol{T}$ as in (C.36) and cutting off the unobservable part leads to

$$\boldsymbol{Y}^{-1}\boldsymbol{X} = \left[\begin{array}{c|c} \boldsymbol{A}' - \boldsymbol{B}_Y\boldsymbol{D}_Y^{-1}\boldsymbol{C}' & \boldsymbol{B}_X - \boldsymbol{B}_Y\boldsymbol{D}_Y^{-1}\boldsymbol{D}_X \\ \hline \boldsymbol{D}_Y^{-1}\boldsymbol{C}' & \boldsymbol{D}_Y^{-1}\boldsymbol{D}_X \end{array}\right] \text{ with } \boldsymbol{T} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{1} \end{bmatrix}$$

Comparing this to

$$\boldsymbol{P} = \boldsymbol{Y}^{-1}\boldsymbol{X} = \left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right]$$

provides a solution:

$$\boldsymbol{C} = \boldsymbol{D}_Y^{-1}\boldsymbol{C}' \Rightarrow \boldsymbol{A} = \boldsymbol{A}' - \boldsymbol{B}_Y\boldsymbol{C},\ \boldsymbol{D} = \boldsymbol{D}_Y^{-1}\boldsymbol{D}_X \Rightarrow \boldsymbol{B} = \boldsymbol{B}_X - \boldsymbol{B}_Y\boldsymbol{D}$$

with free parameters $\boldsymbol{L} = \boldsymbol{B}_Y, \boldsymbol{V} = \boldsymbol{D}_Y^{-1}$ and $\boldsymbol{V}$ invertible since $\boldsymbol{D}_Y$ is supposed to be invertible, this immediately gives (C.61). To verify the McMillan degree let (C.61) be unobservable so that by using inobservability from (C.33) on (C.61) there exists $\boldsymbol{x}$ so that

$$\begin{bmatrix} \boldsymbol{A} + \boldsymbol{L}\boldsymbol{C} \\ \boldsymbol{V}\boldsymbol{C} \end{bmatrix} \boldsymbol{x} = \lambda \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix}$$

This simplifies to

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \boldsymbol{x} = \lambda \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix}$$

since $\boldsymbol{V}$ is invertible. This means that if (C.61) is unobservable then $\boldsymbol{P}$ is unobservable or alternatively if $\boldsymbol{P}$ is observable then (C.61) is observable. Now let (C.61) be uncontrollable so that by using incontrollability from (C.31) on (C.61) there exists $\boldsymbol{x}$ so that

$$\begin{bmatrix} \boldsymbol{A} + \boldsymbol{L}\boldsymbol{C} & \boldsymbol{B} + \boldsymbol{L}\boldsymbol{D} & \boldsymbol{L} \end{bmatrix}^T \boldsymbol{x} = \lambda \begin{bmatrix} \boldsymbol{x} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

This simplifies to

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}^T \boldsymbol{x} = \lambda \begin{bmatrix} \boldsymbol{x} & \boldsymbol{0} \end{bmatrix}$$

This means that if (C.61) is uncontrollable then $\boldsymbol{P}$ is uncontrollable or alternatively if $\boldsymbol{P}$ is controllable then (C.61) is controllable. Therefore if $\boldsymbol{P}$ is minimal then (C.61) is minimal. This means there may be cases where $\boldsymbol{P}$ is not minimal with (C.61) minimal. Since (C.61) has the same state dimensions than $\boldsymbol{P}$, (C.62) follows.

List of Figures

Bibliography

[1] J. Abedor, K. Nagpal, K. Poolla: "A Linear Matrix Inequality Approach to Peak to Peak Gain Minimization", International Journal of Robust and Nonlinear Control, 1996

[2] D. Alpay, L. Baratchart, A. Gombani: "On the Differential Structure of Matrix-valued Rational Inner Functions", Operator Theory: Advances and Applications, Vol. 73, Pp. 30-66, 1994

[3] P. Apkarian, D. Noll: "Nonsmooth H∞ Synthesis", IEEE Transactions on Automatic Control, Vol. 51, No. 1, 2006

[4] J. F. Bonnans, A. Shapiro: "Perturbation Analysis of Optimization Problems", Springer Series in Operations Research, 2000

[5] S. Boyd, L. Vandenberghe: "Convex Optimization", Cambridge University Press, 2004

[6] P. A. Fuhrmann: "A Polynomial Approach to Linear Algebra", Springer, 1996

[7] P. A. Fuhrmann, R. J. Ober: "State Space Formulas for coprime Factorizations", Operator theory: Advances and Applications: Contributions to Operator Theory and its Applications, The T. A. Ando Birthday Volume, pages 39–75. Birkh?auser Verlag, 1993

[8] S. Gumussoy, D. Henrion, M. Millstone, M. L. Overton: "Multiobjective Robust Control with HIFOO 2.0", Proceedings of the 6th IFAC Symposium on Robust Control Design, 2009

[9] X. F. Ji, H. Y. Su, J. Chu: "Peak to Peak Gain Minimization for Uncertain Linear Discrete Systems: A Matrix Inequality Approach", Acta Automatica Sinica, Vol. 33, No. 7, 2007

[10] K. Nagpal, J. Abedor, K. Poolla: "An LMI Approach to Peak-to-Peak Gain Minimization: Filtering and Control", Proceedings of the American Control Conference, 1994

[11] M. Olivi: "Parametrization of Rational Lossless Matrices with Applications to Linear Systems Theory", Habilitation at Université de Nice Sophia Antipolis, 2010

[12] B. T. Polyak, A. V. Nazin, M. V. Topunov, S. A. Nazin: "Rejection of Bounded Disturbances via Invariant Ellipsoids Technique", Proceedings of the 45th IEEE Conference on Decision and Control, 2006

[13] M. Vidyasagar: "Control System Synthesis: A Factorization Approach", MIT Press, 1985

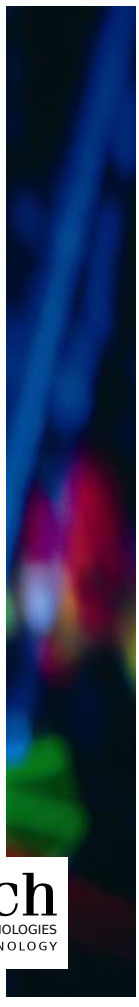
[14] K. Zhou, J. C. Doyle, K. Glover: "Robust and Optimal Control", Prentice Hall, 1996

Contrôle Multiobjectif d'ordre réduit

Résumé : Cette thèse présente une méthode permettant d'améliorer un régulateur existant relativement à des spécifications multiples formulées en termes de normes de système L_*, L_2/H_2 et H_∞. Trois éléments clés distinguent l'approche proposée. Premièrement, un paramétrage des régulateurs à ordre fixé au moyen d'un facteur stable et observable de la boucle fermée. Deuxièmement, un paramétrage minimal de toutes les paires stables observables fondé sur la théorie des systèmes conservatifs et permettant un ajustement infinitésimal du régulateur à améliorer. Et troisièmement, l'utilisation d'une information de sensibilité pour calculer le sous gradient local du critère d'optimisation et construire une suite minimisante.
Mots clés : Contrôle multi-objectifs, Systèmes conservatifs, Paramétrage du régulateur, Factorisation

Reduced Order Multiobjective Control

Abstract: A way to improve an existing controller with respect to a multiobjective specification,written in terms of the L_*, L_2/H_2 and H_∞ system norms, is presented. Three key elements distinguish the proposed approach. First, a parametrization of fixed order controllers by a stable and observable controller parameter. Second, a parametrization of all stable observable pairs based on the theory of lossless systems that allows an infinitesimal adjustment of the controller parameter. And third, the use of a sensitivity information to calculate the local subgradient of the optimization problem so that the controller parameter can be adjusted the right way.
Keywords: Multiobjective Control, Lossless Systems, Controller Parametrization, Factorization



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY