



**HAL**  
open science

# Stochastic optimization problems : decomposition and coordination under risk

Henri Gérard

► **To cite this version:**

Henri Gérard. Stochastic optimization problems : decomposition and coordination under risk. Optimization and Control [math.OC]. Université Paris-Est, 2018. English. NNT : 2018PESC1111 . tel-02067326

**HAL Id: tel-02067326**

**<https://pastel.hal.science/tel-02067326>**

Submitted on 14 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE MSTIC  
MATHÉMATIQUES, SCIENCES ET TECHNOLOGIES  
DE L'INFORMATION ET DE LA COMMUNICATION

## THÈSE

pour obtenir le titre de

Docteur de l'Université Paris-Est  
Spécialité : Mathématiques Appliquées

Présentée et soutenue par

Henri GÉRARD

## Stochastic optimization problems: Decomposition and coordination under risk

Thèse co-dirigée par Michel De Lara et Jean-Christophe Pesquet

*Préparée au sein de l'équipe OPTIMISATION ET SYSTÈMES  
du laboratoire CERMICS de l'ÉCOLE DES PONTS PARISTECH*

soutenue le 26 Octobre 2018

### JURY

---

---

<i>Rapporteurs :</i>	Andrzej Ruszczyński Marianne Akian	Rutgers University INRIA
<i>Président :</i>	Ralph Tyrrell Rockafellar	University of Washington
<i>Directeurs :</i>	Michel De Lara Jean-Christophe Pesquet	École des Ponts ParisTech École Centrale Paris
<i>Examineurs :</i>	Jean-Philippe Chancelier Andy Philpott Radu Ioan Boț	CERMICS University of Auckland University of Vienna



## Résumé

Nous considérons des problèmes d'optimisation stochastique, de théorie des jeux et d'apprentissage automatique avec des mesures de risque.

Dans une première partie, nous mettons l'accent sur la cohérence temporelle en optimisation stochastique. Nous commençons par prouver une équivalence entre cohérence temporelle et l'existence d'une formule imbriquée pour des fonctions. Motivés par des exemples bien connus dans les mesures de risque, nous étudions trois classes de fonctions: les fonctions invariantes par translation, les transformées de Fenchel-Moreau et les fonction supremum. Ensuite, nous étendons le concept de cohérence temporelle à la cohérence entre joueurs, en remplaçant le temps séquentiel par un ensemble non ordonné et les fonctions par des relations binaires. Enfin, nous montrons comment la cohérence entre joueurs est liée à des formes de décomposition séquentielles et parallèles en optimisation.

Dans une seconde partie, nous étudions l'impact des mesures de risque sur la multiplicité des équilibres dans les problèmes de jeux dynamiques dans les marchés complets et incomplets. Nous concevons un exemple où l'introduction de mesures de risque conduit à l'existence de trois équilibres au lieu d'un dans le cas risque neutre. Nous analysons la capacité de deux algorithmes différents à trouver les différents équilibres. Nous discutons des liens entre la cohérence des joueurs et les problèmes d'équilibre dans les jeux.

Dans une troisième partie, nous étudions l'optimisation robuste pour l'apprentissage automatique. En utilisant des mesures de risque convexes, nous fournissons un cadre unifié et proposons un algorithme adapté couvrant trois ensembles d'ensembles d'ambiguïté étudiés dans la littérature.



## Abstract

We consider stochastic optimization, game theory and machine learning problems with risk measures.

In a first part, we focus on time consistency in stochastic optimization. We begin by proving an equivalence between time consistent mappings and the existence of a nested formula. Motivated by well known examples in risk measures, we investigate three classes of mappings: translation invariant, Fenchel-Moreau transform and supremum mappings. Then, we extend the concept of time consistency to player consistency, by replacing the sequential time by any unordered set and mappings by any relations. Finally, we show how player consistency relates to sequential and parallel forms of decomposition in optimization.

In a second part, we study how risk measures impact the multiplicity of equilibria in dynamic game problems in complete and incomplete markets. We design an example where the introduction of risk measures leads to the existence of three equilibria instead of one in the risk neutral case. We analyze the ability of two different algorithms to recover the different equilibria. We discuss links between player consistency and equilibrium problems in games.

In a third part, we study distributionally robust optimization in machine learning. Using convex risk measures, we provide a unified framework and propose an adapted algorithm covering three ambiguity sets discussed in the literature.



## Notation

Here are the main notations used in the manuscript:

$\llbracket a, b \rrbracket$	set of integers between $a$ and $b$
w.r.t	with respect to
$\bigwedge_{x \in \mathbb{X}}$	infimum if it exists over the lattice $\mathbb{X}$
$\bigvee_{x \in \mathbb{X}}$	supremum if it exists over the lattice $\mathbb{X}$
$\Omega$	set of scenarios
$\mathbf{X}$	random variables over the set of scenarios $\Omega$
$\sigma(\mathbf{X})$	$\sigma$ -field generated by the random variable $\mathbf{X}$
$(\Omega, \mathcal{T}, \mathbb{P})$	probability space endowed with a $\sigma$ -field $\mathcal{T}$ and a probability $\mathbb{P}$
$\Delta(\Omega, \mathcal{T})$	set of all probability distributions over $(\Omega, \mathcal{T})$
$\mathcal{Q}$	subset of $\Delta(\Omega)$
$\mathbb{Q}$	element of $\mathcal{Q}$
$\overline{\mathbb{R}}$	set of extended reals $\mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$
$\mathbb{E}_{\mathbb{P}}[\cdot]$	mathematical expectation w.r.t. probability $\mathbb{P}$
$\rho(\cdot)$	mathematical risk mapping
$A$	aggregator mapping
$F$	factor mapping
$S^{A,F}$	subaggregator mapping
$\mathbb{H}$	set of head elements
$\mathbb{T}$	set of tail elements
$\mathcal{R}$	binary relation
$\leq$	order





## Remerciements

Au travers des quelques lignes qui suivent, je vais tenter de remercier toutes les personnes qui m'ont aidé et soutenu pendant ces trois dernières années, me permettant de délivrer ce travail aujourd'hui. Je ne pourrais malheureusement remercier tout le monde et je m'excuse par avance auprès de ceux qui se sentiraient offensés.

Je remercie tout d'abord mes deux rapporteurs Andrzej Ruszczyński et Marianne Akian qui ont accepté de consacrer leur temps à la relecture de ma thèse et ont ainsi fourni de précieux commentaires.

Ayant travaillé dans trois domaines distincts des mathématiques, il me semble important de signifier également ma gratitude à chacun de mes encadrants.

Je tiens particulièrement à remercier Michel De Lara, directeur de ma thèse, pour m'avoir guidé durant ces trois années et pour m'avoir transmis l'importance de la notion de clarté. Je pense également à Jean-Philippe Chancelier et Pierre Carpentier qui en plus d'être de grands scientifiques, sont des personnes très chaleureuses. J'en viens à présent à remercier Andy Philpott pour son accueil et son encadrement en Nouvelle-Zélande. J'ai beaucoup apprécié sa pédagogie, ses conseils linguistiques et son accueil amical sur place. J'ai également collaboré avec le Centre de Vision Numérique de l'école Centrale-Supélec et je remercie du fond du coeur Jean-Christophe Pesquet, mon second directeur de thèse, ainsi que Émilie Chouzenoux, pour leurs conseils avisés et leurs écoutes.

Je remercie également le jury pour leur présence lors de ma soutenance et leurs remarques fructueuses.

J'aimerais à présent me concentrer sur mes amis et collègues du CERMICS. Un grand merci à Isabelle et Fatna pour m'avoir permis de m'affranchir des contraintes administratives pour me concentrer sur ma recherche. Je pense également à Frédéric et Bernard qui m'ont permis d'enseigner à l'École des Ponts ce qui s'est révélé être une formidable expérience. Je n'oublie pas mes deux camarades de thèse François et Tristan qui étaient à mes côtés lors des réunions d'équipe et des diverses conférences. Je ne pourrais mentionner toute l'équipe des doctorants et stagiaires ici bien que chacun ait contribué à me sentir heureux dans ce laboratoire. Néanmoins, mes camarades ayant commencé leurs thèses en même temps que moi au 2ème étage méritent une mention spéciale: Marion, Alexandre, Étienne et Gustave.

Mais bien sûr, au delà du travail, il y a tous les amis qui m'ont accompagné voire ont supporté mes élucubrations mathématiques par moment. Merci à mes deux colocataires mais avant tout amis, Vincent et Guillaume : merci à Guillaume pour nos débats politiques et instructifs et à Vincent pour le goût de l'entrepreneuriat qu'il m'a transmis. Merci à Charles pour m'avoir donné l'ambition de viser haut au cours de nos trois années à Télécom. Merci à Nicolas, Maxime, Axel, Baptiste d'être toujours là quand il s'agit de passer de bon moments, de voyager ou de faire du kite surf. Je remercie également tous ceux que j'ai rencontrés en Nouvelle-Zélande: Oscar qui m'a réservé le meilleur accueil, Pablo et Aurélien pour nos excursions folles, et bien sûr à Marie pour nos randonnées exotiques et pour m'avoir partagé sa passion du fromage. Merci à Hugo, ami d'enfance qui a toujours été là pour refaire le monde dans un bar, partager un verre dans un théâtre, partir en weekend à l'étranger ou tout simplement pour être ensemble sans rien avoir à se dire. Je lui

souhaite le meilleur dans la poursuite de sa carrière et qu'elle soit à la hauteur de son talent.

Je gardais finalement le meilleur pour la fin. Merci à ma famille d'être là dans les bons comme les mauvais moments, de m'offrir des temps de détente et des bons petits plats bourguignons. Il sont là depuis ma plus tendre enfance et sont ceux qui m'ont donné le goût de la découverte et de la prise de risque en m'assurant une zone de réconfort où je peux toujours retourner.

# Contents

<b>I. French and English introductions</b>	<b>17</b>
1. Introduction (French)	19
2. Introduction (English)	31
<b>II. Player consistency for ranking and optimization</b>	<b>41</b>
3. Time consistent mappings	45
4. Three classes of time consistent mappings	67
5. Player consistency and decomposition of optimization problems	89
<b>III. Risk averse competitive equilibrium</b>	<b>105</b>
6. On risk averse competitive equilibrium	107
7. Equilibrium and player consistency	125
<b>IV. Risk averse classification and estimation problems</b>	<b>129</b>
8. General risk measures for robust machine learning	131
<b>V. Appendix</b>	<b>155</b>
9. Recalls	157
10. Translation invariant mappings on ordered group lattices	169
<b>VI. Bibliography</b>	<b>173</b>



# Contents detailed

<b>I. French and English introductions</b>	<b>17</b>
<b>1. Introduction (French)</b>	<b>19</b>
1.1. Contexte de la thèse . . . . .	20
1.2. Résumé et contribution de la cohérence entre joueurs pour classer et optimiser . . . . .	20
1.3. Résumé et contributions des équilibres compétitifs avec risque . . . . .	25
1.4. Résumé et contributions des problèmes d'estimation et de classification avec sensibilité au risque . . . . .	27
1.5. Résumé des contributions de ce manuscrit et perspectives de travail . . . . .	29
<b>2. Introduction (English)</b>	<b>31</b>
2.1. Context of the thesis . . . . .	32
2.2. Summary and contribution of player consistency for ranking and optimization . . . . .	32
2.3. Summary and contributions on risk averse competitive equilibrium . . . . .	36
2.4. Summary and contribution on risk averse classification and estimation problems . . . . .	39
<b>II. Player consistency for ranking and optimization</b>	<b>41</b>
<b>3. Time consistent mappings</b>	<b>45</b>
3.1. Introduction . . . . .	46
3.2. Review of the literature . . . . .	48
3.3. Main result: equivalence between time consistency and nested formula . . . . .	49
3.4. Revisiting the literature . . . . .	58
3.5. Additional characterizations of time consistency . . . . .	61
3.6. Conclusion . . . . .	65
<b>4. Three classes of time consistent mappings</b>	<b>67</b>
4.1. Introduction . . . . .	68
4.2. Time consistent translation invariant mappings . . . . .	68
4.3. Time consistent convex mappings . . . . .	75
4.4. Time consistent supremum mappings . . . . .	81
4.5. Conclusion and perspectives . . . . .	87
<b>5. Player consistency and decomposition of optimization problems</b>	<b>89</b>
5.1. Introduction . . . . .	90
5.2. From time consistency to player consistency . . . . .	90

5.3. Application of player consistency to decomposition for optimization . . . . .	96
5.4. Conclusion . . . . .	102
<b>III. Risk averse competitive equilibrium</b>	<b>105</b>
<b>6. On risk averse competitive equilibrium</b>	<b>107</b>
6.1. Introduction . . . . .	108
6.2. Statement of problem . . . . .	109
6.3. Some equivalences between social planner and equilibrium problems . . . . .	113
6.4. Multiple risk averse equilibrium . . . . .	116
6.5. Analytical results . . . . .	119
6.6. Unstability of equilibrium . . . . .	123
6.7. Conclusion . . . . .	123
<b>7. Equilibrium and player consistency</b>	<b>125</b>
7.1. Introduction . . . . .	126
7.2. Player consistency and equilibrium with risk neutral players . . . . .	126
7.3. Player consistency and equilibrium with risk averse players . . . . .	128
7.4. Conclusion . . . . .	128
<b>IV. Risk averse classification and estimation problems</b>	<b>129</b>
<b>8. General risk measures for robust machine learning</b>	<b>131</b>
8.1. Introduction . . . . .	132
8.2. Problem statement . . . . .	133
8.3. Convex formulation of robust inference problems using risk measures . . . . .	134
8.4. Examples of risks measures . . . . .	138
8.5. Numerical solution . . . . .	146
8.6. Simulation example . . . . .	148
8.7. Conclusion . . . . .	151
<b>V. Appendix</b>	<b>155</b>
<b>9. Recalls</b>	<b>157</b>
9.1. Recall on order and relations . . . . .	157
9.2. Recall on lattices and partitions . . . . .	160
9.3. Recall on stochastic kernels . . . . .	161
9.4. Recall on star-difference . . . . .	162
9.5. Recall on risk measures . . . . .	163
9.6. Recall on Moreau's transform and Moreau's addition . . . . .	166
<b>10. Translation invariant mappings on ordered group lattices</b>	<b>169</b>
10.1. Translation invariant mappings on lattices . . . . .	169

**VI. Bibliography**

**173**





## **Part I.**

### **French and English introductions**



# 1. Introduction (French)

## Contents

---

1.1. Contexte de la thèse . . . . .	20
1.2. Résumé et contribution de la cohérence entre joueurs pour classer et optimiser . . . . .	20
1.3. Résumé et contributions des équilibres compétitifs avec risque . . . . .	25
1.4. Résumé et contributions des problèmes d'estimation et de classification avec sensibilité au risque . . . . .	27
1.5. Résumé des contributions de ce manuscrit et perspec- tives de travail . . . . .	29

---

## 1.1. Contexte de la thèse

Cette thèse présente le travail que j'ai réalisé durant trois années, d'octobre 2015 à octobre 2018, sous la direction de Michel De Lara et Jean-Christophe Pesquet.

J'ai été co-financé par l'École Nationale des Ponts et Chaussées et par le Labex Bézout (Laboratoire d'Excellence). J'ai effectué la majeure partie de mon travail au Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique, allant périodiquement au Centre de Vision Numérique de Centrale Supélec. Pendant ma thèse, j'ai eu l'opportunité d'aller en Nouvelle-Zélande d'octobre 2016 à avril 2017 où j'ai travaillé sous la direction d'Andy Philpott. Ce projet de collaboration a été rendu possible grâce au financement de l'ambassade de France en Nouvelle-Zélande.

Dans cette thèse, nous considérons des problèmes d'optimisation stochastique, de théorie des jeux et d'apprentissage automatique avec des mesures de risque.

Dans une première partie, nous mettons l'accent sur la cohérence temporelle en optimisation stochastique. Nous commençons par prouver une équivalence entre cohérence temporelle et existence d'une formule imbriquée pour des fonctions. Motivés par des exemples bien connus dans les mesures de risque, nous étudions trois classes de fonctions : les fonctions *invariantes par translation*, les *transformées de Fenchel-Moreau* et les fonctions *supremum*. Ensuite, nous étendons le concept de cohérence temporelle à la cohérence entre joueurs, en remplaçant le temps séquentiel par un ensemble non ordonné et les fonctions par des relations binaires. Enfin, nous montrons comment la cohérence entre joueurs est liée à des formes de décomposition séquentielle et parallèle usuelles en optimisation.

Dans une seconde partie, nous étudions l'impact des mesures de risque sur la multiplicité des équilibres, dans les problèmes de jeux dynamiques, quand les marchés sont complets ou incomplets. Nous concevons un exemple où l'introduction de mesures de risque conduit à l'existence de trois équilibres au lieu d'un dans le cas risque neutre. Nous analysons la capacité de deux algorithmes différents à trouver les trois équilibres. Enfin, nous discutons des liens entre cohérence entre joueurs, système de prix et problèmes d'équilibre dans les jeux.

Dans une troisième partie, nous étudions l'application de l'optimisation robuste à l'apprentissage automatique. En utilisant des mesures de risque convexes, nous fournissons un cadre unifié et proposons un algorithme adapté couvrant trois catégories d'ensembles d'ambiguïté étudiés dans la littérature.

## 1.2. Résumé et contribution de la cohérence entre joueurs pour classer et optimiser

Les problèmes d'optimisation stochastique sont naturellement de grande taille car ils sont indexés par les scénarios et le temps (et éventuellement une structure spatiale). Pour les résoudre numériquement, on peut exploiter des schémas de décomposition. Au lieu de résoudre un problème de grande taille, on résout séquentiellement ou en parallèle un ensemble de sous-problèmes plus petits.

Les méthodes de décomposition sont beaucoup étudiées avec une espérance mathématique et un critère additif. Nous présentons ici comme motivation deux schémas

## 1.2. Résumé et contribution de la cohérence entre joueurs pour classer et optimiser

simples et évidents de décomposition : un pour la décomposition parallèle et un pour la décomposition séquentielle. Nous présentons ensuite un cadre plus abstrait impliquant des mesures de risque. Enfin, nous présentons notre contribution dans le cas où l'espérance est remplacée par une application non-linéaire et où le critère n'est pas nécessairement additif.

**Méthodes de décomposition avec espérance et critère additif.** Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace de probabilité et soit  $\mathcal{X} = L^\infty(\Omega, \mathcal{T}, \mathbb{P})$  l'ensemble des variables aléatoires bornées. Considérons un problème d'optimisation stochastique dont la formulation est

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] , \quad (1.1)$$

où  $c_1 : \mathbb{R} \rightarrow \mathbb{R}$  et  $c_2 : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions.

Il existe deux options naturelles pour décomposer le problème (1.1) en utilisant les propriétés de l'espérance. Tout d'abord, par linéarité de l'espérance, on obtient de façon évidente un *schéma de décomposition parallèle* qui s'exprime ainsi

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] = \min_{\mathbf{X}_1 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1)] + \min_{\mathbf{X}_2 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_2(\mathbf{X}_2)] . \quad (1.2)$$

Au lieu de résoudre un problème incluant deux variables  $\mathbf{X}_1$  et  $\mathbf{X}_2$ , on peut résoudre simultanément le problème  $\min_{\mathbf{X}_1 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1)]$  et le problème  $\min_{\mathbf{X}_2 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_2(\mathbf{X}_2)]$ . On agrège ensuite les valeurs optimales de chaque problème pour obtenir la valeur optimale du problème initial (1.1).

On considère à présent le problème

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, \mathbf{X}_2)] , \quad (1.3)$$

où  $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  est une fonction,  $\mathcal{F}_1 \subset \mathcal{F}_2$  sont deux tribus et où  $\mathbf{X}_1$  est  $\mathcal{F}_1$ -mesurable et  $\mathbf{X}_2$  est  $\mathcal{F}_2$ -mesurable. Une seconde approche de décomposition se fonde sur la propriété de formule imbriquée de l'espérance conditionnelle

$$\mathbb{E}_{\mathbb{P}} [\mathbf{X}] = \mathbb{E}_{\mathbb{P}} [\mathbb{E}_{\mathbb{P}} [\mathbf{X} \mid \mathcal{F}]] ,$$

où  $\mathcal{F}$  est une tribu incluse dans la tribu la plus fine  $\mathcal{T}$ . En utilisant cette propriété, on obtient de façon évidente un *schéma de décomposition séquentielle* qui s'exprime ainsi:

$$\min_{\mathbf{X}_1 \preceq \mathcal{F}_1, \mathbf{X}_2 \preceq \mathcal{F}_2} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, \mathbf{X}_2)] = \min_{\mathbf{X}_1 \preceq \mathcal{F}_1} \mathbb{E}_{\mathbb{P}} \left[ \min_{x_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2) \mid \mathcal{F}_2] \right] . \quad (1.4)$$

Le problème (1.4) est décomposable en temps en utilisant la programmation dynamique (Bellman, 1957; Bertsekas, 1976, 1987). On calcule tout d'abord pour chaque  $\mathbf{X}_1$ , la solution optimale du problème intérieur  $\min_{x_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2) \mid \mathcal{F}_2]$ . Cela conduit à la solution optimale  $x_2^*(\mathbf{X}_1)$  qui est paramétrisée par  $\mathbf{X}_1$  (problèmes de mesurabilité mis à part). Nous pouvons alors résoudre le problème d'optimisation externe

$$\min_{\mathbf{X}_1 \in \mathcal{F}_1} \mathbb{E}_{\mathbb{P}} \left[ \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2^*(\mathbf{X}_1)) \mid \mathcal{F}_2] \right] . \quad (1.5)$$

## 1. Introduction (French)

**Introduction de la notion de risque et de critères non additifs.** Les problèmes d’optimisation (1.1) and (1.3) peuvent être incorporés dans un cadre plus général et s’écrivent alors

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \rho(c_1(\mathbf{X}_1) \oplus c_2(\mathbf{X}_2)) , \quad (1.6a)$$

ou

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}, \mathbf{X}_2 \preceq_{\mathcal{F}_2}} \rho(c(\mathbf{X}_1, \mathbf{X}_2)) . \quad (1.6b)$$

où  $\rho : L^\infty(\Omega, \mathcal{T}, \mathbb{P}) \rightarrow \mathbb{R}$  est une mesure de risque et  $\oplus$  est un opérateur qui “agrège” les variables aléatoires. À titre d’illustration, dans la gestion des réseaux électriques, un critère non additif évalué sous une mesure de risque non linéaire peut améliorer la sécurité du réseau et conduire à des politiques plus fiables (Laetitia Andrieu, 2009), par exemple pour intégrer les énergies renouvelables.

Détaillons un exemple d’agrégateur  $\oplus$  et de mesure de risque  $\rho$  pour motiver l’emploi de ces notations. En lieu et place de minimiser une somme de coûts, on peut s’intéresser à

- minimiser le maximum des coûts des joueurs; dans ce cas, l’agrégateur  $\oplus$  est l’opérateur maximum  $\max$ ;
- minimiser un coût final actualisé avec un taux d’intérêt contrôlé; dans ce cas, l’agrégateur  $\oplus$  est l’opérateur produit  $\times$ .

Ainsi, les formulations (1.6) nous permet de prendre en compte une classe plus large de critère de modélisation.

Nous détaillons à présent un exemple de mesure de risque  $\rho$ . Une mesure de risque très utilisée en optimisation stochastique est l’*Average Value at Risk de niveau*  $\beta \in [0, 1)$  qui est donnée par la formule (Rockafellar and Uryasev, 2000a)

$$\text{AV@R}_\beta[\mathbf{X}] = \inf_{\alpha} \left\{ \alpha + \frac{1}{1-\beta} \mathbb{E}_{\mathbb{P}}[\max(0, \mathbf{X} - \alpha)] \right\} . \quad (1.7)$$

Une telle mesure de risque se concentre sur les  $100 \cdot (1 - \beta)\%$  pires scénarios, c’est-à-dire les scénarios conduisant aux coûts les plus élevés. Si  $\beta = 0$ , nous retrouvons la définition de l’espérance mathématique et si  $\beta$  tend vers 1, nous retrouvons l’expression du pire des cas comme limite. Optimiser en utilisant la mesure de risque AV@R conduit à une solution plus robuste dans le sens où, si un “mauvais” scénario se réalise, notre solution sera meilleure que celle donnée par le problème (1.1). Cependant, la solution obtenue en résolvant (1.6a) aura un coût moyen plus élevé. On dit alors qu’on est *averse* au risque: on préfère payer un petit peu plus dans chaque scénario, si cela diminue les coûts les plus élevés.

Maintenant que nous avons présenté les ingrédients principaux, une question se pose: peut-on adapter les méthodes de décompositions parallèles additives (1.2) et séquentielles imbriquées (1.4) aux nouveaux problèmes (1.6). Premièrement, une mesure de risque quelconque n’est pas linéaire ce qui conduit à l’inégalité

$$\rho(\mathbf{X}_1 + \mathbf{X}_2) \neq \rho(\mathbf{X}_1) + \rho(\mathbf{X}_2) . \quad (1.8)$$

## 1.2. Résumé et contribution de la cohérence entre joueurs pour classer et optimiser

Deuxièmement, il est connu que

$$AV@R_\beta[\mathbf{X}] \neq AV@R_\beta[AV@R_\beta[\mathbf{X} \mid \mathcal{F}]] , \quad (1.9)$$

où l'application  $AV@R_\beta[\cdot \mid \mathcal{F}]$  est l'AV@R conditionnelle que nous ne détaillons pas ici. Ainsi, une décomposition additive et une formule imbriquée ne sont pas triviales et les décompositions parallèles et séquentielles nécessitent plus de travail que dans le cas de l'espérance avec critère additif.

**Mélanger les mesures de risque et les méthodes de décomposition.** Motivés par la question “quelles sont les mesures de risque compatibles avec la décomposition en optimisation ?”, nous avons en premier lieu étudié la notion de cohérence temporelle. Cela fournit en effet une formule imbriquée qui rend la décomposition séquentielle possible, comme expliqué par exemple dans Shapiro (2009) et Ruszczyński (2010). Nous avons pu tracer la notion de cohérence temporelle jusqu'à Koopmans (1960). C'est un sujet largement discuté dans la littérature comme l'atteste l'étude Bielecki, Cialenco, and Pitera (2017). Cette étude montre aussi que les cadres théoriques mis en place pour parler de cohérence temporelle sont très disparates, mixant cohérence temporelle proprement dite avec des hypothèses additionnelles.

Notre contribution est résumée ci-dessous.

- Dans le Chap. 3, nous présentons un cadre mathématique pour la cohérence temporelle entre les fonctions. Ce cadre englobe des résultats obtenus par différents auteurs. Considérons un objet mathématique qui est divisé entre un début  $h$  et une fin  $t$  (par exemple, un processus stochastique entre les dates 0 et  $T$  peut être découpé en un début entre les dates 0 et  $t$  et une fin entre les dates  $t$  et  $T$  avec  $0 \leq t \leq T$ ). Étant donné une application  $A$  qui agrège le début et la fin de l'objet et étant donné une application  $F$  qui évalue seulement la fin de l'objet, nous cherchons à satisfaire la propriété suivante :

$$F(t) = F(t') \Rightarrow A(h, t) = A(h, t') , \quad \forall h \in \mathbb{H} , \quad \forall (t, t') \in \mathbb{T}^2 . \quad (1.10)$$

Autrement dit, nous considérons deux processus ayant le même début  $h$ . Si l'évaluation de la fin est égale pour les deux processus, alors la propriété de cohérence temporelle stipule que les évaluations des deux processus en entier doivent être égales. Notre contribution principale est d'avoir montré une équivalence entre la propriété de cohérence temporelle (1.10) et la formule imbriquée suivante

$$A(h, t) = S^{A,F}(h, F(t)) , \quad (1.11)$$

où  $S^{A,F}$  est une application bien choisie. (Nous détaillons brièvement les équations (2.10) et (2.11) dans le cas risque neutre afin d'illustrer notre propos. Nous avons  $A(h, t) = \mathbb{E}_\mathbb{P}[h + t]$ ,  $F(t) = \mathbb{E}_\mathbb{P}[t \mid \mathcal{F}]$  et  $S^{A,F}(h, f) = \mathbb{E}_\mathbb{P}[h + f]$ ). Nous revisitons ensuite la littérature à la lumière de ce cadre mathématique. Cela nous permet de mettre en avant la contribution des différents auteurs. Nous montrons en effet que telle hypothèse additionnelle conduit à tel résultat supplémentaire. Par exemple, Kreps et Porteus ((Kreps and Porteus, 1978a), (Kreps and Porteus, 1979)) énoncent un axiome de cohérence temporelle. Avec



## 1. Introduction (French)

notre cadre abstrait, nous déduisons directement l'existence d'une application  $S^{A,F}$  croissante en son deuxième argument ainsi qu'une formule imbriquée; tandis que Kreps et Porteus obtiennent un résultat plus fort sous des hypothèses plus fortes. En effet, ils ajoutent des hypothèses de continuité, de substitution (reliée à la convexité) et utilisent des inégalités strictes. Cela leur permet d'obtenir que l'application  $S^{A,F}$  est continue et strictement croissante en son deuxième argument.

- Dans Chap. 4, nous nous intéressons à la cohérence temporelle pour trois classes de fonctions. Premièrement, nous caractérisons la cohérence temporelle pour les fonctions qui sont invariantes par translation en utilisant leur ensemble de niveau 0 <sup>1</sup>. La cohérence temporelle pour cette classe de fonctions englobe dans un cadre abstrait les résultats sur les mesures de risque monétaires. Cependant, nous obtenons une caractérisation qui s'avère difficile à utiliser dans des exemples pratiques. En deuxième lieu, nous étudions les fonctions définies comme transformées de Fenchel-Moreau. Nous donnons des conditions pour que de telles fonctions satisfassent une formule imbriquée. Cela nous permet de retrouver les résultats de [Ruszczyński and Shapiro \(2006a\)](#) et [Föllmer and Schied \(2016\)](#). En troisième lieu, nous étudions comment la propriété de cohérence temporelle se transmet pour les fonctions définies comme des supremums.
- Dans Chap. 5, nous introduisons la notion de cohérence sans parler de la notion de temps. Le fait est que dans des problèmes à plusieurs étapes, la cohérence temporelle est liée à l'existence d'un coordinateur inter-temporel. Nous pouvons donc interpréter un problème à plusieurs étapes comme plusieurs joueurs jouant successivement. Nous définissons ainsi la notion de cohérence entre joueurs. Dans (1.10) et (1.11), l'ensemble  $\mathbb{H}$  représente les éléments qui "se produisent avant" les éléments de l'ensemble  $\mathbb{T}$ . Nous oublions ici cette idée et considérons symétriquement les ensembles  $\mathbb{H}$  et  $\mathbb{T}$ . Nous cherchons une propriété que nous appelons la cohérence entre joueurs. Celle-ci est définie entre trois fonctions  $A$ ,  $F_{\mathbb{H}}$  et  $F_{\mathbb{T}}$  et s'énonce

$$\left. \begin{array}{l} F_{\mathbb{H}}(h) = F_{\mathbb{H}}(h') \\ F_{\mathbb{T}}(t) = F_{\mathbb{T}}(t') \end{array} \right\} \Rightarrow A(h, t) = A(h', t'). \quad (1.12)$$

Sous des hypothèses techniques que nous ne détaillons pas, nous pouvons prouver que l'Équation (1.12) est équivalente à l'existence d'une fonction  $S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}$  telle que

$$A(h, t) = S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}(F_{\mathbb{H}}(h), F_{\mathbb{T}}(t)). \quad (1.13)$$

La notion de cohérence temporelle et de *cohérence entre joueurs* sont toutes deux un cas particulier du cadre plus général que nous dénommons la  $\mathcal{R}$ -cohérence. Étant donné un ensemble  $P$  de joueurs, un  $n$ -uplet de relations

---

<sup>1</sup>Dans le cadre théorique des mesures de risque, cet ensemble de niveau est appelé ensemble acceptable.

### 1.3. Résumé et contributions des équilibres compétitifs avec risque

$(\mathcal{S})_{p \in P}$  et une relation produit  $\mathcal{R}$ , nous définissons la  $\mathcal{R}$ -cohérence pour les relations binaires comme suit

$$\underbrace{\prod_{p \in P} \mathcal{S}_p}_{\text{produit de relations}} \subset \underbrace{\mathcal{R}}_{\text{relation sur un ensemble produit}} . \quad (1.14)$$

Nous détaillons ensuite comment manipuler la cohérence entre joueurs pour décomposer des problèmes d'optimisation de manière séquentielle et parallèle. Nous sommes ainsi capables, sous des hypothèses spécifiques, d'obtenir la décomposition imbriquée suivante

$$\inf_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h,t) = \inf_{h \in \mathbb{H}} S^{A,F}(h, \inf_{t \in \mathbb{T}} F(t)) , \quad (1.15)$$

et la décomposition parallèle suivante

$$\inf_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h,t) = S^{A,F_{\mathbb{H}},F_{\mathbb{T}}}\left(\inf_{h \in \mathbb{H}} F_{\mathbb{H}}(h), \inf_{t \in \mathbb{T}} F_{\mathbb{T}}(t)\right) . \quad (1.16)$$

Nous décrivons aussi comment appliquer ces décompositions de façon récursive.

## 1.3. Résumé et contributions des équilibres compétitifs avec risque

Nous avons introduit la notion de cohérence entre joueurs et nous avons expliqué comment cette notion peut aider à décomposer des problèmes d'optimisation stochastique séquentiellement et parallèlement. Une méthode classique pour décomposer en parallèle un problème d'optimisation est de le reformuler comme un problème d'équilibre. L'idée est de dualiser les contraintes de couplage puis d'interpréter les multiplicateurs comme des prix qui coordonnent différents joueurs. Nous présentons, dans un but d'illustration, le cas risque neutre. Ensuite, nous détaillons nos contributions.

**Équilibre risque neutre et décomposition d'un problème d'optimisation stochastique.** Considérons un problème jouet. Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace de probabilités où  $\Omega$  est fini et où  $\mathbb{P}$  charge tous les points. Soit  $L^\infty(\Omega, \mathcal{T}, \mathbf{X})$  l'espace des variables aléatoires. Soit  $\mathcal{X}_1$  et  $\mathcal{X}_2$  deux sous-ensembles de  $L^\infty(\Omega, \mathcal{T}, \mathbf{X})$ . Nous considérons le problème d'optimisation stochastique suivant

$$\min_{(\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{X}_1 \times \mathcal{X}_2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] , \quad (1.17a)$$

$$\text{s.c. } g_1(\mathbf{X}_1) + g_2(\mathbf{X}_2) \geq \mathbf{D} , \quad \mathbb{P} - \text{p.s.} \quad (1.17b)$$

où

- les fonctions  $c_1 : \mathbb{R} \rightarrow \mathbb{R}$  et  $c_2 : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions coût,
- les fonctions  $g_1 : \mathbb{R} \rightarrow \mathbb{R}$  et  $g_2 : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions production,

## 1. Introduction (French)

- la variable aléatoire  $\mathbf{D}$  représente une demande agrégée.

Nous considérons à présent le problème d'équilibre risque neutre qui s'écrit

$$\min_{\mathbf{X}_1 \in \mathcal{X}_1} \mathbb{E}_{\mathbb{P}}[c_1(\mathbf{X}_1)] \quad (1.18)$$

$$\min_{\mathbf{X}_2 \in \mathcal{X}_2} \mathbb{E}_{\mathbb{P}}[c_2(\mathbf{X}_2)] \quad (1.19)$$

$$0 \leq \boldsymbol{\lambda} \perp \mathbf{D} - (g_1(\mathbf{X}_1) + g_2(\mathbf{X}_2)) \geq 0, \quad (1.20)$$

où  $\boldsymbol{\lambda}$  sont les variables de prix de marché. La contrainte (1.20) d'écart complémentaire exprime le fait que les prix doivent être choisis de telle sorte que l'offre soit égale à la demande.

Le premier théorème du bien-être énonce que les problèmes (2.17) et (2.18) sont équivalents, c'est-à-dire, le marché "trouve" une solution qui est optimale d'un point de vue social. Ainsi, on obtient, pour le problème d'optimisation (1.17), une méthode de décomposition en un problème d'équilibre où chaque joueur résout son sous-problème d'optimisation.

L'idée derrière le premier théorème du bien-être est la suivante. Sous des hypothèses techniques, le problème (1.17) est équivalent à son dual qui se formule

$$\max_{\boldsymbol{\lambda}} \min_{(\mathbf{X}_1, \mathbf{X}_2)} \mathbb{E}_{\mathbb{P}}[c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] + \sum_{\omega \in \Omega} \boldsymbol{\lambda}(\mathbf{D}(\omega) - \omega)(g_1(\mathbf{X}_1)(\omega) + g_2(\mathbf{X}_2)(\omega)). \quad (1.21)$$

Le problème interne peut à présent être résolu joueur par joueur. Cette méthode de décomposition est appelée méthode de décomposition par les prix. À l'optimum, on satisfait la condition des écarts complémentaires (1.20).

Nous étudions ce qui se passe quand on remplace l'espérance mathématique par une mesure de risque cohérente.

**Équilibre averse au risque et décomposition d'un problème d'optimisation stochastique.** Nous étudions les équilibres averse au risque motivés par les marchés de l'électricité. En effet, la plupart des régions industrialisées du monde ont, au cours des trente dernières années, établi des marchés de gros de l'électricité qui prennent la forme d'une vente aux enchères correspondant à l'offre et à la demande. Ces marchés sont souvent appelés systèmes à deux règlements.

Si les distributions de probabilité pour les modes de production intermittents (éolien, solaire, ...) sont connus, il est possible de maximiser le bien-être total des producteurs et des consommateurs dans chaque scénario. Après de nombreuses répétitions du modèle, un bénéfice total à long terme est maximisé. Maximiser le bien-être attendu peut être modélisé comme un programme stochastique en deux étapes. La méthode de calcul des prix et des mécanismes de paiement dans un marché stochastique est décrite dans un certain nombre d'articles (voir [Pritchard et al. \(2010\)](#), [Wong and Fuller \(2007\)](#) et [Zakeri et al. \(2016\)](#)). Lorsqu'elle est évaluée en utilisant la distribution de probabilité supposée sur l'offre, il est possible de démontrer que la compensation est plus efficace que les systèmes à deux règlements.

Si les agents de ces systèmes sont réticents à prendre des risques, on peut aussi chercher à maximiser le "bien-être social averse au risque". Dans ce contexte, le

calcul des prix et des paiements aux agents devient plus compliqué. Si les agents utilisent des mesures de risque cohérentes, il est possible de définir un marché complet du risque dans un sens précis. Si le marché est complet, un équilibre partiel parfaitement concurrentiel maximisera également le bien-être social averse au risque, c'est-à-dire qu'il est efficace. D'un autre côté, si le marché du risque n'est pas complet, alors un équilibre partiel parfaitement compétitif peut être inefficace. Cela a été exploré dans un certain nombre d'articles (voir par exemple de [Maere d'Aertrycke et al. \(2017\)](#), [Ehrenmann and Smeers \(2011\)](#) et [Ralph and Smeers \(2015\)](#)).

- Dans le Chap. 6, nous étudions une classe de mécanismes de répartition et de tarification stochastiques en supposant que les agents tenteront de maximiser, à prix donné, leur bien-être averse au risque. Les agents ont des mesures de risque cohérentes et sont supposés agir comme s'ils n'avaient aucune influence sur les prix sur les marchés de l'énergie et du risque. Nous visons à éclairer certaines difficultés qui surviennent lorsque les marchés à risque ne sont pas complets. Nous décrivons une instance simple d'un marché stochastique qui a trois équilibres différents. Deux de ces points sont stables dans le sens de [Samuelson \(1941\)](#) et sont des attracteurs *d'algorithmes de tâtonnement*. Le troisième équilibre est instable, mais la solution est fournie par le solveur PATH bien connu dans GAMS (Voir [Ferris and Munson \(2000\)](#)). Notre exemple illustre la délicatesse de la recherche de solutions numériques pour les équilibres dans les marchés incomplets. Puisque ceux-ci sont utilisés pour justifier des décisions, la non-unicité des solutions dans ce cadre est indésirable. En effet, nous ne pouvons pas contrôler vers quel équilibre les différents algorithmes vont converger.
- Dans Chap. 7 nous discutons des liens entre la cohérence entre joueurs et les problèmes d'équilibre. Nous suggérons une nouvelle interprétation des prix des marchés en tant qu'outils pour obtenir de la cohérence entre les joueurs et le planificateur central. En un sens, les prix ne sont pas seulement des mécanismes de coordination efficaces mais aussi des mécanismes qui induisent de la cohérence.

## 1.4. Résumé et contributions des problèmes d'estimation et de classification avec sensibilité au risque

Dans [Rockafellar and Uryasev \(2013\)](#) et [Rockafellar, Uryasev, and Zabarankin \(2008\)](#), les auteurs présentent un quadrangle fondamental qui fournit un lien entre problèmes d'estimation et de contrôle. Nous suivons leur chemin pour fournir des liens entre l'optimisation averse au risque et l'optimisation robuste. Plus précisément, nous montrons que les problèmes d'optimisation robustes sont des problèmes d'optimisation stochastique avec une mesure de risque bien choisie qui possède des propriétés agréables pour l'optimisation.

- Dans Chap. 8, nous nous intéressons à l'étude du risque dans l'apprentissage

## 1. Introduction (French)

automatique. Face à un problème d'apprentissage machine, où l'on doit classer certaines données ou prédire certaines quantités, la robustesse des solutions reste un problème. Aucune borne de l'erreur n'est généralement fournie et [Papernot, McDaniel, and Goodfellow \(2016\)](#) et [Kurakin, Goodfellow, and Bengio \(2016\)](#) présentent des exemples où de petites modifications sur l'entrée peuvent changer radicalement la solution en sortie.

Le manque de robustesse se produit également dans les problèmes d'optimisation face à l'incertitude sur les paramètres. Dans [Ben-Tal and Nemirovski \(2000\)](#), les auteurs ont montré qu'une petite modification des paramètres peut transformer une solution réalisable en une solution irréalisable.

La robustesse apparaît ainsi comme un moyen de contrôler les performances hors échantillon. Il existe une vaste littérature traitant des formulations de problèmes robustes et nous renvoyons le lecteur à [Ben-Tal, El Ghaoui, and Nemirovski \(2009\)](#). Une première approche consiste à introduire des contraintes en probabilité. Sous certaines conditions, ceci est équivalent à une approche basée sur des ensembles d'ambiguïtés. Nous détaillons pour plus de clarté la formulation d'un problème robuste reposant sur des ensembles d'ambiguïté. Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace de probabilité et  $\mathcal{X}$  un ensemble de variables aléatoires sur cet espace. Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction. Le problème d'optimisation classique s'écrit

$$\inf_{x \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{X})], \quad (1.22)$$

et son équivalent robuste avec ensemble d'ambiguïté s'écrit

$$\inf_{x \in \mathcal{X}} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{X})], \quad (1.23)$$

où  $\mathcal{Q}$  est un ensemble de lois de probabilité sur  $\Omega$ .

[Ben-Tal, Den Hertog, De Waegenare, Melenberg, and Rennen \(2013\)](#); [Hu and Hong \(2013\)](#); [Duchi, Glynn, and Namkoong \(2016\)](#); [Moghaddam and Mahlooji \(2016\)](#) et [Namkoong and Duchi \(2016\)](#) ont largement contribué dans le domaine des ensembles d'ambiguïté. Dans [Esfahani and Kuhn \(2015\)](#) et [Esfahani, Shafieezadeh-Abadeh, Hanasusanto, and Kuhn \(2017\)](#), les auteurs présentent un cadre théorique pour l'optimisation robuste fondé sur la distance de Wasserstein. On définit l'ensemble  $\mathcal{Q}$  dans (1.23) comme une boule de rayon  $\epsilon$  centrée sur la loi de référence  $\mathbb{P}$  en utilisant la distance de Wasserstein. Ensuite, on minimise sur ces lois de probabilité en essayant de se prémunir du pire des coûts.

Cette idée de se prémunir du pire des coûts est bien connue en finance. La représentation robuste des mesures de risque fournit un cadre théorique adéquat. Une classe de mesures de risque appropriée est celle des mesures de risque cohérentes qui furent introduites dans l'article précurseur de [Artzner, Delbaen, Eber, and Heath \(1999\)](#). [Föllmer and Schied \(2016\)](#) étudient une plus large classe de mesures de risque qui sont les mesures de risque convexes. Cela permet d'étendre certains résultats.

## 1.5. Résumé des contributions de ce manuscrit et perspectives de travail

Nous avons choisi pour ce chapitre l'approche de [Esfahani and Kuhn \(2015\)](#) qui est de reformuler un problème d'optimisation robuste avec ensemble d'ambiguïté comme un problème convexe.

Notre contribution est double. Premièrement, nous clarifions le lien qui existe entre mesure de risque et optimisation robuste. Cela permet d'utiliser des résultats de finance en apprentissage automatique. Deuxièmement, nous adaptons un algorithme de projection défini dans [Combettes \(2003\)](#) pour être capable de résoudre avec le même algorithme des problèmes robustes de grande dimension avec des ensembles  $\mathcal{Q}$  d'ambiguïté différents. Nous avons conduit des expériences numériques.

## 1.5. Résumé des contributions de ce manuscrit et perspectives de travail

### Résumé des contributions

Ce manuscrit est une contribution au domaine de l'optimisation avec risque et avec méthodes de décomposition.

- Dans le Chap. 3, nous avons présenté un cadre abstrait pour la cohérence temporelle après avoir revisité la littérature. Notre contribution principale est d'énoncer une équivalence entre la cohérence temporelle et une formule imbriquée.
- Dans Chap. 4, nous avons étudié trois classes de fonctions cohérentes temporellement : les fonctions invariantes par translation, les transformées de Fenchel-Moreau et les fonctions définies comme supremum. Pour chaque classe, nous avons donné des caractérisations de la cohérence temporelle ou des conditions pour obtenir de la cohérence temporelle.
- Dans Chap. 5, nous avons étendu le cadre de cohérence temporelle à la cohérence pour les relations binaires. Par la suite, cela nous a permis de définir une notion de cohérence entre joueurs. Nous avons ensuite détaillé comment la notion de cohérence entre joueurs peut être utilisée pour une décomposition séquentielle et parallèle en optimisation.
- Dans Chap. 6, nous avons étudié l'équilibre avec risque dans des marchés. Nous avons présenté un exemple qui montre des équilibres multiples même avec des hypothèses de stricte concavité. Le contenu de Chap. 6 a été publié dans *Operation Research Letter* (à modifications mineures près).
- Dans Chap. 7, nous avons donné une interprétation des prix du marché comme un outil pour créer des fonctions qui sont cohérentes entre joueurs.
- Dans Chap. 8, nous avons fourni un lien entre l'optimisation avec mesure de risque et l'optimisation robuste en apprentissage automatique. Nous avons proposé un algorithme qui permet de résoudre plusieurs problèmes différents

## 1. Introduction (French)

et nous avons montré que la projection différentielle est un outil qui permet d'obtenir du parallélisme.

### Perspectives de travail

Sans être exhaustif, nous récapitulons ici le travail qui pourrait être fait et les questions qui restent ouvertes.

- Dans le prolongement du Chap. 4, nous pourrions discuter de la possibilité de mélanger les différentes classes de fonctions présentées pour obtenir un meilleur aperçu de la cohérence temporelle. Nous devrions également nous concentrer sur la recherche d'exemples plus concrets pour motiver les différentes généralisations que nous avons étudiées.
- Les conditions que doivent satisfaire les fonctions convexes généralisées pour obtenir de la cohérence temporelle et une formule imbriquée sont difficiles à vérifier. Exhiber des fonctions qui ne sont pas convexes mais qui satisfont ces conditions reste une question ouverte. Nous pouvons étudier ce qu'il se passe lorsqu'on mixe les fonctions invariantes par translation, les transformées de Fenchel-Moreau et les fonction supremum pour construire de la cohérence temporelle.
- En apprentissage automatique, les diagrammes d'influence capturent l'interaction dynamique entre plusieurs joueurs. Ils pourraient fournir une application adaptée aux différentes méthodes de décomposition.
- Dans l'extension de Chap. 6, nous pourrions étudier l'algorithme de tâtonnement sur un problème d'équilibre à grande échelle en utilisant des algorithmes d'approximation classiques.
- Dans l'extension de Chap. 7, nous pourrions expliquer comment la complétion du marché est lié à la cohérence entre joueurs.
- Dans l'extension de Chap. 8, nous pourrions comparer notre algorithme avec d'autres méthodes existantes afin de comprendre les avantages et les inconvénients de ce que nous avons proposé.

## 2. Introduction (English)

### Contents

---

2.1. Context of the thesis . . . . .	32
2.2. Summary and contribution of player consistency for ranking and optimization . . . . .	32
2.3. Summary and contributions on risk averse competitive equilibrium . . . . .	36
2.4. Summary and contribution on risk averse classification and estimation problems . . . . .	39

---



## 2.1. Context of the thesis

This manuscript presents the work that I have done during three years of PhD, from October 2015 to October 2018, under the supervision of Michel De Lara and Jean-Christophe Pesquet.

I was cofinanced by École Nationale des Ponts et Chaussées and by Labex Bézout (Laboratories of Excellence). I did most of my work at Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique visiting periodically the Center for computer vision of Centrale Supélec. During my PhD, I had the opportunity to go to New-Zealand for 6 months between October 01, 2016 and March 31, 2017 where I have worked with Andy Philpott. At that time, I was financed by the French embassy in New Zealand.

In this thesis, we consider stochastic optimization, game theory and machine learning problems with risk measures.

In a first part, we focus on time consistency in stochastic optimization. We begin by proving an equivalence between time consistent mappings and the existence of a nested formula. Motivated by well known examples in risk measures, we investigate three classes of mappings: translation invariant, Fenchel-Moreau transform and supremum mappings. Then, we extend the concept of time consistency to player consistency, by replacing the sequential time by any unordered set and mappings by any relations. Finally, we show how player consistency relates to sequential and parallel forms of decomposition in optimization.

In a second part, we study how risk measures impact the multiplicity of equilibria in dynamic game problems in complete and incomplete markets. We design an example where the introduction of risk measures leads to the existence of three equilibria instead of one in the risk neutral case. We analyze the ability of two different algorithms to recover the different equilibria. We discuss links between player consistency, system of prices and equilibrium problems in games.

In a third part, we study distributionally robust optimization in machine learning. Using convex risk measures, we provide a unified framework and propose an adapted algorithm covering three ambiguity sets discussed in the literature.

## 2.2. Summary and contribution of player consistency for ranking and optimization

Stochastic optimization problems are naturally large scale since they are indexed scenarios and time (and possibly a spatial structure). To solve them numerically, one can exploit decomposition schemes. Instead of solving one big problem, you solve, sequentially or in parallel, a set of smaller subproblems.

Decomposition methods are widely studied with mathematical expectation taken together with an additive criterion. We present here two simple and obvious schemes for parallel decomposition and for sequential decomposition as a motivation. Then we present a more abstract framework involving risk measures. Finally we present our contribution with non-linear mappings and a criterion that is not necessarily additive.

## 2.2. Summary and contribution of player consistency for ranking and optimization

### Decomposition schemes with mathematical expectation and additive criterion

Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a probability space and let  $\mathcal{X} = L^\infty(\Omega, \mathcal{T}, \mathbb{P})$  be the set of bounded real random variables. Consider a stochastic optimization problem formulated as

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] , \quad (2.1)$$

where  $c_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $c_2 : \mathbb{R} \rightarrow \mathbb{R}$  are functions.

Let us show two natural options to decompose problem (2.1) using properties of the mathematical expectation. First, by linearity of the mathematical expectation, we obtain a *parallel decomposition scheme*, that is, obviously

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] = \min_{\mathbf{X}_1 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1)] + \min_{\mathbf{X}_2 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_2(\mathbf{X}_2)] . \quad (2.2)$$

Instead of solving one problem involving two variables  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , we can solve simultaneously the problem  $\min_{\mathbf{X}_1 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1)]$  and the problem  $\min_{\mathbf{X}_2 \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} [c_2(\mathbf{X}_2)]$ . Then we aggregate the optimal value of each problem to get the optimal value of the original problem (2.1).

We now consider the problem

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}, \mathbf{X}_2 \preceq_{\mathcal{F}_2}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, \mathbf{X}_2)] , \quad (2.3)$$

where  $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is a function,  $\mathcal{F}_1 \subset \mathcal{F}_2$  are two  $\sigma$ -fields and where  $\mathbf{X}_1$  is  $\mathcal{F}_1$ -measurable and  $\mathbf{X}_2$  is  $\mathcal{F}_2$ -measurable. A second approach is based on the tower property of conditional expectation, that is, the nested formula

$$\mathbb{E}_{\mathbb{P}}[\mathbf{X}] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[\mathbf{X} \mid \mathcal{F}]] ,$$

where  $\mathcal{F}$  is a  $\sigma$ -field that is included in  $\mathcal{T}$ . Using this property, we obtain a *sequential decomposition scheme*, that is, obviously,

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}, \mathbf{X}_2 \preceq_{\mathcal{F}_2}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, \mathbf{X}_2)] = \min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}} \mathbb{E}_{\mathbb{P}} \left[ \min_{x_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2) \mid \mathcal{F}_2] \right] . \quad (2.4)$$

Problem (2.4) is time decomposable using dynamic programming (Bellman, 1957; Bertsekas, 1976, 1987). We first compute for each possible  $\mathbf{X}_1$ , an optimal solution of the inner problem  $\min_{x_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2) \mid \mathcal{F}_2]$ . This leads to an optimal solution  $x_2^*(\mathbf{X}_1)$  that is parametrized by  $\mathbf{X}_1$  (measurability problems apart). Then we solve the outer optimization problem

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}} \mathbb{E}_{\mathbb{P}} \left[ \mathbb{E}_{\mathbb{P}} [c(\mathbf{X}_1, x_2^*(\mathbf{X}_1)) \mid \mathcal{F}_2] \right] . \quad (2.5)$$

**Introducing risk measure and non additive criterion** Problems (2.1) and (2.3) can be casted in a more general framework that reads

$$\min_{\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}^2} \rho(c_1(\mathbf{X}_1) \oplus c_2(\mathbf{X}_2)) , \quad (2.6a)$$

or

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}, \mathbf{X}_2 \preceq_{\mathcal{F}_2}} \rho(c(\mathbf{X}_1, \mathbf{X}_2)) . \quad (2.6b)$$

## 2. Introduction (English)

where  $\rho : L^\infty(\Omega, \mathcal{T}, \mathbb{P}) \rightarrow \mathbb{R}$  is a risk measure and  $\oplus$  is an ‘‘aggregator’’ of random variable. In the management of electricity transmission, a non additive criterion evaluated under a non linear risk measure can improve the safety of the network and lead to more reliable policies (Laetitia Andrieu, 2009), for example to integrate renewable energy. Let us detail examples of aggregator  $\oplus$  and risk measure  $\rho$  to motivate this framework.

Instead of minimizing a sum of costs, one can look for

- minimizing the maximum of the cost of the two players; in this case  $\oplus$  is the maximum operator  $\max$ ;
- minimizing a discounted final cost with a controlled interest rate; in this case,  $\oplus$  is the product operator  $\times$ .

Hence, formulations (2.6) account for a broader class of modeling criteria.

We now provide examples of a risk measure  $\rho$ . A popular risk measure in stochastic optimization is the so called Average Value at Risk of level  $\beta \in [0, 1)$  given by (Rockafellar and Uryasev, 2000a)

$$\text{AV@R}_\beta[\mathbf{X}] = \inf_{\alpha} \left\{ \alpha + \frac{1}{1-\beta} \mathbb{E}_{\mathbb{P}}[\max(0, \mathbf{X} - \alpha)] \right\}. \quad (2.7)$$

Such a risk measure focuses on the  $100 \cdot (1-\beta)\%$  worst scenarios, that is, the scenarios that lead to the highest cost. If  $\beta = 0$ , we retrieve the mathematical expectation and, if  $\beta$  tends to 1, we retrieve the worst case risk measure. Optimizing using this AV@R risk measure will lead to a more robust solution in the sense that if ‘‘bad’’ scenarios occurs, our solution will be better than the one given by problem (2.1). However, the solution obtain by solving (2.6a) will have a higher *expected* cost. We then say that we are *risk averse*: we prefer to pay a little bit more at the beginning to avoid high cost in extreme situations.

Now that we have presented main ingredients, a question arises: can we adapt the decomposition methods parallel additive (2.2) and nested sequential (2.4) to new problems (2.6) ? First, a general risk measure is not linear which leads to

$$\rho(\mathbf{X}_1 + \mathbf{X}_2) \neq \rho(\mathbf{X}_1) + \rho(\mathbf{X}_2). \quad (2.8)$$

Second, it is well known that

$$\text{AV@R}_\beta[\mathbf{X}] \neq \text{AV@R}_\beta[\text{AV@R}_\beta[\mathbf{X} \mid \mathcal{F}]], \quad (2.9)$$

where the mapping  $\text{AV@R}_\beta[\cdot \mid \mathcal{F}]$  is the conditional average value at risk that we do not detail here. Hence, a nested formula or an additive decomposition are not trivial and to obtain parallel or sequential decomposition requires more work.

**Mixing risk measures and decomposition schemes** Driven by the question: ‘‘what are the risk measures that are compatible with decomposition in optimization ? ‘‘, we first focus on the notion of time consistency as it provides a nested formula and makes sequential decomposition possible as detailed for example in Shapiro (2009)

## 2.2. Summary and contribution of player consistency for ranking and optimization

and [Ruszczyński \(2010\)](#). We could trace the notion of time consistency to [Koopmans \(1960\)](#) and it is widely discussed in the literature as attests the survey [Bielecki, Cialenco, and Pitera \(2017\)](#). This survey also shows that the frameworks to discuss time consistency are quite disparate, mixing time consistency with different assumptions.

Our contribution are summarized below.

- In Chap. 3, we present a framework for time consistency for mappings to bring into a common framework results obtained by different authors. Consider a mathematical object that is divided between a head part  $h$  and a tail part  $t$ , for example stochastic processes. Given a mapping  $A$  that "aggregates" the head part and the tail part and a mapping  $F$  that evaluates only the tail part, we look for the property that

$$F(t) = F(t') \Rightarrow A(h, t) = A(h, t'), \quad \forall h \in \mathbb{H}, \quad \forall (t, t') \in \mathbb{T}^2. \quad (2.10)$$

Our main contribution is to show an equivalence between Equation (2.10) and the following nested formula

$$A(h, t) = S^{A,F}(h, F(t)), \quad (2.11)$$

where  $S^{A,F}$  is a well chosen mapping. (We detail briefly Equations (2.10) and (2.11) in the risk neutral case for illustration. In a nutshell, we have that  $A(h, t) = \mathbb{E}_{\mathbb{P}}[h + t]$ ,  $F(t) = \mathbb{E}_{\mathbb{P}}[t \mid \mathcal{F}]$  and  $S^{A,F}(h, f) = \mathbb{E}_{\mathbb{P}}[h + f]$ . We then revisit the literature at the light of this framework. This allows us to highlight the contribution of the different authors making clearer which additional assumption leads to which additional result. For example, Kreps and Porteus ([\(Kreps and Porteus, 1978a\)](#), [\(Kreps and Porteus, 1979\)](#)) state a temporal consistency axiom. With our abstract setting, we directly deduce the existence of a mapping  $S^{A,F}$  increasing in its second argument and a Nested Formula, whereas they obtain a stronger result under stronger assumptions. Indeed, they add assumptions of continuity, substitution (related to convexity) using strict inequalities. This enables them to obtain a mapping  $S^{A,F}$  which is continuous and strictly increasing in its second argument.

- In Chap. 4, we investigate time consistency for three classes of mappings. First, we characterize time consistency for mappings that are translation invariant using their level sets of level 0 <sup>1</sup>. Time consistency for this class of mappings encompasses, in an abstract setting, results on monetary risk measures. However, it gives a time consistency characterization which is quite involved since using it on practical examples is difficult. Second, we study mappings that are defined as Fenchel-Moreau transforms. We give conditions on mappings to satisfy a nested formula. This allows to recover well known result of [Ruszczyński and Shapiro \(2006a\)](#) and [Föllmer and Schied \(2016\)](#). Third, we study inheritance of time consistency for mappings that are defined as supremum.

---

<sup>1</sup>In the framework of monetary risk measure, this set is called acceptance set

## 2. Introduction (English)

- In Chap. 5, we introduce the notion of consistency without relying on the notion of time. We are motivated by the idea that, in multistage problems, time consistency is linked to the existence of an intertemporal coordinator and we can interpret a multistage problem as several players playing successively. We hence define the notion of players consistency. In (2.10) and (2.11), the head set  $\mathbb{H}$  represents elements that "occurs before" elements of the tail set  $\mathbb{T}$ . Now, we forget this idea and we consider the sets  $\mathbb{H}$  and  $\mathbb{T}$  symmetrically. We look for a property, that we call *player consistency*, between three mappings  $A$ ,  $F_{\mathbb{H}}$  and  $F_{\mathbb{T}}$  that reads

$$\left. \begin{array}{l} F_{\mathbb{H}}(h) = F_{\mathbb{H}}(h') \\ F_{\mathbb{T}}(t) = F_{\mathbb{T}}(t') \end{array} \right\} \Rightarrow A(h, t) = A(h', t') . \quad (2.12)$$

Under mild assumption, we are able to prove that Equation (2.12) is equivalent to the existence of a mapping  $S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}$  such that

$$A(h, t) = S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}(F_{\mathbb{H}}(h), F_{\mathbb{T}}(t)) . \quad (2.13)$$

The notion of time consistency and player consistency are both encompassed in the definition of  $\mathcal{R}$ -consistency for binary relations. Given a set  $P$  of players, a tuple of relations  $(\mathcal{S})_{p \in P}$  and a product relation  $\mathcal{R}$ , we define  $\mathcal{R}$ -consistency for binary relations as follows

$$\underbrace{\prod_{p \in P} \mathcal{S}_p}_{\text{product of relations}} \subset \underbrace{\mathcal{R}}_{\text{relation over product set}} . \quad (2.14)$$

We then detail how to use player consistency to decompose optimization problems sequentially and in parallel. We are able to derive under suitable assumption the nested optimization problem

$$\inf_{(h, t) \in \mathbb{H} \times \mathbb{T}} A(h, t) = \inf_{h \in \mathbb{H}} S^{A, F}(h, \inf_{t \in \mathbb{T}} F(t)) , \quad (2.15)$$

for sequential decomposition and

$$\inf_{(h, t) \in \mathbb{H} \times \mathbb{T}} A(h, t) = S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}\left(\inf_{h \in \mathbb{H}} F_{\mathbb{H}}(h), \inf_{t \in \mathbb{T}} F_{\mathbb{T}}(t)\right) . \quad (2.16)$$

for parallel decomposition. We also describe how to apply these two schemes recursively.

### 2.3. Summary and contributions on risk averse competitive equilibrium

We have introduced the notion of player consistency and we have explained how this notion may help to decompose stochastic optimization problems sequentially and in a parallel manner. A common way to decompose an optimization problem in parallel is to reformulate it as an equilibrium problem. The idea is to dualize coupling constraints and then interpret multipliers as prices that coordinates different agents. We present, for the purpose of illustration, the risk neutral case. Then we detail our contributions.

**Risk neutral equilibrium and decomposition of a stochastic optimization problem** Let us consider a toy problem example. Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a probability space where  $\Omega$  is finite and  $\mathbb{P}$  charges all points. and let  $L^\infty(\Omega, \mathcal{T}, \mathbf{X})$  be the space of bounded random variables. Let  $\mathcal{X}_1$  and  $\mathcal{X}_2$  be two subsets of  $L^\infty(\Omega, \mathcal{T}, \mathbf{X})$ . We consider the following stochastic optimization problem

$$\min_{(\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{X}_1 \times \mathcal{X}_2} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] , \quad (2.17a)$$

$$\text{s.t. } g_1(\mathbf{X}_1) + g_2(\mathbf{X}_2) \geq \mathbf{D} , \mathbb{P}\text{-a.s.} \quad (2.17b)$$

where

- the functions  $c_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $c_2 : \mathbb{R} \rightarrow \mathbb{R}$  are costs functions,
- the functions  $g_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $g_2 : \mathbb{R} \rightarrow \mathbb{R}$  are supply functions,
- the random variable  $\mathbf{D}$  represents an aggregated demand.

We now turn on the statement of a risk neutral equilibrium. An equilibrium problem reads

$$\min_{(\mathbf{X}_1)} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1)] \quad (2.18)$$

$$\min_{(\mathbf{X}_2)} \mathbb{E}_{\mathbb{P}} [c_2(\mathbf{X}_2)] \quad (2.19)$$

$$0 \leq \boldsymbol{\lambda} \perp \mathbf{D} - (g_1(\mathbf{X}_1) + g_2(\mathbf{X}_2)) \geq 0 , \quad (2.20)$$

where  $\boldsymbol{\lambda}$  is a market price variable. The complementary slackness constraint (2.20) express the fact that prices should be chosen such that supply equals to demand.

The first welfare theorem indicates that Problems (2.17) and (2.18) are "equivalent", that is, the market "finds" a solution that is socially optimal. Thus, it provides a decomposition scheme for the optimization problem (2.17) into an equilibrium problem.

The idea behind the first welfare theorem is the following. Under mild assumptions, Problem (2.17) is equivalent to a reformulation of its dual that reads (we omit sets over which optimum are taken to focus on decomposition)

$$\max_{\boldsymbol{\lambda}} \min_{(\mathbf{X}_1, \mathbf{X}_2)} \mathbb{E}_{\mathbb{P}} [c_1(\mathbf{X}_1) + c_2(\mathbf{X}_2)] + \sum_{\omega \in \Omega} \boldsymbol{\lambda}(\mathbf{D}(\omega) - \omega)(g_1(\mathbf{X}_1)(\omega) + g_2(\mathbf{X}_2)(\omega)) . \quad (2.21)$$

The inner problem can now be solved player by player. This decomposition is known as price decomposition scheme. At optimum, we want to satisfy the complementarity slackness conditions (2.20). The connection with equilibrium problem follows naturally.

We now want to investigate what happens when we replace the mathematical expectation by a coherent risk measure.

## 2. Introduction (English)

### **Risk averse equilibrium and decomposition of a stochastic optimization problem**

We investigate risk averse equilibrium motivated by electricity markets. Indeed, most industrialized regions of the world have over the last thirty years have established wholesale electricity markets, that take the form of an auction that matches supply and demand. These are often called *two-settlement* markets.

If probability distributions for intermittent supply are known for these systems then it makes sense to maximize the expected total welfare of producers and consumers in each dispatch. Then many repetitions of this will yield a long run total benefit that is maximized. Maximizing expected welfare can be modeled as a two-stage stochastic program. Methods for computing prices and single-settlement payment mechanisms for such a *stochastic market clearing* mechanism are described in a number of papers (see [Pritchard et al. \(2010\)](#), [Wong and Fuller \(2007\)](#) and [Zakeri et al. \(2016\)](#)). When evaluated using the assumed probability distribution on supply, stochastic market clearing can be shown to be more efficient than two-settlement systems.

If agents in these systems are risk averse, then one might also seek to maximize some risk-adjusted social welfare. In this setting the computation of prices and payments to the agents becomes more complicated. If agents use coherent risk measures then it is possible to define a complete market for risk in a precise sense. If the market is complete then a perfectly competitive partial equilibrium will also maximize risk-adjusted social welfare, i.e. it is efficient. On the other hand if the market for risk is not complete, then perfectly competitive partial equilibrium can be inefficient. This has been explored in a number of papers (see e.g. [de Maere d’Aertrycke et al. \(2017\)](#), [Ehrenmann and Smeers \(2011\)](#) and [Ralph and Smeers \(2015\)](#)).

- In Chap. 6 we study a class of stochastic dispatch and pricing mechanisms under the assumption that agents will attempt to maximize their risk-adjusted welfare at these prices. Agents have coherent risk measures and are assumed to behave as price takers in the energy and risk markets. We aim at enlightening some difficulties that arise when risk markets are not complete. We describe a simple instance of a stochastic market that has three different equilibria. Two of these points are stable in the sense of [Samuelson \(1941\)](#) and are attractors of *tâtonnement algorithms*. The third equilibrium is unstable, yet is the solution yielded by the well-known PATH solver in GAMS (See [Ferris and Munson \(2000\)](#)). Our example illustrates the delicacy of seeking numerical solutions for equilibria in incomplete markets. Since these are used for justifying decisions, the nonuniqueness of solutions in this setting is undesirable. Indeed, we cannot control to which equilibrium the different algorithms are going to converge.
- In Chap. 7 we discuss links between player consistency and equilibrium problems. We suggest an interpretation of markets prices as tools to obtain consistency between players and social planner. In a sense, prices are not only efficient coordinator mechanisms, but also consistent coordinator mechanisms.



## 2.4. Summary and contribution on risk averse classification and estimation problems

In [Rockafellar and Uryasev \(2013\)](#) and [Rockafellar, Uryasev, and Zabrankin \(2008\)](#), the authors present a fundamental quadrangle that provides link between estimation and control problems. We follow their path to provide links between risk averse optimization and distributionally robust optimization. More precisely, we show that common distributionally robust optimization problems are risk averse problems with a well chosen risk measure that displays nice properties.

- In machine learning, the robustness of the solutions obtained for classification and prediction tasks remains a main issue. In [Papernot, McDaniel, and Goodfellow \(2016\)](#) and [Kurakin, Goodfellow, and Bengio \(2016\)](#), some examples are provided where small modifications of the input data can completely change the resulting solution.

This kind of problems also occurs in optimal control when there exist uncertainties on parameters. In [Ben-Tal and Nemirovski \(2000\)](#), the authors showed that a small perturbation on the parameters can turn a feasible solution into an infeasible one.

In this context, robust approaches appear as a way of controlling out-of-sample performance. There is an extensive literature dealing with robust problems and the reader is referred to [Ben-Tal, El Ghaoui, and Nemirovski \(2009\)](#) for a survey. One of the main approaches consists of introducing constraints on the probability distribution of the unknown data. Under some conditions, this approach is equivalent to deal with ambiguity sets or a modified loss function. Let us detail for clarity the formulation of a robust problem based on ambiguity set. Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a probability space and  $\mathcal{X}$  be a set of random variables on this space. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a function. The usual stochastic optimization problem reads

$$\inf_{\mathbf{X} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{X})], \quad (2.22)$$

and its robust counterpart based on ambiguity set reads

$$\inf_{\mathbf{X} \in \mathcal{X}} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{X})], \quad (2.23)$$

where  $\mathcal{Q}$  is a set of probability distributions on  $\Omega$ . The works in [Ben-Tal, Den Hertog, De Waegenaere, Melenberg, and Rennen \(2013\)](#); [Hu and Hong \(2013\)](#); [Duchi, Glynn, and Namkoong \(2016\)](#); [Moghaddam and Mahlooji \(2016\)](#) and [Namkoong and Duchi \(2016\)](#) have brought more insight on ambiguity sets. In [Esfahani and Kuhn \(2015\)](#) and [Esfahani, Shafieezadeh-Abadeh, Hanasusanto, and Kuhn \(2017\)](#), the authors present a distributionally robust optimization framework based on the Wasserstein distance. A set of probability distributions is defined as a ball centered on the reference probability with respect to the Wasserstein distance, then the optimization is carried out for the worst cost over this probability set.



## 2. Introduction (English)

This idea of minimizing the worst cost over a given probability set is well-known in quantitative finance. The robust representation of risk measures provides a theoretical framework to do so. A good class of risk measures is the class of coherent ones which were introduced in the seminal paper by [Artzner, Delbaen, Eber, and Heath \(1999\)](#). In [Föllmer and Schied \(2016\)](#), a broader class of so-called convex risk measures was investigated, for which a large number of results were established.

In this paper, we follow the line of [Esfahani and Kuhn \(2015\)](#), which aims at reformulating robust problems using ambiguity sets as convex minimization problems. Our contribution is threefold. First we clarify the links existing between risk measures and robust optimization. This allows us to transpose results from finance to machine learning. Second, we propose a unifying convex optimization setting for dealing with various risk measures, including those based on  $\varphi$ -divergences or the Wasserstein distance. Finally, we propose an accelerated algorithm grounded on the subgradient projection method proposed in [Combettes \(2003\)](#). We show that the proposed algorithm is able to solve efficiently large-scale robust problems.

## **Part II.**

# **Player consistency for ranking and optimization**



# Introduction

This chapter deals with time consistency, its extension to player consistency and its use in optimization. Time consistency is a formalization of the idea of sticking to one's plan. More precisely, if I decide today of an optimal policy for the future, and if I recompute in the future an optimal policy, then I should end up with the same policies than the one given by the original strategy.

In the risk neutral case, the notion of time consistency is related to the nested formula

$$\mathbb{E}_{\mathbb{P}}[\mathbf{X}] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[\mathbf{X} \mid \mathcal{F}]] , \quad (2.24)$$

where  $\mathcal{F}$  is a given  $\sigma$ -field. Such a nested formula allows dynamic programming and allows to decompose a stochastic problem sequentially (Bellman, 1957; Bertsekas, 1976, 1987).

We then discuss a generalization of time consistency that we call player consistency. Consider a finite number of decision makers, each of them endowed with their own preference on their own decision set. A leader supervises the team of players. He is endowed with his own preference over the product of players decision sets. Keeping in mind that an individual playing at two different time steps is represented by two distinct players, we extend the notion of time consistency to one of player consistency as follows: when the players, one by one, are indifferent between two decisions, so is the team leader with respect to the two collections of decisions. With this approach, we expect to deal with optimization problems with a general structure of information not necessarily processes and filtration, which can account for decentralized information among players.

We first consider in Chap. 3 the case of a unique individual playing at different time steps. This is the well known framework of Time Consistency. Behind the words “time consistency”, there is a disparate literature. We could trace the notion first in Kreps and Porteus (1978b). Other classical references include Epstein and Schneider (2003b), Artzner, Delbaen, Eber, Heath, and Ku (2007b), Ruszczyński (2010). Time consistency means that if a tail  $t'$  of a stochastic process is greater than a tail  $t$  of an other process, and if the heads of the processes are both equal to  $h$ , then the whole process  $(h, t')$  should be greater than the whole process  $(h, t)$ . When preferences are represented by numerical functions, we prove that time consistency is equivalent to a nested formula. We introduce three notions of time consistency: a weak, a usual and a strong one. We provide for each of these three notions two characterizations: one in terms of a set valued mapping and one in terms of level sets.

Then, in Chap. 4, we discuss applications to different class of mappings inspiring ourselves from results on risk measures (Föllmer and Schied, 2016). We first consider translation invariant mappings. For such mappings, the notion of acceptance set is crucial. We show that the time consistency property between two mappings is

equivalent to the inclusion of an acceptance set in a starshape operation between an acceptance set and a residual set. In certain case, this operation collapses to an inclusion between two acceptance sets. We then consider mappings that are defined as Fenchel-Moreau's transform. We illustrate that time consistency is equivalent to a form of rectangularity property in the dual space, taken together with an additive decomposition of the Fenchel transform of the risk mappings.

In Chap. 5, we unlink time and consistency and define  $\mathcal{R}$ -consistency for binary relations and player consistency. We then detail why player consistency is an adapted tool for decomposing optimization problems sequentially and in parallel.

To conclude this part, we illustrate in Chap. 5.3 how our framework can be applied to optimization problems. We provide two results: one specialized to Time Consistency and one, more general, for Player Consistency.

The following table 2.1 sums up the contributions of this part.

- Time consistency is equivalent to a nested formula that leads to sequential decomposition.
- Player consistency is equivalent to a nested formula that leads to parallel decomposition

Time consistency	Player consistency
$F(t) \leq F(t') \Rightarrow A(h, t) \leq A(h, t')$	$\left. \begin{array}{l} F_{\mathbb{H}}(h') \leq F_{\mathbb{H}}(h') \\ F_{\mathbb{T}}(t) \leq F_{\mathbb{T}}(t') \end{array} \right\} \Rightarrow A(h, t) \leq A(h', t')$
$\Downarrow$	$\Downarrow$
$A(h, t) = S^{A,F}(h, F(t))$	$A(h, t) = S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}}(F_{\mathbb{H}}(h), F_{\mathbb{T}}(t))$
$\Downarrow$	$\Downarrow$
Sequential optimization	Parallel optimization
$\Downarrow$	$\Downarrow$
$\bigwedge_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h, t) = \bigwedge_{h \in \mathbb{H}} S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t))$	$\bigwedge_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h, t) = S^{A,F}(\bigwedge_{h \in \mathbb{H}} F_{\mathbb{H}}(h), \bigwedge_{t \in \mathbb{T}} F_{\mathbb{T}}(t))$

Table 2.1.: summary of the contribution of Part II.

# 3. Time consistent mappings

This chapter provides the content of [Gérard, De Lara, and Chancelier \(2017\)](#) with some changes of notation. Section Sect. 3.5 is not part of the article but has been added since it provides complementary results that will be useful in the rest of the manuscript.

## Contents

---

<b>3.1. Introduction</b>	<b>46</b>
<b>3.2. Review of the literature</b>	<b>48</b>
3.2.1. Axiomatic for Time Consistency (TC)	48
3.2.2. Axiomatic for Nested Formulas (NF)	49
<b>3.3. Main result: equivalence between time consistency and nested formula</b>	<b>49</b>
3.3.1. Weak Time Consistency	50
3.3.2. Extensions to Usual and Strong Time Consistency	52
3.3.3. Analytical properties of time consistent mappings	54
<b>3.4. Revisiting the literature</b>	<b>58</b>
3.4.1. Axiomatic for Time Consistency (TC)	59
3.4.2. Axiomatic for Nested formulas (NF)	60
<b>3.5. Additional characterizations of time consistency</b>	<b>61</b>
3.5.1. Characterization of WTC in terms of partitions	61
3.5.2. Characterizations of UTC in terms of level sets	63
3.5.3. Characterizations of STC in terms of level sets	64
<b>3.6. Conclusion</b>	<b>65</b>

---

### 3. Time consistent mappings

You are a financial analyst. At the beginning of every week, you are able to rank every pair of stochastic processes starting from that week up to the horizon. Suppose that two processes are equal at the beginning of the week. Your ranking procedure is time consistent if the ranking does not change between this week and the next one. In this paper, we propose a minimalist definition of Time Consistency (TC) between two (assessment) mappings. With very few assumptions, we are able to prove an equivalence between Time Consistency and a Nested Formula (NF) between the two mappings. Thus, in a sense, two assessments are consistent if and only if one is factored into the other. We review the literature and observe that the various definitions of TC (or of NF) are special cases of ours, as they always include additional assumptions. By stripping off these additional assumptions, we present an overview of the literature where the contribution of each author is enlightened.

## 3.1. Introduction

Behind the words “Time Consistency” and “Nested Formula”, one can find a vast literature resorting to economics, dynamical risk measures and stochastic optimization.

Let us start with economics. In a dynamic bargaining problem, a group of agents has to agree on a common path of actions. As time goes on and information is progressively revealed, they can all reconsider the past agreement, and possibly make new assessments leading to new actions. Stability is the property that the agents will stick to their previous commitment. Time consistency is a form of stability when an individual makes a deal between his different selves (agents) along time. The notion of “consistent course of action” (see [Peleg and Yaari, 1973](#)) is well-known in the field of economics, with the seminal work of ([Strotz, 1955-1956](#)): an individual having planned his consumption trajectory is consistent if, reevaluating his plans later on, he does not deviate from the originally chosen plan. This idea of consistency as “sticking to one’s plan” may be extended to the uncertain case where plans are replaced by decision rules (“Do thus-and-thus if you find yourself in this portion of state space with this amount of time left”, Richard Bellman cited in ([Dreyfus, 2002](#))); ([Hammond, 1976](#)) addresses “consistency” and “coherent dynamic choice”, ([Kreps and Porteus, 1978a](#)) refers to “temporal consistency”. Another classical reference in economics is ([Epstein and Schneider, 2003a](#)).

Dynamic or Time Consistency has been introduced in the context of dynamical risk measures (see [Riedel, 2004](#); [Detlefsen and Scandolo, 2005](#); [Cheridito et al., 2006](#); [Artzner et al., 2007a](#), for definitions and properties of coherent and consistent dynamic risk measures).

In the field of stochastic optimization, Time Consistency has then been studied in the stochastic programming framework by ([Shapiro, 2009](#)) and for Markov Decision Processes by ([Ruszczyński, 2010](#)).

These different origins of Time Consistency contribute to a disparate literature. First, as Nested Formulas lead naturally to Time Consistency, some authors study the conditions to obtain Nested Formulas, whereas others focus on the axiomatics of Time Consistency and obtain Nested Formulas. Second, many definitions co-

habit. For instance, (Ruszczyński, 2010) add translation invariant property with additive criterion, (Shapiro, 2016; Artzner, Delbaen, Eber, Heath, and Ku, 2007a) add assumptions of coherent risk measures, and many authors focus on a particular structure of information (filtration). In this disconnected landscape, (De Lara and Leclère, 2016) tries to make the connection between “dynamic consistency” for optimal control problems (economics, stochastic optimization) and “time consistency” for dynamic risk measures. In this paper, we will focus on Time Consistency, motivated by dynamic risk measures — where the future assessment of a tail of a process is consistent with the initial assessment of the whole process, head and tail — but not limited to them. Below, we sketch our definitions of TC and NF. Our main contribution will be proving their equivalence.

**Axiomatic for Time Consistency.** We start presenting axiomatic of Time Consistency in a nutshell. Depending on the authors, the objects that are manipulated are either processes (Riedel, 2004; Detlefsen and Scandolo, 2005; Cheridito, Delbaen, and Kupper, 2006; Artzner, Delbaen, Eber, Heath, and Ku, 2007a) or lotteries (Kreps and Porteus, 1978a; Epstein and Schneider, 2003a). These objects are divided into two parts: a head  $h$  and a tail  $t$ . On the one hand, we have a way to assess any tail  $t$  by means of a mapping  $F$  (factor), yielding  $F(t)$ . On the other hand, we have a way to assess any couple head-tail  $(h, t)$  by means of a mapping  $A$  (aggregator), yielding  $A(h, t)$ .

We look for a consistency property between these two ranking mapping  $F$  and  $A$ : if a tail  $t$  is equivalent to a tail  $t'$ , then the two elements  $(h, t)$  and  $(h, t')$  — that share the same head — must be such that  $(h, t)$  is equivalent to  $(h, t')$ . This can be written mathematically as

$$F(t) = F(t') \Rightarrow A(h, t) = A(h, t') , \quad \forall (h, t, t') \in \mathbb{H} \times \mathbb{T}^2 . \quad (\text{TC})$$

**Axiomatic for Nested Formulas.** Some authors focus on sufficient conditions to obtain a Nested Formula (Shapiro, 2016; Ruszczyński and Shapiro, 2006a). In a Nested Formula, the assessment  $F(t)$  of any tail  $t$  is factored inside the assessment  $A(h, t)$  of any head-tail  $(h, t)$  by means of a surrogate mapping  $S^{A,F}$  as follows:

$$A(h, t) = S^{A,F}(h, F(t)) . \quad (\text{NF})$$

Of course, (NF) implies (TC). We will prove the reverse: (TC) implies that there exists a mapping  $S^{A,F}$  such that (NF) holds true.

In Sect. 3.2, we go through the literature, with the goal of extracting the following components: what kind of objects are treated, what are the heads and the tails, how are these objects ranked. In Sect. 3.3, we formally state our definitions of Time Consistency (TC) and Nested Formula (NF), and we prove their equivalence. We also provide conditions to obtain analytical properties of the mapping  $S^{A,F}$  appearing in the Nested Formula, such as monotonicity, continuity, convexity, positive homogeneity and translation invariance. In Sect. 3.4, we show that our framework covers the different frameworks reviewed in Sect. 3.2.



### 3. Time consistent mappings

## 3.2. Review of the literature

We have screened a selection of papers, in mathematics and economics, touching Time Consistency and Nested Formula in various settings. Depending on the setting, we identify the following components, as introduced in Sect. 3.1: what kind of objects are treated, what are the heads and the tails, how are these objects ranked. Table 3.1 sums up our survey.

	Article	Objects	Head	Tail	Assessment
Time Consistency	Kreps and Porteus	Lottery	Lottery from 1 to $s$	Lottery from $s + 1$ to $T$	Expected utility
	Epstein and Schneider	Lottery	Lottery from 1 to $s$	Lottery from $s + 1$ to $T$	Not necessarily expected utility
	Ruszczyński	Process	Process from 1 to $s$	Process from $s + 1$ to $T$	Dynamic risk measure
	Artzner et al.	Process	Process from 1 to $\tau$	Process from $\tau$ to $T$ , $\tau$ stopping time	Coherent risk measure
Nested Formula	Shapiro	Process	Process from 1 to $s$	Process from $s + 1$ to $T$	Coherent risk measure
	Ruszczyński and Shapiro	Process	Process from 1 to $s$	Process from $s + 1$ to $T$	Coherent risk measure
	De Lara and Leclère	Process	Process from 1 to $s$	Process from $s + 1$ to $T$	Dynamic risk measure

Table 3.1.: Sketch of papers selected on Time Consistency and Nested Formulas

### 3.2.1. Axiomatic for Time Consistency (TC)

The first group of authors is subdivided between economists, who deal with lotteries and preferences, and probabilists who deal with stochastic processes and dynamical risk measures.

#### Lotteries and preferences

In (Kreps and Porteus, 1978a), (Kreps and Porteus, 1979) and (Epstein and Schneider, 2003a), the authors deal with lotteries and preferences. A preference is a total, transitive and reflexive relation. Proper assumptions make it possible that the preference relation can be represented by a numerical evaluation. Assumptions of monotonicity and convexity are also made.

In (Kreps and Porteus, 1978a), Kreps and Porteus propose axioms that make that the preference is represented by an expected utility formula.

By contrast, more general numerical representations are studied in (Epstein and Schneider, 2003a), even if Epstein and Schneider add an hypothesis of additive criterion. A summary of the assumptions can be found in Table 3.2.

#### Dynamic risk measures and processes

In (Ruszczyński, 2010) and (Artzner, Delbaen, Eber, Heath, and Ku, 2007a), the authors deal with stochastic processes assessed by dynamical risk measures.

### 3.3. Main result: equivalence between time consistency and nested formula

In (Ruszczyński, 2010), Ruszczyński studies a family of conditional risk measures which are monotonic, invariant by translation and homogeneous. The criterion is additive.

In (Artzner, Delbaen, Eber, Heath, and Ku, 2007a), Artzner, Delbaen, Eber, Heath and Ku focus on the value of the stochastic process at the final time step. They use as assessment a particular class of risk measures, the so-called coherent risk measures.

#### 3.2.2. Axiomatic for Nested Formulas (NF)

In (Shapiro, 2016), (Ruszczyński and Shapiro, 2006a) and (De Lara and Leclère, 2016), the focus is on exhibiting sufficient conditions to obtain Nested Formulas. All authors study stochastic processes, with an assumption of monotonicity for the assessment, but there are some differences.

In (Ruszczyński and Shapiro, 2006a), Ruszczyński and Shapiro study coherent risk measures in their dual form (hence with properties of convexity, invariance by translation and additive criterion).

In (Shapiro, 2016), Shapiro focuses on assessing the value of the process at the final step with coherent risk measures.

In (De Lara and Leclère, 2016), De Lara and Leclère study how commutation properties between time aggregators and uncertainty aggregators make it possible to obtain Nested Formulas.

	Article	Monotonicity	Translation invariance	Convexity
Time Consistency	(Kreps and Porteus, 1978a)	Yes	No	Yes
	(Kreps and Porteus, 1979)	Yes	No	Yes
	(Epstein and Schneider, 2003a)	Yes	No	Yes
	(Ruszczyński, 2010)	Yes	Yes	No
	(Artzner, Delbaen, Eber, Heath, and Ku, 2007a)	Yes	Yes	Yes
Nested Formula	(Shapiro, 2016)	Yes	Yes	Yes
	(Ruszczyński and Shapiro, 2006a)	Yes	Yes	Yes
	(De Lara and Leclère, 2016)	Yes	No	No

Table 3.2.: Most common assumptions in the selection of papers on Time Consistency and Nested Formula

### 3.3. Main result: equivalence between time consistency and nested formula

In Sect. 3.1, we have sketched the notions of Time Consistency and Nested Formula. Now, in §3.3.1, we properly define Weak Time Consistency — with minimal assumptions — and we prove that it is equivalent to a Nested Formula. In §3.3.2, we extend definitions and results to Usual and Strong Time Consistency: by adding

### 3. Time consistent mappings

order structures, we obtain additional properties. In §3.3.3, we provide conditions to obtain analytical properties of the mapping appearing in the Nested Formula, such as monotonicity, continuity, convexity, positive homogeneity and translation invariance. Let us introduce basic notations.

Let  $\mathbb{H}$  and  $\mathbb{T}$  be two sets, respectively called *head set* and *tail set*. Let  $\mathbb{A}$ ,  $\mathbb{F}$  be two sets and let  $A$  and  $F$  be two mappings as follows:

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A} , \quad F : \mathbb{T} \rightarrow \mathbb{F} . \quad (3.1)$$

The mapping  $A$  is called an *aggregator*, as it aggregates head-tail in  $\mathbb{H} \times \mathbb{T}$  into an element of  $\mathbb{A}$ . The mapping  $F$  is called a *factor* because of the Nested Formula (NF) in Sect. 3.1.

**Definition 3.3.1.** With the couple aggregator-factor  $(A, F)$  in (3.1) we associate the set-valued mapping

$$\begin{aligned} S^{A,F} : \mathbb{H} \times \text{Im}(F) &\rightrightarrows \mathbb{A} \\ (h, f) &\mapsto S^{A,F}(h, f) = \{A(h, t) \mid t \in F^{-1}(f)\} , \end{aligned} \quad (3.2)$$

where  $\text{Im}(F) = F(\mathbb{T})$ . We call  $S^{A,F}$  the *subaggregator* of the couple  $(A, F)$ .

#### 3.3.1. Weak Time Consistency

**Definition 3.3.2** (Weak Time Consistency). The couple aggregator-factor  $(A, F)$  in (3.1) is said to satisfy *Weak Time Consistency (WTC)* if we have

$$F(t) = F(t') \Rightarrow A(h, t) = A(h, t') , \quad \forall h \in \mathbb{H} , \quad \forall (t, t') \in \mathbb{T}^2 . \quad (3.3)$$

Here is our main result where we characterize the WTC property in terms of the subaggregator in (3.2).

**Theorem 3.3.3** (Nested decomposition of WTC mappings). *The couple aggregator-factor  $(A, F)$  in (3.1) is WTC if and only if the subaggregator set valued mapping  $S^{A,F}$  in (3.2) is a mapping. In that case, the following Nested Formula between mappings holds true:*

$$A(h, t) = S^{A,F}(h, F(t)) , \quad \forall h \in \mathbb{H} , \quad \forall t \in \mathbb{T} . \quad (3.4)$$

*Proof.* Note that we always have by Equation (3.2) that

$$A(h, t) \in S^{A,F}(h, F(t)) . \quad (3.5)$$

1. We suppose that the couple  $(A, F)$  is Weak Time Consistent. Consider  $(h, f)$  fixed in  $\mathbb{H} \times \text{Im}(F)$ . We are going to show that the set valued mapping  $S^{A,F}$  is in fact a mapping, by proving that the set  $S^{A,F}(h, f)$ , defined in (3.2), is reduced to a singleton.

We consider two elements  $a = A(h, t)$  and  $a' = A(h, t')$  in the set  $S^{A,F}(h, f)$ . By definition (3.2), we have  $F(t) = F(t') = f$ . Then, using the Weak Time Consistency property (3.3), we deduce  $A(h, t) = A(h, t')$ . Thus,  $S^{A,F}(h, f)$  is reduced to one value for  $f \in \text{Im}(F)$ . The set valued mapping  $S^{A,F}$  is thus a mapping and, using Equation (3.5), we obtain  $A(h, t) = S^{A,F}(h, F(t))$ .

### 3.3. Main result: equivalence between time consistency and nested formula

2. We suppose now that the set valued mapping  $S^{A,F}$ , defined in (3.2), is a mapping. Since  $S^{A,F}$  is a mapping, we deduce by Equation (3.5) that  $A(h, t) = S^{A,F}(h, F(t))$  for all  $t \in \mathbb{T}$ . Therefore, we have the implications:  $F(t) = F(t') \Rightarrow S^{A,F}(h, F(t)) = S^{A,F}(h, F(t')) \Rightarrow A(h, t) = A(h, t')$ . We conclude that the weak time consistency property (3.3) is satisfied.

In both cases, we have shown that Equation (3.4) holds true.  $\square$

**Example 3.3.4** (The couple  $(\text{AV@R}_\beta[\cdot + \cdot], \text{AV@R}_\beta[\cdot | \mathcal{F}])$  is not Weak Time Consistent). We now give an example inspired from (Pflug and Pichler, 2014, Sect. 5.3.2, p. 188) and involving the well known Average Value at Risk. It helps to illustrate our main result and the notions we have introduced so far.

Let  $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ , that we equip with the uniform probability distribution  $\mathbb{P} = \frac{1}{4}\delta_{\omega_1} + \frac{1}{4}\delta_{\omega_2} + \frac{1}{4}\delta_{\omega_3} + \frac{1}{4}\delta_{\omega_4}$ .

We introduce the sets  $\mathbb{H} = \mathbb{T} = \mathbb{R}^{|\Omega|} = \mathbb{R}^4$ . On this finite space  $\Omega$ , the Average Value at Risk of level  $\beta$  ( $0 \leq \beta \leq 1$ ) of a random variable  $\mathbf{X} : \Omega \rightarrow \mathbb{R}$  is defined by (Rockafellar and Uryasev, 2000a)

$$\text{AV@R}_\beta(\mathbf{X}) = \min_{\alpha \in \mathbb{R}} \left\{ \alpha + \frac{1}{1-\beta} \mathbb{E}_{\mathbb{P}} [[\mathbf{X} - \alpha]^+] \right\}. \quad (3.6)$$

Let  $\mathcal{F} = \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\}$  be a  $\sigma$ -field on the space  $\Omega$ . The Conditional Average Value at Risk of level  $\beta$ , of a random variable  $\mathbf{X} : \Omega \rightarrow \mathbb{R}$  with respect to the  $\sigma$ -field  $\mathcal{F}$  is defined by (Pflug and Pichler, 2014)

$$\text{AV@R}_\beta(\mathbf{X} | \mathcal{F}) = \inf \{ \mathbb{E}_{\mathbb{P}}[\mathbf{X}Z | \mathcal{F}] : \mathbb{E}_{\mathbb{P}}[Z | \mathcal{F}] = \mathbf{1}, 0 \leq Z, \beta Z \leq 1 \} \quad (3.7)$$

We define two mappings

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{R} \quad F : \mathbb{T} \rightarrow \mathbb{R}^2 \quad (3.8a)$$

$$(h, t) \mapsto \text{AV@R}_{0.5}[h + t], \quad t \mapsto \text{AV@R}_{0.5}[t | \mathcal{F}]. \quad (3.8b)$$

Consider four elements: a head  $h_0 = (0, 0, 0, 0) \in \mathbb{H}$ , a first tail  $t_0 = (3, 3, 2, 1) \in \mathbb{T}$ , a second tail  $t'_0 = (1, 3, 2, 2) \in \mathbb{T}$  and an element of the factor's image  $f_0 = (3, 2) \in \mathbb{F}$ . On the one hand, the elements  $F(t_0)$  and  $F(t'_0)$  are equal, because

$$\text{AV@R}_{0.5}[t_0 | \mathcal{F}] = \underbrace{(3; 2)}_{f_0} = \text{AV@R}_{0.5}[t'_0 | \mathcal{F}]. \quad (3.9)$$

On the other hand, the elements  $A(h_0, t_0)$  and  $A(h_0, t'_0)$  are not equal, because

$$3 = \text{AV@R}_{0.5}[h_0 + t_0] \neq \text{AV@R}_{0.5}[h_0 + t'_0] = 2.5. \quad (3.10)$$

The subaggregator  $S^{A,F}$  in (3.2) is not a mapping since

$$S^{A,F}(h_0, f_0) = \{ \text{AV@R}_{0.5}[h_0 + t] | \text{AV@R}_{0.5}[t | \mathcal{F}] = f_0 \} \supset \{2.5; 3\}, \quad (3.11)$$

and therefore the couple  $(A, F)$  in (3.8) is not Weak Time Consistent.

### 3. Time consistent mappings

#### 3.3.2. Extensions to Usual and Strong Time Consistency

With additional order structures on the image sets  $\mathbb{A}$  and  $\mathbb{F}$  of the aggregator  $A$  and of the factor  $F$ , and possibly on the head set  $\mathbb{H}$  — all presented in (3.1) — we define two additional notions of Time Consistency, usual and strong.

##### Usual Time Consistency (UTC)

Suppose that the image sets  $\mathbb{A}$  and  $\mathbb{F}$  are equipped with orders, denoted by  $\leq$ .

**Definition 3.3.5** (Definition of Usual Time Consistency). The couple aggregator-factor  $(A, F)$  in (3.1) is said to satisfy *Usual Time Consistency (UTC)* if we have

$$F(t) \leq F(t') \Rightarrow A(h, t) \leq A(h, t'), \quad \forall h \in \mathbb{H}, \quad \forall (t, t') \in \mathbb{T}^2. \quad (3.12)$$

We extend the result of Theorem 3.3.3 as follows.

**Proposition 3.3.6** (Nested decomposition of UTC mappings). *The couple  $(A, F)$  in (3.1) is UTC if and only if the set valued mapping  $S^{A,F}$  in (3.2) is a mapping and is increasing<sup>1</sup> in its second argument. In that case, the Nested Formula (3.4) holds true.*

*Proof.*

1. We suppose that the couple  $(A, F)$  is Usual Time Consistent. In particular, it is Weak Time Consistent and Theorem 3.3.3 gives us that the set-valued mapping  $S^{A,F}$  in (3.2) is a mapping and that  $A(h, t) = S^{A,F}(h, F(t))$ , for all  $h \in \mathbb{H}$  and  $t \in \mathbb{T}$ .

There remains to show that the subaggregator  $S^{A,F}$  in (3.2) is increasing in its second argument. Let  $f = F(t)$  and  $f' = F(t')$  be two fixed elements in the set  $\text{Im}(F)$ , such that  $f \leq f'$ , hence  $F(t) \leq F(t')$ . By the Usual Time Consistency definition (3.12), we deduce that  $A(h, t) \leq A(h, t')$ , for all  $h \in \mathbb{H}$ . This leads to  $S^{A,F}(h, F(t)) \leq S^{A,F}(h, F(t'))$ , by the nested formula (3.4). By identification, we obtain that  $S^{A,F}(h, f) \leq S^{A,F}(h, f')$ , for all  $h \in \mathbb{H}$ . Therefore, we have proved that the subaggregator  $S^{A,F}$  is increasing in its second argument.

2. We suppose now that the set valued mapping  $S^{A,F}$  is a mapping, increasing in its second argument. Following the proof of Theorem 3.3.3, we know that, as  $S^{A,F}$  is a mapping, we have  $A(h, t) = S^{A,F}(h, F(t))$ , for all  $(h, t) \in \mathbb{H} \times \mathbb{T}$ . We consider  $f = F(t)$  and  $f' = F(t')$  in  $\text{Im}(F)$  such that  $f \leq f'$ . As the mapping  $S^{A,F}$  is increasing in its second argument, we obtain that  $S^{A,F}(h, F(t)) \leq S^{A,F}(h, F(t'))$ . This leads to  $A(h, t) \leq A(h, t')$ , for all  $h \in \mathbb{H}$ , and we recover the Usual Time Consistency property (3.12).

In both cases, we have shown that Equation (3.4) holds true.  $\square$

---

<sup>1</sup>Let  $\mathbb{X}$  and  $\mathbb{Y}$  be sets endowed with orders denoted by  $\leq$ . A mapping  $M : \mathbb{X} \rightarrow \mathbb{Y}$  is said to be *increasing* if  $x \leq x' \Rightarrow M(x) \leq M(x')$ .

### Strong Time Consistency (STC)

Suppose that the head set  $\mathbb{H}$  and the image sets  $\mathbb{A}$  and  $\mathbb{F}$  are equipped with orders, denoted by  $\leq$ .

**Definition 3.3.7** (Definition of Strong Time Consistency). The couple  $(A, F)$  in Equation (3.1) is said to satisfy *Strong Time Consistency* (STC) if we have

$$\left. \begin{array}{l} F(t) \leq F(t') \\ h \leq h' \end{array} \right\} \Rightarrow A(h, t) \leq A(h', t'), \quad \forall (h, h', t, t') \in \mathbb{H}^2 \times \mathbb{T}^2. \quad (3.13)$$

We extend the results of Theorem 3.3.3 as follows.

**Proposition 3.3.8** (Nested decomposition for STC mappings). *The couple  $(A, F)$  in (3.1) is STC if and only if the set valued mapping  $S^{A,F}$  is a mapping increasing in its first and second arguments. In that case, the Nested Formula (3.4) holds true.*

*Proof.*

1. We suppose that the couple  $(A, F)$  is Strong Time Consistent. In particular, it is Usual Time Consistent and Proposition 3.3.6 gives us that the set-valued mapping  $S^{A,F}$  is a mapping, increasing in its second argument, and that  $A(h, t) = S^{A,F}(h, F(t))$ , for all  $h \in \mathbb{H}$  and for all  $t \in \mathbb{T}$ .

There remains to show that the subaggregator  $S^{A,F}$  is increasing in its first argument. We consider  $f = F(t)$  in the set  $\text{Im}(F)$  and two elements  $h$  and  $h'$  of  $\mathbb{H}$ , such that  $h \leq h'$ . By the Strong Time Consistency property (3.13), we have that  $A(h, t) \leq A(h', t)$ . This leads to  $S^{A,F}(h, F(t)) \leq S^{A,F}(h', F(t))$  by (3.4). As  $f = F(t)$ , we obtain that  $S^{A,F}(h, f) \leq S^{A,F}(h', f)$ . We conclude that the subaggregator  $S^{A,F}$  is increasing in its first argument.

2. We suppose now that set-valued mapping  $S^{A,F}$  in (3.2) is a mapping which is increasing in its first and in its second argument. Following the proof of Proposition 3.3.6, we know that (3.4) holds true for all  $(h, t) \in \mathbb{H} \times \mathbb{T}$ . We also have that the couple  $(A, F)$  is Usual Time Consistent. We consider  $f = F(t)$  and  $f' = F(t')$  in  $\text{Im}(F)$  such that  $f \leq f'$ . We also consider  $(h, h') \in \mathbb{H}^2$  such that  $h \leq h'$ . We have

$$F(t) \leq F(t') \Rightarrow A(h, t) \leq A(h, t'), \quad (3.14a)$$

$$\text{by Usual Time Consistency property (3.12),} \quad (3.14b)$$

$$\Rightarrow A(h, t) \leq S^{A,F}(h, F(t')), \quad (3.14c)$$

$$\text{by the Nested Formula (3.4),} \quad (3.14d)$$

$$\Rightarrow A(h, t) \leq S^{A,F}(h', F(t')), \quad (3.14e)$$

$$\text{by monotonicity in the first argument of } S^{A,F}, \quad (3.14f)$$

$$\Rightarrow A(h, t) \leq A(h', t'), \quad (3.14g)$$

$$\text{by the Nested Formula (3.4).} \quad (3.14h)$$

The reasoning is true for any couple  $(h, h') \in \mathbb{H}^2$  such that  $h \leq h'$ . We conclude that the couple  $(A, F)$  satisfies Strong Time Consistency.

In both cases, we have shown that Equation (3.4) holds true.  $\square$

### 3. Time consistent mappings

#### Summing up results about WTC, UTC and STC

In §3.3.1 and §3.3.2, we have introduced three notions of Time Consistency, from the weakest to the strongest. Of course, we have that a Strong Time Consistent couple is also Usual Time Consistent, and that a Usual Time Consistent couple is also Weak Time Consistent. We sum up the different definitions and results in Table 3.3.

	Weak (3.3)	⇐	Usual (3.12)	⇐	Strong (3.13)
Definition	$F(t) = F(t')$ ↓ $A(h, t) = A(h, t')$		$F(t) \leq F(t')$ ↓ $A(h, t) \leq A(h, t')$		$h \leq h'$ , $F(t) \leq F(t')$ ↓ $A(h, t) \leq A(h', t')$
Characterization in terms of subaggregator	$S^{A,F}$ is a mapping		$S^{A,F}$ is a mapping increasing in its second argument		$S^{A,F}$ is a mapping increasing in both arguments

Table 3.3.: Characterization of Time Consistency in terms of subaggregator

#### 3.3.3. Analytical properties of time consistent mappings

Here, we study properties inherited by the subaggregator  $S^{A,F}$  in (3.2) when it is a mapping, that is, when the couple  $(A, F)$  is Weak Time Consistent (see Theorem 3.3.3).

We focus on monotonicity, continuity, convexity, positive homogeneity and translation invariance.

##### Monotony

We suppose that the head set  $\mathbb{H}$ , the tail set  $\mathbb{T}$ , and the image sets  $\mathbb{A}$  and  $\mathbb{F}$  — all presented in (3.1) — are equipped with orders, denoted by  $\leq$ . The proof of the following proposition is left to the reader as a direct application of the Nested Formula (3.4).

**Proposition 3.3.9** (Monotony). *Let the couple  $(A, F)$  be Weak Time Consistent, as in Definition 3.3.2. If the mapping  $A$  is increasing in its first argument, then the subaggregator  $S^{A,F}$  in (3.2) is increasing in its first argument.*

##### Continuity

We suppose that the head set  $\mathbb{H}$ , the tail set  $\mathbb{T}$ , and the image sets  $\mathbb{A}$  and  $\mathbb{F}$  are metric spaces.

**Proposition 3.3.10** (Continuity). *Let the couple  $(A, F)$  be Weak Time Consistent, as in Definition 3.3.2. Assume that the tail set  $\mathbb{T}$  is compact. If the factor  $F$  is continuous and if the aggregator  $A$  is continuous with a compact image  $\text{Im}(A) = A(\mathbb{H} \times \mathbb{T})$ , then the subaggregator  $S^{A,F}$  in (3.2) is continuous on  $\mathbb{H} \times \text{Im}(F)$ .*



### 3.3. Main result: equivalence between time consistency and nested formula

*Proof.* We prove the continuity of the subaggregator  $S^{A,F}$  on  $\mathbb{H} \times \text{Im}(F)$  by using the sequential characterization of the continuity on metric spaces. For this purpose, we consider, on the one hand,  $(\bar{h}, \bar{f})$  element of  $\mathbb{H} \times \text{Im}(F)$  and, on the other hand,  $(h_n)_{n \in \mathbb{N}}$  a sequence of elements of  $\mathbb{H}$  converging to  $\bar{h}$  and  $(f_n)_{n \in \mathbb{N}}$  a sequence of elements of  $\text{Im}(F)$  converging to  $\bar{f}$ . We will show that  $S^{A,F}(h_n, f_n)$  converges to  $S^{A,F}(\bar{h}, \bar{f})$ . We introduce the notation  $\mathcal{L}(\{u_n\})$  to denote the set of limit points of a sequence  $(u_n)_{n \in \mathbb{N}}$ .

As  $f_n \in \text{Im}(F)$ , there exists an element  $t_n \in \mathbb{T}$  such that  $F(t_n) = f_n$  for each  $n$ . By the Nested Formula (3.4), we deduce that

$$A(h_n, t_n) = S^{A,F}(h_n, F(t_n)) = S^{A,F}(h_n, f_n). \quad (3.15)$$

We will now show that the set  $\mathcal{L}(\{A(h_n, t_n)\})$  of limit points is reduced to the singleton  $\{S^{A,F}(\bar{h}, \bar{f})\}$ . The proof is in several steps as follows:

1.  $\mathcal{L}(\{A(h_n, t_n)\}) \neq \emptyset$ ,
2.  $\mathcal{L}(\{A(h_n, t_n)\}) \subset A(\bar{h}, \mathcal{L}(\{t_n\}))$ ,
3.  $A(\bar{h}, \mathcal{L}(\{t_n\}))$  is reduced to the singleton  $\{S^{A,F}(\bar{h}, \bar{f})\}$ ,
4.  $\mathcal{L}(\{A(h_n, t_n)\}) = \{S^{A,F}(\bar{h}, \bar{f})\}$ .

Here is the proof.

1. As the sequence  $(A(h_n, t_n))_{n \in \mathbb{N}}$  takes value in the compact set  $\text{Im}(A)$ , we have that  $\mathcal{L}(\{A(h_n, t_n)\}) \neq \emptyset$ .
2. We prove that  $\mathcal{L}(\{A(h_n, t_n)\}) \subset A(\bar{h}, \mathcal{L}(\{t_n\}))$ . Let  $a$  be an element of the set  $\mathcal{L}(\{A(h_n, t_n)\})$ . By definition of this latter set, there exists a subsequence  $(A(h_{\Phi(n)}, t_{\Phi(n)}))_{n \in \mathbb{N}}$  converging to  $a$ . Now, we know that  $(h_{\Phi(n)})_{n \in \mathbb{N}}$  converges to  $\bar{h}$ , but it is not necessarily the case that  $(t_{\Phi(n)})_{n \in \mathbb{N}}$  converges. However, by compactness of the tail set  $\mathbb{T}$ , there exist a subsequence  $(t_{\Psi \circ \Phi(n)})_{n \in \mathbb{N}}$  converging to a certain  $\bar{t} \in \mathcal{L}(\{t_n\})$ . As the sequence  $(A(h_{\Phi(n)}, t_{\Phi(n)}))_{n \in \mathbb{N}}$  is converging to  $a$ , the subsequence  $(A(h_{\Psi \circ \Phi(n)}, t_{\Psi \circ \Phi(n)}))_{n \in \mathbb{N}}$  is also converging to  $a$ . Now that both inner subsequences converge, we use the continuity of the mapping  $A$ , and obtain that  $a = \lim_{n \rightarrow \infty} A(h_{\Psi \circ \Phi(n)}, t_{\Psi \circ \Phi(n)}) = A(\bar{h}, \bar{t}) \in A(\bar{h}, \mathcal{L}(\{t_n\}))$ .
3. We prove the equality  $A(\bar{h}, \mathcal{L}(\{t_n\})) = \{S^{A,F}(\bar{h}, \bar{f})\}$ . Since the set  $\mathcal{L}(\{t_n\})$  is not empty by compactness of  $\mathbb{T}$ , we consider  $(\bar{t}, \bar{t}') \in \mathcal{L}(\{t_n\})^2$  any two limits points of the sequence  $(t_n)_{n \in \mathbb{N}}$ . As  $F(t_n) = f_n$  and  $\lim_{n \rightarrow \infty} f_n = \bar{f}$ , we deduce that  $F(\bar{t}) = \bar{f} = F(\bar{t}')$ , by continuity of the factor mapping  $F$ . The Nested Formula (3.4) gives

$$A(\bar{h}, \bar{t}) = S^{A,F}(\bar{h}, F(\bar{t})) = S^{A,F}(\bar{h}, \bar{f}) = S^{A,F}(\bar{h}, F(\bar{t}')) = A(\bar{h}, \bar{t}').$$

This proves that  $A(\bar{h}, \mathcal{L}(\{t_n\})) = \{S^{A,F}(\bar{h}, \bar{f})\}$ .



### 3. Time consistent mappings

4. Gathering up the previous results, we obtain that

$$\emptyset \neq \mathcal{L}(\{A(h_n, t_n)\}) \subset A(\bar{h}, \mathcal{L}(\{t_n\})) = \{S^{A,F}(\bar{h}, \bar{f})\}. \quad (3.16)$$

We conclude that  $\mathcal{L}(\{A(h_n, t_n)\}) = \{S^{A,F}(\bar{h}, \bar{f})\}$ .

From Equation (3.15), we have the equalities  $\mathcal{L}(\{S^{A,F}(h_n, f_n)\}) = \mathcal{L}(\{A(h_n, t_n)\}) = \{S^{A,F}(\bar{h}, \bar{f})\}$ . Therefore, the sequence  $S^{A,F}(h_n, f_n)$  converges to  $S^{A,F}(\bar{h}, \bar{f})$ . This ends the proof.  $\square$

### Convexity

As we are dealing with convexity property, we assume that the sets  $\mathbb{H}, \mathbb{T}$  and  $\mathbb{F}$  in (3.1) are vector spaces. We also suppose that the aggregator  $A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$  in (3.1) takes extended real values, that is,  $\mathbb{A} = \mathbb{R} \cup \{-\infty, +\infty\}$ .

**Proposition 3.3.11.** *Let the couple  $(A, F)$  be Weak Time Consistent, as in Definition 3.3.2. If there exists a nonempty convex subset  $\bar{\mathbb{T}} \subset \mathbb{T}$  such that  $F(\bar{\mathbb{T}}) = \text{Im}(F)$  and that the restricted function  $F|_{\bar{\mathbb{T}}}$  is affine, and if the aggregator  $A$  is jointly convex, then the subaggregator  $S^{A,F}$  in (3.2) is jointly convex on  $\mathbb{H} \times \text{Im}(F)$ .*

*Proof.* We introduce the notation  $\text{epi}(M)$  to denote the epigraph<sup>2</sup> of a mapping  $M$ . We prove that the subaggregator  $S^{A,F}$  is jointly convex by showing that its epigraph is jointly convex.

Let  $((h_1, f_1), a_1)$  and  $((h_2, f_2), a_2)$  be two elements of the epigraph  $\text{epi}(S^{A,F})$  of the subaggregator. We consequently have  $a_1 \geq S^{A,F}(h_1, f_1)$  and  $a_2 \geq S^{A,F}(h_2, f_2)$  which by addition to

$$\lambda a_1 + (1 - \lambda)a_2 \geq \lambda S^{A,F}(h_1, f_1) + (1 - \lambda)S^{A,F}(h_2, f_2), \quad (3.17)$$

where  $\lambda$  is an element of  $[0, 1]$ . As, by assumption,  $F(\bar{\mathbb{T}}) = \text{Im}(F)$ , there exist two elements  $(\bar{t}_1, \bar{t}_2) \in \bar{\mathbb{T}}^2$  such that

$$F(\bar{t}_1) = f_1 \text{ and } F(\bar{t}_2) = f_2. \quad (3.18)$$

We have the succession of equalities and inequality

$$\lambda a_1 + (1 - \lambda)a_2 \geq \lambda S^{A,F}(h_1, f_1) + (1 - \lambda)S^{A,F}(h_2, f_2),$$

by Equation (3.17),

$$= \lambda S^{A,F}(h_1, F(\bar{t}_1)) + (1 - \lambda)S^{A,F}(h_2, F(\bar{t}_2)),$$

by Equation (3.18),

$$= \lambda A(h_1, \bar{t}_1) + (1 - \lambda)A(h_2, \bar{t}_2),$$

---

<sup>2</sup>Let  $\mathbb{X}$  be a set. The epigraph of the mapping  $M : \mathbb{X} \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$  is defined by  $\text{epi}(M) = \{(x, y) \in \mathbb{X} \times \mathbb{R} : M(x) \leq y\}$  where  $y$  is real valued.

### 3.3. Main result: equivalence between time consistency and nested formula

by the Nested Formula (3.4),

$$\geq A(\lambda h_1 + (1 - \lambda)h_2, \lambda \bar{t}_1 + (1 - \lambda)\bar{t}_2) ,$$

by convexity of  $A$ ,

$$= S^{A,F} \left( \lambda h_1 + (1 - \lambda)h_2, F(\lambda \bar{t}_1 + (1 - \lambda)\bar{t}_2) \right) ,$$

by the Nested Formula (3.4),

$$= S^{A,F}(\lambda h_1 + (1 - \lambda)h_2, \lambda F(\bar{t}_1) + (1 - \lambda)F(\bar{t}_2)) ,$$

by affinity of  $F$  on  $\bar{\mathbb{T}}$ ,

$$= S^{A,F}(\lambda h_1 + (1 - \lambda)h_2, \lambda f_1 + (1 - \lambda)f_2) ,$$

by Equation (3.18).

We deduce that the element  $\left( (\lambda h_1 + (1 - \lambda)h_2, \lambda f_1 + (1 - \lambda)f_2), \lambda a_1 + (1 - \lambda)a_2 \right)$  is in the epigraph  $\text{epi}(S^{A,F})$  of the subaggregator. This ends the proof.  $\square$

For example, if  $F$  is a projection of  $\mathbb{T}$  on  $\mathbb{F}$ , that is, if  $F^2 = F$ , then  $F$  is the identity mapping on  $\mathbb{F}$ , so that  $F$  satisfies the conditions of Proposition 3.3.11.

Notice that, if the factor  $F$  is only convex, we cannot conclude in general. For example, let  $A(h, t) = h + t$  be an aggregator and let  $F(t) = \exp(t)$  be a factor. Then the couple  $(A, F)$  is Weak Time Consistent with an associated subaggregator  $S^{A,F}(h, f) = h + \ln(f)$  which is not convex.

#### Homogeneity

As we are dealing with homogeneity property, we assume that the sets  $\mathbb{H}, \mathbb{T}, \mathbb{A}$  and  $\mathbb{F}$  in (3.1) are endowed with an external multiplication with the scalar field  $\mathbb{R}$ .

**Proposition 3.3.12** (Positive homogeneity). *Let the couple  $(A, F)$  be Weak Time Consistent, as in Definition 3.3.2. If the mapping  $A$  is jointly positively homogeneous and if the mapping  $F$  is positively homogeneous, then the subaggregator  $S^{A,F}$  is jointly positively homogeneous.*

*Proof.* Let  $(h, t)$  be element of  $\mathbb{H} \times \mathbb{T}$ . Let  $\lambda \in \mathbb{R}^+$ . We have the following equalities

$$S^{A,F}(\lambda h, \lambda F(t)) = S^{A,F}(\lambda h, F(\lambda t)) ,$$

by positive homogeneity of  $F$

$$= A(\lambda h, \lambda t) ,$$

by the Nested Formula (3.4)

$$= \lambda A(h, t) ,$$

by positive homogeneity of  $A$

$$= \lambda S^{A,F}(h, F(t)) ,$$

by the Nested Formula (3.4).

This ends the proof.  $\square$

### 3. Time consistent mappings

#### Translation invariance

As we are dealing with translation invariance, we assume that the sets  $\mathbb{H}, \mathbb{T}, \mathbb{A}$  and  $\mathbb{F}$  in (3.1) are endowed with an addition  $+$ . We also assume that there exists a set  $\mathbb{I}$  of invariants which is a common subspace of  $\mathbb{H}, \mathbb{T}, \mathbb{A}$  and  $\mathbb{F}$ , as follows.

**Definition 3.3.13.** Let  $\mathbb{X}$  and  $\mathbb{Y}$  be sets equipped with an addition  $+$ . Let  $\mathbb{I} \subset \mathbb{X} \cap \mathbb{Y}$  be a common subset of  $\mathbb{X}$  and  $\mathbb{Y}$ . A mapping  $M : \mathbb{X} \rightarrow \mathbb{Y}$  is said to be  $\mathbb{I}$ -translation invariant if

$$M(x + i) = M(x) + i^3, \quad \forall x \in \mathbb{X}, \quad \forall i \in \mathbb{I}. \quad (3.19)$$

**Proposition 3.3.14.** Let the couple  $(A, F)$  be Weak Time Consistent, as in Definition 3.3.2. If the mapping  $A$  is jointly translation invariant and if the mapping  $F$  is translation invariant then the subaggregator  $S^{A,F}$  is jointly translation invariant.

*Proof.* Let  $(h, t)$  be an element of  $\mathbb{H} \times \mathbb{T}$ . Let  $i \in \mathbb{I}$ . We have the following equalities:

$$S^{A,F}(h + i, F(t) + i) = S^{A,F}(h + i, F(t + i)),$$

by translation invariance of  $F$

$$= A(h + i, t + i),$$

by the Nested Formula (3.4)

$$= A(h, t) + i,$$

by translation invariance of  $A$

$$= S^{A,F}(h, F(t)) + i,$$

by the Nested Formula (3.4). We conclude that the subaggregator  $S^{A,F}$  is jointly translation invariant.  $\square$

## 3.4. Revisiting the literature

In Sect. 3.2, we have gone through a selection of papers, touching Time Consistency and Nested Formula in various settings. In Sect. 3.3, we have formally stated our (abstract) definitions of Time Consistency (TC) and Nested Formula (NF), and we have proven their equivalence. We have also provided conditions to obtain analytical properties of the mapping  $S^{A,F}$  appearing in the Nested Formula, such as monotonicity, continuity, convexity, positive homogeneity and translation invariance.

Now, we return to the literature that we have briefly reviewed in Sect. 3.2, and we show how our framework applies. For this purpose, we go through each article and try to answer two questions.

---

<sup>3</sup> To recover general formulation of the form  $M((i_1, 0, \dots, 0)) = i_1$ , we can extend Definition 3.3.13 by saying that the mapping  $M$  is translation invariant if  $M(x + i) = M(x) + \gamma(i)$  where  $\gamma : \mathbb{I} \rightarrow \mathbb{J}$  is an injection from a set  $\mathbb{I} \subset \mathbb{X}$  to a set  $\mathbb{J} \subset \mathbb{Y}$ . We prefer to stick to the Definition 3.3.13 to enhance the clarity of the paper.

First, what are the core assumptions that relate to our minimal notions of Time Consistency or Nested Formula? In particular, what are the heads and the tails and how are the Time Consistency axiom or the Nested Formula formulated? We will recover that the various definitions in the selection appear as special cases of ours.

Second, what are the assumptions that are additional to the core TC or NF formulations, and what do they imply for the subaggregator in the Nested Formula? We will extract the additional assumptions specific to each author and hence highlight their additional contribution.

### 3.4.1. Axiomatic for Time Consistency (TC)

We start our survey with the group of authors stating Time Consistency axiomatic. This group is subdivided between economists, who deal with lotteries and preferences, and probabilists, who deal with stochastic processes and dynamical risk measures.

#### Lotteries and preferences

Kreps and Porteus ((Kreps and Porteus, 1978a), (Kreps and Porteus, 1979)) state a temporal consistency axiom (Axiom 2.1) in the first paper. In the second paper, they focus on the particular case of two stage problems. Their axiomatic is an instance of our Definition 3.3.5 of Usual Time Consistency. With our Proposition 3.3.6, we directly deduce the existence of a subaggregator increasing in its second argument and a Nested Formula, whereas they obtain a stronger result under stronger assumptions. Indeed, they add assumptions of continuity, substitution (related to convexity) and focus on Usual Time Consistency with strict inequalities. This enables them to obtain a subaggregator which is continuous and strictly increasing in its second argument and is defined by ((Lemma 4, Theorem 2) and Proposition 1 respectively)

$$u_{y_t} : \{(z, \gamma) \in Z_t \times \mathbb{R} : \gamma = U_{y_t, z}(x) \text{ for some } x \in \mathbb{X}_{t+1}\} \rightarrow \mathbb{R}. \quad (3.20)$$

Epstein and Schneider (Epstein and Schneider, 2003a) state an axiom of Dynamic Consistency (Axiom 4: DC) which is a particular case of our Definition 3.3.5 of Usual Time Consistency. With our Proposition 3.3.6, we directly deduce the existence of a subaggregator increasing in its second argument and a Nested Formula, whereas they obtain a stronger result under stronger assumptions. Indeed, they introduce four additional axioms — Conditional Preferences (CP), Multiple Priors (MP), Risk Preference (RP) and Full Support (FS) — that ensure a particular form of the subaggregator. MP and CP ensure that the subaggregator can be represented as a minimum of expectation over a rectangular set of probabilities which is closed and convex. MP and RP ensure that the criterion is additive over time. FS ensures that the probability measures have full support. Epstein and Schneider obtain the

### 3. Time consistent mappings

following Nested Formula<sup>4</sup> associated to Time Consistency (Theorem 3.2):

$$V_t(h, \omega) = \min_{m \in \mathcal{P}_t^{+1}(\omega)} \int \left[ u(h_t(\omega)) + \beta V_{t+1}(h) \right] dm . \quad (3.21)$$

#### Dynamic risk measures and processes

Ruszczyński studies (Ruszczyński, 2010) dynamic risk measures  $\{\rho_{s,T}\}_{s=1}^T$ . Time Consistency (his Definition 3), appears as a particular case of our Usual Time Consistency Definition 3.3.5. With our Proposition 3.3.6, we directly deduce the existence of a subaggregator increasing in its second argument and a Nested Formula, whereas Ruszczyński obtains a stronger result under stronger assumptions. Indeed, he adds assumptions that induce a particular form for the subaggregator. From a conditional risk measure  $\rho_{s,T}$ , he defines mappings  $\rho_{s,s'}$  with  $s \leq s' \leq T$ . With our notations for aggregator  $A$  and factor  $F$ , he then focuses on the case where the initial assessment is  $A = \rho_{s,T}$  and the future assessment is  $F = \rho_{s',T}$ . With two additional assumptions of invariance by translation and normalization ( $\rho_{s,T}(0) = 0$ ), Ruszczyński is able to state that the subaggregator has the specific form (Theorem 1)

$$S^{A,F} = \rho_{s,s'} . \quad (3.22)$$

In (Artzner, Delbaen, Eber, Heath, and Ku, 2007a), Artzner, Delbaen, Eber, Heath and Ku present Time Consistency (their Definition 4.1) which appears as an instance of our Definition 3.3.5 of Usual Time Consistency. With our Proposition 3.3.6, we directly deduce the existence of a subaggregator increasing in its second argument and a Nested Formula, whereas they obtain a stronger result under stronger assumptions. Indeed, they study particular mappings of the form  $\Psi_t = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\cdot | \mathcal{F}_t]$ , where  $\mathcal{P}$  is a subset of probabilities and  $(\mathcal{F}_t)_{t=0}^T$  is a filtration. They make an intermediary step before presenting a Nested Formula. They use a tool that they name stability by pasting (rectangularity) of the set  $\mathcal{P}$  of probability distributions. With our notations for aggregator  $A$  and factor  $F$ , this enables them to obtain, for  $s \leq s'$ , that if  $A = \Psi_s$  and  $F = \Psi_{s'}$  then the subaggregator has the specific form (Theorem 4.2):

$$S^{A,F}(h, \cdot) = \Psi_s(h + \cdot) . \quad (3.23)$$

#### 3.4.2. Axiomatic for Nested formulas (NF)

Shapiro and Ruszczyński (Ruszczyński and Shapiro, 2006a) study a family of conditional risk mapping  $\rho_t = \rho_{x_2|x_1} \circ \dots \circ \rho_{x_t|x_{t-1}}$  (Equation (5.8)). Each  $\rho_t$  is increasing and is associated with a  $\sigma$ -algebra  $\mathcal{F}_t$ , where  $(\mathcal{F}_t)_{t=2}^T$  is a filtration. As these mappings  $\rho_t$  are instances of the mappings in our Nested Formula (3.4), they are Usual Time Consistent, by using our Proposition 3.3.6. With our notations for aggregator  $A$  and factor  $F$ , and with additional assumptions of monotonicity, translation

<sup>4</sup> Equation (3.21) is the original transcription of the formula in (Epstein and Schneider, 2003a), to which we refer the reader for a better understanding. By laying it out, we only want to stress the Nested Formula between  $V_t$  and  $V_{t+1}$ .

### 3.5. Additional characterizations of time consistency

invariance, convexity and homogeneity, Shapiro and Ruszczyński obtain that, if the initial assessment is  $A = \rho_t$  and the future assessment is  $F = \rho_{t+1}$ , then the subaggregator is (Theorem 5.1)

$$S^{A,F} = \rho_{x_{t+1}|x_t} . \quad (3.24)$$

Shapiro ((Shapiro, 2016)) focuses on a future assessment and on a subaggregator of the form (Definition 2.1)

$$F = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [\dots \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [\cdot | \mathcal{F}_{T-1}] | \mathcal{F}_0] , \quad S^{A,F} = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [\cdot] .$$

With our notations for aggregator  $A$  and factor  $F$ , this Nested Formula is an instance of our Nested Formula (3.4). We can define a natural initial assessment which is Usual Time Consistent with the future assessment, by using our Proposition 3.3.6. With additional assumptions of finiteness, Shapiro obtains that there exists a bounded set  $\hat{\mathcal{P}}$  of probability distributions such that the initial assessment has the specific form (Theorem 2.1)

$$A = \sup_{\mathbb{P} \in \hat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}} [\cdot] . \quad (3.25)$$

Besides, with additional assumption (Theorem 2.2) that  $\mathcal{P}$  is convex, bounded and weakly closed, Shapiro establishes that  $\mathcal{P} = \hat{\mathcal{P}}$ .

De Lara and Leclère (De Lara and Leclère, 2016) study composition of one time step aggregators. They make a distinction between uncertainty aggregator and time step aggregator, and they write a Nested Formula (Equation (11)) which is an instance of our Formula (3.4). We can naturally define an initial assessment from this composition operation which is time consistent with the one time step aggregator, by using our Proposition 3.3.6. They add an additional hypothesis of monotonicity and one of commutation between uncertainty aggregator and time aggregator. They deduce that the initial assessment can be defined as the composition between a one time step aggregator (subaggregator) and a future assessment (Theorem 9).

## 3.5. Additional characterizations of time consistency

Section § 3.5 is not part of the preprint but has been added since it provides complementary results, useful in the rest of the manuscript.

### 3.5.1. Characterization of WTC in terms of partitions

We characterize Weak Time Consistent factors as defined in (3.1) in term of partitions. For this purpose, we introduce the following definition.

**Definition 3.5.1.** With any  $A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{R}$ , we associate the family  $(A_h)_{h \in \mathbb{H}}$  of mappings defined by

$$\forall h \in \mathbb{H} , \quad A_h : \mathbb{T} \rightarrow \mathbb{A} , \quad t \mapsto A(h, t) . \quad (3.26)$$

### 3. Time consistent mappings

We denote by  $\mathbb{T}/A_h$  the partition of the domain  $\mathbb{T}$  induced by the mapping  $A_h$ , that is  $t, t'$  belong to the same equivalence class if and only if  $A(h, t) = A(h, t')$ . We denote by  $\bigvee_{h \in \mathbb{H}} \mathbb{T}/A_h$  the (less fine) partition of  $\mathbb{T}$  induced by the family  $(A_h)_{h \in \mathbb{H}}$  of mappings (where the notation  $\bigvee$  is defined in Definition 9.2.2 in Appendix 9.2).

We show that a couple  $(A, F)$  is WTC if and only if the partition  $\mathbb{T}/F$  induced by the mapping  $F : \mathbb{T} \rightarrow \mathbb{F}$  is finer than any of the partitions  $\mathbb{T}/A_h$ .

**Proposition 3.5.2.** *The couple aggregator-factor  $(A, F)$  in (3.1) is Weak Time Consistent if and only if*

$$\bigvee_{h \in \mathbb{H}} \mathbb{T}/A_h \preceq \mathbb{T}/F, \quad (3.27)$$

which reads "the partition  $\mathbb{T}/F$  is finer than the partition  $\bigvee_{h \in \mathbb{H}} \mathbb{T}/A_h$ " and where notation  $\preceq$  is defined in Definition 9.2.4 (See Appendix 9.1).

*Proof.*

1. Suppose that the couple  $(A, F)$  is weak time consistent. We fix  $h \in \mathbb{H}$  and we consider  $(t, t') \in \mathbb{T}^2$  such that  $F(t) = F(t')$ . Then, by the weak time consistency property (3.3), we have  $A_h(t) = A_h(t')$ . This means that when two elements are in the same atoms (see Appendix 9.1) of the partition  $\mathbb{T}/F$ , they are also in the same atoms in the partition  $\mathbb{T}/A_h$ . As it is defined in Appendix 9.1, we write  $\mathbb{T}/A_h \preceq \mathbb{T}/F$ . The previous implications are true for all  $h \in \mathbb{H}$ , so Equation (3.27) is satisfied.
2. We now suppose that Equation (3.27) is satisfied. We have in particular that

$$\mathbb{T}/A_h \preceq \mathbb{T}/F, \quad (3.28)$$

for a chosen  $h \in \mathbb{H}$ . For this particular element  $h$ , Equation (3.28) states, by definition, that  $F(t) = F(t') \Rightarrow A_h(t) = A_h(t')$ . The previous reasoning can be done with every element  $h \in \mathbb{H}$ .

This ends the proof. □

We state a corollary using the subaggregator  $S^{A,F}$  defined in (3.2).

**Corollary 3.5.3.** *If the couple aggregator-factor  $(A, F)$  in (3.1) is Weak Time Consistent then*

$$\bigvee_{h \in \mathbb{H}} \mathbb{T}/(S_h^{A,F} \circ F) \preceq \mathbb{T}/F, \quad (3.29)$$

where  $S_h^{A,F} : \mathbb{F} \rightarrow \mathbb{A}$  is a mapping defined for all  $h \in \mathbb{H}$  by  $S_h^{A,F}(f) = S^{A,F}(h, f)$ .

*Proof.* Suppose the couple  $(A, F)$  is WTC. Then we know by Theorem 3.3.3 that  $A(h, t) = S^{A,F}(h, F(t))$  and hence

$$\mathbb{T}/A_h = \mathbb{T}/(S_h^{A,F} \circ F), \quad \forall h \in \mathbb{H}. \quad (3.30)$$

From Equation (3.30) and Proposition 3.5.2 we get Equation 3.29. This ends the proof. □

### 3.5. Additional characterizations of time consistency

The following corollary, whose proof is left to the reader, is a direct consequence of Proposition 3.5.2 and Proposition 3.5.3.

**Corollary 3.5.4.** *The couple aggregator-factor  $(A, F)$  in (3.1) is Weak Time Consistent if and only if*

$$\bigvee_{h \in \mathbb{H}} \mathbb{T} / (S_h^{A, F} \circ F) \preceq \mathbb{T} / F . \quad (3.31)$$

#### 3.5.2. Characterizations of UTC in terms of level sets

In the Weak Time Consistency property (3.3), we made no assumptions to order the values of the mappings  $F$  and  $A$ . This is not the case for the Usual Time Consistency property (3.12) anymore. The characterization of Weak Time Consistency by partitions in Proposition 3.5.2 can be extended to Usual Time Consistency as a characterization in terms of level sets.

We first define the level sets of a mapping  $M$ .

**Definition 3.5.5.** Let  $M : \mathbb{X} \rightarrow \mathbb{Y}$  be a mapping where  $\mathbb{Y}$  is equipped with an order denoted by  $\leq$ . The level set of value  $y$  of the mapping  $M$  is defined by

$$M_{\leq y} = \{x \in \mathbb{X} \mid M(x) \leq y\} . \quad (3.32)$$

Note that  $M$  is not necessarily real valued.

The following proposition states that the couple  $(A, F)$  is Usual Time Consistent if and only if every level set of the mapping  $F$  is included in an intersection of level sets induced by the family of mappings  $(A_h)_{h \in \mathbb{H}}$ .

**Proposition 3.5.6.** *The couple aggregator-factor  $(A, F)$  in (3.1) is Usual Time Consistent if and only if*

$$F_{\leq F(t')} \subset \bigcap_{h \in \mathbb{H}} A_{h \leq A_h(t')} , \quad \forall t' \in \mathbb{T} . \quad (3.33)$$

*Proof.*

1. We suppose that the couple  $(A, F)$  is usual time consistent. We fix  $h \in \mathbb{H}$  and  $t' \in \mathbb{T}$ . We consider  $t \in \mathbb{T}$  such that  $F(t) \leq F(t')$ . Then, by usual time consistency property (3.12), we have  $A_h(t) \leq A_h(t')$  and thus  $F_{\leq F(t')} \subset A_{h \leq A_h(t')}$  for all  $(h, t') \in \mathbb{H} \times \mathbb{T}$  by (3.32). The previous implications are true for all  $h \in \mathbb{H}$ , so Equation (3.33) is satisfied.
2. We consider now that Equation (3.33) holds true so that

$$F_{\leq F(t')} \subset A_{h \leq A_h(t')} , \quad (3.34)$$

for a chosen couple  $(h, t') \in \mathbb{H} \times \mathbb{T}$ . For this particular couple, equation (3.34) states, by definition (3.32), that  $F(t) \leq F(t') \Rightarrow A_h(t) \leq A_h(t')$ . The previous reasoning can be done with every couple  $(h, t')$  of  $\mathbb{H} \times \mathbb{T}$ . This ends the proof.

□



### 3. Time consistent mappings

The following corollary, whose proof is left to the reader, is the counterpart of Corollary 3.5.3.

**Corollary 3.5.7.** *If the couple aggregator-factor  $(A, F)$  in (3.1) is Usual Time Consistent then*

$$F_{\leq F(t')} \subset \bigcap_{h \in \mathbb{H}} (S_h^{A,F} \circ F)_{\leq (S_h^{A,F} \circ F)(t')}, \quad \forall t' \in \mathbb{T}, \quad (3.35)$$

where  $S_h^{A,F} : \mathbb{F} \rightarrow \mathbb{A}$  is a mapping defined for all  $h \in \mathbb{H}$  by  $S_h^{A,F}(f) = S^{A,F}(h, f)$ .

### 3.5.3. Characterizations of STC in terms of level sets

In the usual time consistency property (3.12), we made no assumptions to order the element of the set  $\mathbb{H}$ . This is not the case for the strong time consistency property (3.13). The characterization of UTC in terms of level sets of Proposition 3.5.6 can be extended to STC as a characterization in terms of level sets over well chosen subsets of  $\mathbb{H}$ .

We recall that  $F_{\leq f}$  defined in (3.32) is the level set of value  $f$  for the mapping  $F$ .

**Proposition 3.5.8.** *The couple aggregator-factor  $(A, F)$  in (3.1) is strong time consistent (STC) if and only if*

$$F_{\leq F(t')} \subset \bigcap_{h \leq h'} A_{h \leq A_{h'}(t')}, \quad \forall t' \in \mathbb{T}. \quad (3.36)$$

Note that the intersection of level sets is over all  $h \leq h'$  in Equation (3.36) and over all  $h \in \mathbb{H}$  in Equation (3.33).

*Proof.*

1. We suppose that the couple  $(A, F)$  is strong time consistent. We fix  $h' \in \mathbb{H}$  and  $t' \in \mathbb{T}$ . We consider  $t \in \mathbb{T}$  such that  $F(t) \leq F(t')$  and  $h \in \mathbb{H}$  such that  $h \leq h'$ . Then, by strong time consistency property (3.12), we have  $A_h(t) \leq A_{h'}(t')$  and thus  $F_{\leq F(t')} \subset A_{h \leq A_{h'}(t')}$ . The previous implications are true for all  $h \leq h'$ , so that Equation (3.36) is satisfied.
2. We consider now that Equation (3.36) holds true, so that

$$F_{\leq F(t')} \subset A_{h \leq A_{h'}(t')}, \quad (3.37)$$

for any couple  $(h', t') \in \mathbb{H} \times \mathbb{T}$  and any  $h \leq h'$ . For a given couple and a given element, Equation (3.37) states  $F(t) \leq F(t') \Rightarrow A_h(t) \leq A_{h'}(t')$ . The previous reasoning can be done with every couple  $(h', t')$  of  $\mathbb{H} \times \mathbb{T}$  and any  $h \leq h'$ . This ends the proof. □

## 3.6. Conclusion

Time Consistency is a notion discussed in economics (dynamic optimization, bargaining) and mathematics (dynamical risk measures, multi-stage stochastic optimization). We have gone through a selection of papers that are representative of the different fields; we have tried to separate the common core elements related to Time Consistency from the additional assumptions that make the specific contribution of each author. We have presented a framework of Weak Time Consistency which allows us to prove an equivalence with a Nested Formula, under minimal assumptions. By formulating the core skeleton axioms, we hope to have shed light on the notion of Time Consistency, often melted with other notions in the literature. We believe that this makes the notion more transparent and that it opens the way for possible extensions.

First, in our setting for Weak Time Consistency, we introduced heads and tails with asymmetric roles. More generally, we could consider a symmetric axiomatic of the form

$$\left. \begin{array}{l} F_{\mathbb{H}}(h) = F_{\mathbb{H}}(h') \\ F_{\mathbb{T}}(t) = F_{\mathbb{T}}(t') \end{array} \right\} \Rightarrow A(h, t) = A(h', t') , \quad \forall (h, h', t, t') \in \mathbb{H}^2 \times \mathbb{T}^2 , \quad (3.38)$$

and a Nested Formula

$$A(h, t) = S^{F_{\mathbb{H}}, F_{\mathbb{T}}, A}(F_{\mathbb{H}}(h), F_{\mathbb{T}}(t)) . \quad (3.39)$$

Second, we used mappings to assess and compare objects. More generally, this can be done using general binary relations. Time Consistency now reads

$$\mathcal{H} \times \mathcal{T} \subset \mathcal{A} , \quad (3.40)$$

where the relation  $\mathcal{H}$  compares head elements, the relation  $\mathcal{T}$  compares tail elements, and the relation  $\mathcal{A}$  compares an overall head-tail object.

These possible extensions are presented in Chap. 5 where we do not define explicit time order between heads and tails.

**Acknowledgements.** The authors want to thank Université Paris-Est and Labex Bézout for the financial support. The first author particularly thanks them for the funding of his PhD program.



# 4. Three classes of time consistent mappings

## Contents

---

<b>4.1. Introduction</b>	<b>68</b>
<b>4.2. Time consistent translation invariant mappings</b>	<b>68</b>
4.2.1. Translation invariant mappings on a group	68
4.2.2. Characterization of UTC in terms of acceptance sets	69
<b>4.3. Time consistent convex mappings</b>	<b>75</b>
4.3.1. Basic tools to deal with Fenchel-Moreau's conjugacy	75
4.3.2. Relation between Nested Formula in the "primal" and decomposability in the "dual"	75
4.3.3. Illustration with convex risk measures	78
<b>4.4. Time consistent supremum mappings</b>	<b>81</b>
4.4.1. Background on lattices and their topologies	81
4.4.2. Usual time consistency with mappings defined with supremum	83
<b>4.5. Conclusion and perspectives</b>	<b>87</b>

---

## 4. Three classes of time consistent mappings

### 4.1. Introduction

In this section, we investigate the properties induced by time consistency for three classes of mappings. Indeed, we are motivated by the mappings describing risk measures but we remain as general as possible. In particular, we do not want to stick to an additive criterion, that is, we want to be able to encompass nested formula of the kind

$$\rho_0(\mathbf{X} \oplus \mathbf{Y}) = \rho_1(\mathbf{X} \oplus \rho_2(\mathbf{Y})) , \quad (4.1)$$

where  $\oplus$  is not necessarily the “usual addition”.

Three classes are studied in the world of risk measures. We investigate them in an abstract framework.

- We study in Sect. 4.2.1 translation invariant mappings motivated by the representation of risk measures in terms of acceptance set.
- Then, in Sect. 4.3, we study mappings that are defined as Fenchel-Moreau’s transform motivated by the dual reformulation of convex risk measures.
- Finally, in Sect. 4.4, we study mappings that are defined as supremum motivated by the representation of coherent risk measures (a subset of convex risk measures).

### 4.2. Time consistent translation invariant mappings

We study translation invariant mappings defined on ordered groups. We associate to each such mapping an acceptance set which is the level set of level 0. We prove that time consistency between two mappings is equivalent to an inclusion between acceptance sets.

#### 4.2.1. Translation invariant mappings on a group

We provide here the definition of a translation invariant mapping and the one of an acceptance set. With these notions, we will state our main contribution. We first recall the definition of an ordered group.

**Definition 4.2.1.** The triplet  $(\mathbb{F}, \oplus, \leq)$  is said to be an *ordered group* if  $\mathbb{F}$  is a set and

- $(\mathbb{F}, \oplus)$  is a group,
- $(\mathbb{F}, \leq)$  is an ordered set,
- the order  $\leq$  is compatible with  $\oplus$ , i.e.

$$f_1 \leq f_2 \Rightarrow f_1 \oplus f_3 \leq f_2 \oplus f_3 , \quad \forall (f_1, f_2, f_3) \in \mathbb{F}^3 . \quad (4.2)$$

We now provide the definition of translation invariant mappings on a group.

## 4.2. Time consistent translation invariant mappings

**Definition 4.2.2.** Let  $(\mathbb{T}, \oplus)$  be a commutative group and  $(\mathbb{F}, \oplus)$  be a subgroup of  $(\mathbb{T}, \oplus)$ , that we denote by

$$(\mathbb{F}, \oplus) \subset (\mathbb{T}, \oplus) . \quad (4.3)$$

A  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping is a mapping  $F : \mathbb{T} \rightarrow \mathbb{F}$  that satisfies

$$F(t \oplus f) = F(t) \oplus f , \quad \forall t \in \mathbb{T} , \quad \forall f \in \mathbb{F} . \quad (4.4)$$

In addition, if  $(F, \oplus, \leq)$  is an ordered group, we introduce the notations  $\mathcal{A}_F$  and  $\mathcal{A}_{F|\mathbb{F}}$  to deal with particular level sets of the  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping  $F : \mathbb{T} \rightarrow \mathbb{F}$ :

$$\mathcal{A}_F = \{t \in \mathbb{T} \mid F(t) \leq 0\} , \quad (4.5a)$$

$$\mathcal{A}_{F|\mathbb{F}} = \{f \in \mathbb{F} \mid F(f) \leq 0\} = \mathcal{A}_F \cap \mathbb{F} . \quad (4.5b)$$

### 4.2.2. Characterization of UTC in terms of acceptance sets

Given two translation invariant mappings  $F$  and  $\rho$  as in Definition 4.2.2, we will build an aggregator  $A_\rho$  such that the couple  $(A_\rho, F)$  is time consistent as in Definition 3.3.5.

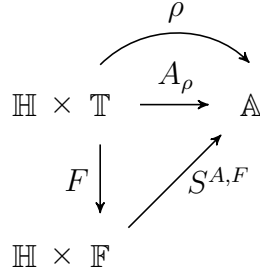


Figure 4.1.: Representation of links between mappings of Proposition 4.2.3

**Proposition 4.2.3.** Let  $(\mathbb{T}, \oplus)$  be a commutative group. Given two subgroups

$$(\mathbb{H}, \oplus) \subset (\mathbb{T}, \oplus) \text{ and } (\mathbb{A}, \oplus) \subset (\mathbb{T}, \oplus) , \quad (4.6)$$

and a  $(\mathbb{T}, \mathbb{A})$ -translation invariant mapping  $\rho : \mathbb{T} \rightarrow \mathbb{A}$ , we define the mapping  $A_\rho : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$  by

$$A_\rho : \mathbb{H} \times \mathbb{T} \rightarrow (\mathbb{A}, \oplus, \leq) , \quad (4.7)$$

$$(h, t) \mapsto \rho(h \oplus t) . \quad (4.8)$$

Let  $F : \mathbb{T} \rightarrow (\mathbb{F}, \oplus, \leq)$  be a  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping where

$$(\mathbb{F}, \oplus) \subset (\mathbb{T}, \oplus) \quad (4.9)$$

Then, the couple  $(A_\rho, F)$  is Usual Time Consistent if and only if

$$\mathcal{A}_F \oplus F(t) \subset \mathcal{A}_\rho * \mathbb{C}_\rho(t) , \quad \forall t \in \mathbb{T} , \quad (4.10)$$

#### 4. Three classes of time consistent mappings

where the starshape operation  $\underline{*}$  between sets is defined in Equation (9.17) and where the set valued mapping  $\mathbb{C}_\rho$  is given by

$$\mathbb{C}_\rho : \mathbb{T} \rightrightarrows \mathbb{T} , \quad (4.11)$$

$$t \mapsto \{h \ominus \rho(h \oplus t) , h \in \mathbb{H}\} . \quad (4.12)$$

*Proof.* We first recall the notation of a level set given in Definition 3.5.5. The level set of value  $y \in \mathbb{Y}$  of a mapping  $M : \mathbb{X} \rightarrow (\mathbb{Y}, \leq)$  is defined by

$$M_{\leq y} = \{x \in \mathbb{X} \mid M(x) \leq y\} . \quad (4.13)$$

We know from Proposition 3.5.6 that the couple  $(A_\rho, F)$  is Usual Time Consistent if and only if

$$F_{\leq F(t)} \subset \bigcap_{h \in \mathbb{H}} \rho(h \oplus \cdot)_{\leq \rho(h \oplus t)} , \quad \forall t' \in \mathbb{T} . \quad (4.14)$$

For  $h \in \mathbb{H}$  and  $t \in \mathbb{T}$  fixed, the level set  $\rho(h \oplus \cdot)_{\leq \rho(h \oplus t)}$  of the mapping  $\rho(h \oplus \cdot)$  is defined by

$$\begin{aligned} \rho(h \oplus \cdot)_{\leq \rho(h \oplus t)} &= \{t' \in \mathbb{T} \mid \rho(h \oplus t') \leq \rho(h \oplus t)\} , && \text{(by Definition 3.5.5,)} \\ &= \{t' \in \mathbb{T} \mid \rho(h \oplus t' \ominus \rho(h \oplus t)) \leq 0\} , && (\rho \text{ is } (\mathbb{T} - \mathbb{A})\text{-translation invariant)} \\ &= \{t' \in \mathbb{T} \mid h \oplus t' \ominus \rho(h \oplus t) \in \mathcal{A}_\rho\} , && \text{(Definition 4.5a)} \\ &= \mathcal{A}_\rho \oplus \rho(h \oplus t) \ominus h . \end{aligned}$$

Thus, for  $t \in \mathbb{T}$ , we have that

$$\begin{aligned} \bigcap_{h \in \mathbb{H}} \rho(h \oplus \cdot)_{\leq \rho(h \oplus t)} &= \bigcap_{h \in \mathbb{H}} \mathcal{A}_\rho \oplus \rho(h \oplus t) \ominus h , \\ &= \bigcap_{c \in \mathbb{C}_\rho(t)} \mathcal{A}_\rho \ominus c , && \text{(by Equation 4.12)} \\ &= \mathcal{A}_\rho \underline{*} \mathbb{C}_\rho(t) . && \text{(by Equation (9.17))} \end{aligned}$$

Using similar arguments, for  $t \in \mathbb{T}$  fixed, the level set  $F_{\leq F(t)}$  of the mapping  $F$  is defined by

$$\begin{aligned} F_{\leq F(t)} &= \{t' \in \mathbb{T} \mid F(t') \leq F(t)\} , && \text{(by Definition 3.5.5,)} \\ &= \{t' \in \mathbb{T} \mid F(t' \ominus F(t)) \leq 0\} , && (F \text{ is } (\mathbb{T} - \mathbb{F})\text{-translation invariant)} \\ &= \{t' \in \mathbb{T} \mid t' \ominus F(t) \in \mathcal{A}_F\} , && \text{(Definition 4.5a)} \\ &= \mathcal{A}_F \oplus F(t) . && (4.17a) \end{aligned}$$

This ends the proof. □

The next proposition and its proof provide, under additional assumptions, a reformulation of (4.10). They are a generalization, with our notations, of Lemma 11.14 and Proposition 11.15 of Föllmer and Schied (2016), since we do not refer to risk measures on probability spaces but to more general sets.

## 4.2. Time consistent translation invariant mappings

**Proposition 4.2.4.** *Let  $(\mathbb{T}, \oplus)$  be a commutative group. Given two subgroups*

$$(\mathbb{H}, \oplus) \subset (\mathbb{A}, \oplus) \subset (\mathbb{T}, \oplus), \quad (4.18)$$

and a  $(\mathbb{T}, \mathbb{A})$ -translation invariant mapping  $\rho : \mathbb{T} \rightarrow \mathbb{A}$ , we define the mapping  $A_\rho : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$  by

$$A_\rho : \mathbb{H} \times \mathbb{T} \rightarrow (\mathbb{A}, \oplus, \leq), \quad (4.19)$$

$$(h, t) \mapsto \rho(h \oplus t). \quad (4.20)$$

Let  $F : \mathbb{T} \rightarrow (\mathbb{F}, \oplus, \leq)$  be a  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping.

1. *The couple  $(A_\rho, F)$  is Usual Time Consistent if and only if*

$$\mathcal{A}_F \oplus F(t) \subset \mathcal{A}_\rho \oplus \rho(t), \quad \forall t \in \mathbb{T}. \quad (4.21)$$

2. *If, in addition, we have that*

- $(\mathbb{H}, \oplus) \subset (\mathbb{A}, \oplus) \subset (\mathbb{F}, \oplus) \subset (\mathbb{T}, \oplus)$ ,
- the  $(\mathbb{T}, \mathbb{A})$ -translation invariant mapping  $\rho : \mathbb{T} \rightarrow \mathbb{A}$  is increasing,
- the  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping  $F : \mathbb{T} \rightarrow \mathbb{F}$  satisfies  $F(0) = 0$  (where 0 is the neutral element of  $(\mathbb{T}, \oplus)$ ),

then the couple of mappings  $(A_\rho, F)$  is Usual Time Consistent if and only if

$$\mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} = \mathcal{A}_\rho, \quad (4.22)$$

where  $\mathcal{A}_F$ ,  $\mathcal{A}_{\rho|_{\mathbb{F}}}$  and  $\mathcal{A}_\rho$  are defined in (4.5a) and (4.5b).

*Proof.* We prove both statements successively.

1. As  $\mathbb{H} \subset \mathbb{A}$ , we have by  $(\mathbb{T}, \mathbb{A})$ -translation invariance of  $\rho : \mathbb{T} \rightarrow \mathbb{A}$  that  $\rho(t) = \rho(t \oplus h \ominus h) = \rho(t \oplus h) \ominus h = \rho(h \oplus t) \ominus h$  for any  $h \in \mathbb{H}$  and  $t \in \mathbb{T}$ . The result of Proposition 4.2.3 reads now

$$\mathcal{A}_F \oplus F(t) \subset \mathcal{A}_\rho * \{\rho(t)\}, \quad \forall t \in \mathbb{T}, \quad (4.23)$$

which, by definition (9.17) of the star difference, gives equation (4.21) and ends the proof of the first statement.

2. The proof goes in three steps as follows:

- a) first, we show that  $t \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} \Leftrightarrow F(t) \in \mathcal{A}_{\rho|_{\mathbb{F}}}$ ,  $\forall t \in \mathbb{T}$ ,
- b) then we use the previous assertion to prove the two following statements:

$$\mathcal{A}_\rho \subset \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} \Leftrightarrow \rho(F(t)) \leq \rho(t), \quad \forall t \in \mathbb{T}, \quad (4.24a)$$

$$\mathcal{A}_\rho \supset \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} \Leftrightarrow \rho(F(t)) \geq \rho(t), \quad \forall t \in \mathbb{T}, \quad (4.24b)$$

- c) finally, we bring all elements together to conclude.



#### 4. Three classes of time consistent mappings

We now detail each step.

a) We prove the implication  $t \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}} \Rightarrow F(t) \in \mathcal{A}_{\rho|\mathbb{F}}$  and the reverse statement  $F(t) \in \mathcal{A}_{\rho|\mathbb{F}} \Rightarrow t \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}}$  successively.

- Let  $t \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}}$  be given. By definition,  $t$  can be decomposed as  $t = t_F \oplus t_\rho$  with  $t_F \in \mathcal{A}_F$  and  $t_\rho \in \mathcal{A}_{\rho|\mathbb{F}}$ . We successively obtain

$$\begin{aligned} F(t) &= F(t_F) \oplus t_\rho, \\ &\quad (\text{as } t_\rho \in \mathbb{F} \text{ and } F \text{ is } (\mathbb{T}, \mathbb{F})\text{-translation invariant}) \\ &\leq t_\rho, \quad (\text{as } t_F \in \mathcal{A}_F = \{t \in \mathbb{T} \mid F(t) \leq 0\}) \end{aligned}$$

which leads to

$$\begin{aligned} \rho(F(t)) &\leq \rho(t_\rho), && (\text{by monotonicity of } \rho) \\ &\leq 0, && (\text{by definition of } t_\rho \in \mathcal{A}_{\rho|\mathbb{F}}) \end{aligned}$$

and hence,  $F(t) \in \mathcal{A}_{\rho|\mathbb{F}}$ .

- We now assume that  $F(t) \in \mathcal{A}_{\rho|\mathbb{F}}$  and recall that for all  $t \in \mathbb{T}$ ,  $F(t \ominus F(t)) = F(t) \ominus F(t) = 0$  by  $(\mathbb{T} - \mathbb{F})$ -translation invariance of the mapping  $F$ . The converse implication follows immediately from the decomposition  $t = t \ominus F(t) \oplus F(t)$  since  $F(t) \in \mathcal{A}_{\rho|\mathbb{F}}$  by assumption and  $t \ominus F(t) \in \mathcal{A}_F$ .

b) We prove statements (4.24a) and (4.24b) successively.

- First, we focus on equation (4.24a):

$$\mathcal{A}_\rho \subset \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}} \Leftrightarrow \rho(F(t)) \leq \rho(t), \quad \forall t \in \mathbb{T}, \quad (4.25)$$

We suppose that left hand side of this equation is satisfied, i.e.  $\mathcal{A}_\rho \subset \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}}$ , and we show that it implies the right hand side of the equation. For that purpose, we fix  $t \in \mathbb{T}$ . We recall that  $\rho(t) \in \mathbb{A} \subset \mathbb{F}$  by definition of the mapping  $\rho : \mathbb{T} \rightarrow \mathbb{A}$  and assumption  $(\mathbb{A}, \oplus) \subset (\mathbb{F}, \oplus)$ . We have that  $F(t) \ominus \rho(t) = F(t \ominus \rho(t))$  by  $(\mathbb{T}, \mathbb{F})$ -translation invariance of the mapping  $F$ . As  $t \ominus \rho(t) \in \mathcal{A}_\rho \subset \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}}$  we get by item 2a just above that  $F(t \ominus \rho(t)) \in \mathcal{A}_{\rho|\mathbb{F}}$  and then  $F(t) \ominus \rho(t) \in \mathcal{A}_{\rho|\mathbb{F}}$ . This implies that

$$\rho(F(t)) \ominus \rho(t) = \rho(F(t) \ominus \rho(t)) \leq 0. \quad (4.26)$$

Assume now that  $\rho(F(t)) \leq \rho(t)$  for all  $t \in \mathbb{T}$  and let  $\tilde{t} \in \mathcal{A}_\rho$ . Then by definition (4.5a) of an acceptance set, we got that  $\rho(\tilde{t}) \leq 0$ . It follows that  $\rho(F(\tilde{t})) \leq 0$  and that  $F(\tilde{t}) \in \mathcal{A}_{\rho|\mathbb{F}}$  and so, by item 2a just above, that  $\tilde{t} \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}}$ .

- Second, we focus on Equation (4.24b)

$$\mathcal{A}_\rho \supset \mathcal{A}_F \oplus \mathcal{A}_{\rho|\mathbb{F}} \Leftrightarrow \rho(F(t)) \geq \rho(t), \quad \forall t \in \mathbb{T}, \quad (4.27)$$

## 4.2. Time consistent translation invariant mappings

We assume  $\mathcal{A}_\rho \supset \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}}$ . Let us fix  $t \in \mathbb{T}$ . Then, by adding and removing the term  $F(t)$  we get

$$t \ominus \rho(F(t)) = \underbrace{t \ominus F(t)}_{\in \mathcal{A}_F} \oplus \underbrace{F(t) \ominus \rho(F(t))}_{\in \mathcal{A}_{\rho|_{\mathbb{F}}}} \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} . \quad (4.28)$$

It follows by left hand side of (4.24b) that  $t \ominus \rho(F(t))$  belongs to  $\mathcal{A}_\rho$ . That implies, taken together with the  $(\mathbb{T}, \mathbb{A})$ -translation invariance of the mapping  $\rho : \mathbb{T} \rightarrow \mathbb{A}$

$$\rho(t) \ominus \rho(F(t)) = \rho(t \ominus \rho(F(t))) \leq 0 . \quad (4.29)$$

To prove the reverse implication of Equation (4.24b), take  $t \in \mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}}$  and assume that  $\rho(F(t)) \geq \rho(t)$ . Using step 2a, we have that  $F(t) \in \mathcal{A}_{\rho|_{\mathbb{F}}}$  and we obtain that

$$\rho(t) \leq \rho(F(t)) \leq 0 , \quad (4.30)$$

which gives  $t \in \mathcal{A}_\rho$  by definition (4.5a) of an acceptance set.

- c) We finally bring all elements together. We know from Theorem 3.3.6 that the couple of mappings  $(A_\rho, F)$  is usual time consistent if and only if the subaggregator  $S^{A_\rho, F}$  defined in (3.2) is a mapping increasing in its second argument and we have the nested formula  $A_\rho(h, t) = S^{A_\rho, F}(h, F(t))$ .

In this case, by Definition 3.2, we have that

$$S^{A_\rho, F}(h, f) = \{A_\rho(h, t) \mid F(t) = f\} , \quad \forall (h, f) \in \mathbb{H} \times \mathbb{F} . \quad (4.31)$$

As the set-valued mapping  $S^{A_\rho, F}$  is a mapping, choosing one element  $t \in \mathbb{T}$  such that  $F(t) = f$  is sufficient to define the value of  $S^{A_\rho, F}(h, f)$ . We notice that, for each element  $f \in \mathbb{F}$ , the following statement holds true

$$F(t) = F(0) \oplus t . \quad (4.32)$$

By  $(\mathbb{T}, \mathbb{F})$ -translation invariance property (4.4), we have that  $F(f \ominus F(0)) = F(0) \oplus (f \ominus F(0)) = f$  for all  $f \in \mathbb{F}$ . We deduce that

$$S^{A_\rho, F}(h, f) = A_\rho(h, f \ominus F(0)) , \quad \forall (h, f) \in \mathbb{H} \times \mathbb{F} . \quad (4.33)$$

Hence, the nested formula  $A_\rho(h, t) = S^{A_\rho, F}(h, F(t))$  reads

$$\begin{aligned} A_\rho(h, t) &= A_\rho(h, F(t) \ominus F(0)) , && \text{(by (4.33))} \\ \rho(h \oplus t) &= \rho(h \oplus F(t) \ominus F(0)) , && \text{(by (4.19) that defines } A_\rho) \\ h \oplus \rho(t) &= h \oplus \rho(F(t) \ominus F(0)) , && \text{(by } (\mathbb{T}, \mathbb{A})\text{-translation invariance)} \\ \rho(t) &= \rho(F(t) \ominus F(0)) , && \text{(by compatibility of } \oplus \text{ with } \leq) \\ \rho(t) &= \rho(F(t)) && \text{(as } F(0) = 0.) \end{aligned}$$

The fact that  $\rho(t) = \rho(F(t))$  taken together with both statements of Equations (4.24) gives the wanted result.

This ends the proof.

#### 4. Three classes of time consistent mappings

□

We have provided a characterization of time consistency for translation invariant mappings in terms of acceptance sets leading to Equation (4.22)

$$\mathcal{A}_F \oplus \mathcal{A}_{\rho|_{\mathbb{F}}} = \mathcal{A}_\rho . \quad (4.34)$$

However, this equation is difficult to solve when the variables are the mappings  $\rho : \mathbb{T} \rightarrow \mathbb{A}$  and  $F : \mathbb{T} \rightarrow \mathbb{F}$  given in Proposition 4.2.4 since it is an implicit equation in  $\rho$ .

### 4.3. Time consistent convex mappings

In this section, we focus on time consistency for mappings that are defined as Fenchel-Moreau transforms. We are motivated by the result on dual representation of convex risk mappings. We first recall Fenchel-Moreau conjugacy with general couplings (not necessarily the classic duality pairing). Then, we state our main theorem that provide a nested formula and hence time consistency of mappings defined as Fenchel-Moreau transforms. We illustrate our contribution on an example with coherent risk measures.

#### 4.3.1. Basic tools to deal with Fenchel-Moreau's conjugacy

The formal tools of coupling and Fenchel-Moreau conjugate were introduced in [Moreau](#). We recall that

$$\overline{\mathbb{R}} = [-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}. \quad (4.35)$$

When we manipulate functions with values in  $\overline{\mathbb{R}} = [-\infty, +\infty]$ , we adopt the Moreau *lower addition* or *upper addition* defined in Equations (9.46a) and (9.47a), depending on whether we deal with sup or inf operations. We invite the reader to refer to Appendix 9.6 for a complete review. We only recall here definitions of coupling and transforms.

**Background on Fenchel-Moreau conjugacy with respect to a coupling.** Let be given two sets  $\mathbb{C}$  and  $\mathbb{C}^\sharp$ . Consider a *coupling* function  $\Phi : \mathbb{C} \times \mathbb{C}^\sharp \rightarrow [-\infty, +\infty]$ . We also use the notation  $\mathbb{C} \overset{\Phi}{\leftrightarrow} \mathbb{C}^\sharp$  for a coupling, so that

$$\mathbb{C} \overset{\Phi}{\leftrightarrow} \mathbb{C}^\sharp \iff \Phi : \mathbb{C} \times \mathbb{C}^\sharp \rightarrow [-\infty, +\infty]. \quad (4.36)$$

**Definition 4.3.1.** The *Fenchel-Moreau conjugate* of a function  $f : \mathbb{C} \rightarrow [-\infty, +\infty]$ , with respect to the coupling  $\Phi$  in (4.36), is the function  $f^\Phi : \mathbb{C}^\sharp \rightarrow [-\infty, +\infty]$  defined by

$$f^\Phi(c^\sharp) = \sup_{c \in \mathbb{C}} \left( \Phi(c, c^\sharp) \dot{+} (-f(c)) \right), \quad \forall c^\sharp \in \mathbb{C}^\sharp. \quad (4.37)$$

#### 4.3.2. Relation between Nested Formula in the “primal” and decomposability in the “dual”

We provide a nested formula between mappings defined as Fenchel-Moreau transforms.

##### Main result: nested formula for Fenchel conjugates

We introduce the notion of decomposable coupling.

**Definition 4.3.2.** Let  $\mathbb{X}$ ,  $\mathbb{Y}$ ,  $\mathbb{Z}$  and  $\mathbb{Y}'$  be four sets and let  $\theta_{\mathbb{X} \times \mathbb{Z}}$ ,  $\theta_{\mathbb{Z}}$  and  $\theta_{\mathbb{X}}$  be three mappings with values in  $\mathbb{Y}'$

$$\theta_{\mathbb{X} \times \mathbb{Z}} : \mathbb{X} \times \mathbb{Z} \rightarrow \mathbb{Y}', \quad (4.38a)$$

$$\theta_{\mathbb{Z}} : \mathbb{Z} \rightarrow \mathbb{Y}', \quad (4.38b)$$

$$\theta_{\mathbb{X}} : \mathbb{X} \rightarrow \mathbb{Y}'. \quad (4.38c)$$

#### 4. Three classes of time consistent mappings

Let  $\varphi : \mathbb{Y}' \times \mathbb{Y} \rightarrow [-\infty, +\infty]$  be a coupling between  $\mathbb{Y}'$  and  $\mathbb{Y}$ .

We say that the coupling  $\varphi$  is  $(\theta_{\mathbb{X} \times \mathbb{Z}}, \theta_{\mathbb{X}}, \theta_{\mathbb{Z}})$ -decomposable if

$$\varphi(\theta_{\mathbb{X}}(x), y) = \sup_{z \in \mathbb{Z}} \left\{ \varphi(\theta_{\mathbb{X} \times \mathbb{Z}}(x, z), y) \dot{+} \left( -\varphi(\theta_{\mathbb{Z}}(z), y) \right) \right\}, \quad (4.39)$$

$$\forall (x, y) \in \mathbb{X} \times \mathbb{Y}.$$

Here is our main proposition.

**Proposition 4.3.3.** *Let  $\mathbb{X}$ ,  $\mathbb{Y}$ ,  $\mathbb{Z}$  and  $\mathbb{Y}'$  be four sets and  $g : \mathbb{Y} \rightarrow [-\infty, +\infty]$  be a numerical function. Let  $\varphi : \mathbb{Y}' \times \mathbb{Y} \rightarrow [-\infty, +\infty]$  be  $(\theta_{\mathbb{X} \times \mathbb{Z}}, \theta_{\mathbb{X}}, \theta_{\mathbb{Z}})$ -decomposable as in Definition 4.3.2.*

Let us define the coupling  $\Phi : \mathbb{X} \times (\mathbb{Y} \times \mathbb{Z}) \rightarrow [-\infty, +\infty]$  by

$$\Phi(x, (y, z)) = \varphi(\theta_{\mathbb{X} \times \mathbb{Z}}(x, z), y), \quad \forall (x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}, \quad (4.40)$$

and the function  $G : \mathbb{Y} \times \mathbb{Z} \rightarrow [-\infty, +\infty]$  by

$$G(y, z) = g(y) \dot{+} \varphi(\theta_{\mathbb{Z}}(z), y), \quad \forall (y, z) \in \mathbb{Y} \times \mathbb{Z}. \quad (4.41)$$

Then, we have the following Nested Formula between Fenchel-Moreau conjugates:

$$G^\Phi = g^\varphi \circ \theta_{\mathbb{X}}. \quad (4.42)$$

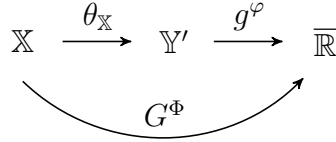


Figure 4.2.: Representation of the Nested Formula (4.42)

*Proof.* We have, for any  $x \in \mathbb{X}$ , the following equalities

$$G^\Phi(x) = \sup_{(y,z) \in \mathbb{Y} \times \mathbb{Z}} \left\{ \Phi(x, (y, z)) \dot{+} \left( -G(y, z) \right) \right\}, \quad (4.43)$$

by Equation (4.37) that expresses the  $\Phi$ -conjugate of  $G$ ,

$$= \sup_{(y,z) \in \mathbb{Y} \times \mathbb{Z}} \left\{ \varphi(\theta_{\mathbb{X} \times \mathbb{Z}}(x, z), y) \dot{+} \left( -g(y) \right) \dot{+} \left( -\varphi(\theta_{\mathbb{Z}}(z), y) \right) \right\}, \quad (4.44)$$

by Equations (4.40) and (4.41) that express particular forms of  $\Phi$  and  $G$ , and by the joint property (9.48b) of Moreau's additions,

$$= \sup_{y \in \mathbb{Y}} \left\{ -g(y) \dot{+} \sup_{z \in \mathbb{Z}} \left\{ \varphi(\theta_{\mathbb{X} \times \mathbb{Z}}(x, z), y) \dot{+} \left( -\varphi(\theta_{\mathbb{Z}}(z), y) \right) \right\} \right\}, \quad (4.45)$$

by property (9.46e) of Moreau's additions,

$$= \sup_{y \in \mathbb{Y}} \left\{ -g(y) \dot{+} \varphi(\theta_{\mathbb{X}}(x), y) \right\}, \quad (4.46)$$

by Equation (4.39) that expresses the supremum,

$$= g^\varphi(\theta_{\mathbb{X}}(x)), \quad (4.47)$$

by Definition 4.3.1 of a Fenchel-Moreau transform. This ends the proof.  $\square$

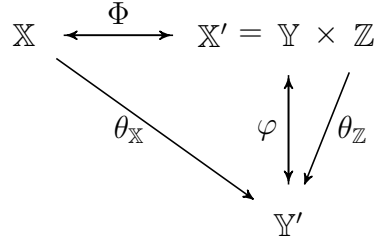


Figure 4.3.: Representation of links between the mappings of Definition 4.3.2

### Special cases of the Nested formula

We now focus on two particular cases: a case of product set and a case of indicator mappings.

**The case of product set.** In that case,  $\mathbb{X} = \mathbb{X}_1 \times \mathbb{X}_2$ . We are motivated by Agent Consistency that will be presented in the next Chapter 5.2. The proof is left to the reader as an application of Proposition 4.3.3.

**Proposition 4.3.4.** *Let  $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}', \mathbb{Y}, \mathbb{Z}, \mathbb{Y}'_1$  and  $\mathbb{Y}'_2$  be six sets and let  $G : \mathbb{Y} \times \mathbb{Z} \rightarrow [-\infty, +\infty]$  be a mapping.*

*Let  $\Phi$  and  $\varphi$  be two couplings as follows*

$$\Phi : (\mathbb{X}_1 \times \mathbb{X}_2) \times (\mathbb{Y} \times \mathbb{Z}) \rightarrow [-\infty, +\infty] , \quad (4.48)$$

$$\varphi : (\mathbb{Y}'_1 \times \mathbb{Y}'_2) \times \mathbb{Y} \rightarrow [-\infty, +\infty] . \quad (4.49)$$

*We assume that there exists six mappings*

$$\theta_{\mathbb{X}_1 \times \mathbb{Z}} : \mathbb{X}_1 \times \mathbb{Z} \rightarrow \mathbb{Y}'_1 , \quad (4.50)$$

$$\theta_{\mathbb{X}_2 \times \mathbb{Z}} : \mathbb{X}_2 \times \mathbb{Z} \rightarrow \mathbb{Y}'_2 , \quad (4.51)$$

$$\theta_{\mathbb{Z},1} : \mathbb{Z} \rightarrow \mathbb{Y}'_1 , \quad (4.52)$$

$$\theta_{\mathbb{Z},2} : \mathbb{Z} \rightarrow \mathbb{Y}'_2 , \quad (4.53)$$

$$\theta_{\mathbb{X}_1} : \mathbb{X}_1 \rightarrow \mathbb{Y}'_1 , \quad (4.54)$$

$$\theta_{\mathbb{X}_2} : \mathbb{X}_2 \rightarrow \mathbb{Y}'_2 , \quad (4.55)$$

*and a numerical function  $g : \mathbb{Y} \rightarrow [-\infty, +\infty]$  such that, for all  $(x_1, x_2, y, z)$  in  $\mathbb{X}_1 \times \mathbb{X}_2 \times \mathbb{Y} \times \mathbb{Z}$*

$$\Phi(x_1, x_2, y, z) = \varphi\left(\left(\theta_{\mathbb{X}_1 \times \mathbb{Z}}(x_1, z), \theta_{\mathbb{X}_2 \times \mathbb{Z}}(x_2, z)\right), y\right) , \quad (4.56)$$

$$G(y, z) = g(y) \dot{+} \varphi\left(\left(\theta_{\mathbb{Z},1}(z), \theta_{\mathbb{Z},2}(z)\right), y\right) , \quad (4.57)$$

$$\begin{aligned} \varphi\left(\left(\theta_{\mathbb{X}_1}(x_1), \theta_{\mathbb{X}_2}(x_2)\right), y\right) &= \sup_{z \in \mathbb{Z}} \left\{ \varphi\left(\left(\theta_{\mathbb{X}_1 \times \mathbb{Z}}(x_1, z), \theta_{\mathbb{X}_1 \times \mathbb{Z}}(x_2, z)\right), y\right) \right. \\ &\quad \left. \dot{+} \varphi\left(\left(\theta_{\mathbb{Z},1}(z), \theta_{\mathbb{Z},2}(z)\right), y\right) \right\} . \end{aligned} \quad (4.58)$$

*Then, we have the following Nested Formula between Moreau-Fenchel conjugates:*

$$G^\Phi = g^\varphi \circ (\theta_{\mathbb{X}_1}, \theta_{\mathbb{X}_2}) . \quad (4.59)$$

#### 4. Three classes of time consistent mappings

**The case of indicator mappings.** This proposition is inspired by coherent risk measures for which the dual representation is a support function.

**Proposition 4.3.5.** *Let the mapping  $g$  in Proposition 4.3.3 be,  $\delta_Y$ , the indicator mapping of the set  $Y \subset \mathbb{Y}$  that is,  $g$  takes the value 0 on  $Y$  and  $+\infty$  otherwise. We assume that there exist two elements  $0_{\mathbb{Y}'}$  and  $+\infty_{\mathbb{Y}'}$  elements of  $\mathbb{Y}'$  such that the coupling  $\varphi$  in Proposition 4.3.3 satisfies*

$$\varphi(0_{\mathbb{Y}'}, y) = 0 \in \overline{\mathbb{R}}, \quad \varphi(-\infty_{\mathbb{Y}'}, y) = -\infty \in \overline{\mathbb{R}}, \quad \forall y \in \mathbb{Y}. \quad (4.60)$$

For a set  $Z \subset \mathbb{Z}$ , we define the mapping  $\theta_Z$  that takes the value  $0_{\mathbb{Y}'}$  on  $Z$  and  $+\infty_{\mathbb{Y}'}$  otherwise. Then, under the assumptions of Proposition 4.3.3 the mapping  $G$  in (4.41) is given by

$$G(y, z) = \delta_Y(y) + \varphi(\tilde{\delta}_Z(z), y) = \delta_{Y \times Z}(y, z). \quad (4.61)$$

If in addition the coupling  $\varphi$  is a bilinear coupling between the vector spaces  $\mathbb{Y}$  and  $\mathbb{Y}'$ , we obtain by (4.42) that the mapping  $G^\Phi$  is the composition of a support function with a mapping.

Previous proposition expresses the fact that the mapping  $G$  is the indicator mapping of the rectangle  $X' = Y \times Z$  included in the “dual” set  $\mathbb{X}' = \mathbb{Y} \times \mathbb{Z}$ .

#### 4.3.3. Illustration with convex risk measures

We illustrate how Proposition 4.3.3 is linked to well known results concerning a class of convex risk measures, defined as support functions by means of probability subsets.

**Basic notions** Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a probability space and  $L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}})$  be the vector space of essentially bounded random variables on  $\Omega$ . We consider  $\mathcal{F}_0 = \{\emptyset, \Omega\}$  the trivial  $\sigma$ -field over  $\Omega$  and  $\mathcal{F}$  a subfield of  $\mathcal{T}$ :

$$\{\emptyset, \Omega\} \subset \mathcal{F} \subset \mathcal{T}. \quad (4.62)$$

We denote by  $\Delta(\Omega, \mathcal{T})$  the set of probability distributions over  $(\Omega, \mathcal{T})$ . We will denote by  $\mathbb{Q}$  a generic element of  $\Delta(\Omega, \mathcal{T})$  and by  $\mathbf{X}$  a generic element of  $L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}})$ . For convenience, we introduce the notation  $\mathcal{M} = \Delta(\Omega, \mathcal{F})$  to denote the set of distributions on the  $\sigma$ -field  $\mathcal{F}$  and denote by  $m$  a general element of  $\mathcal{M}$ . We have

$$\mathcal{M} = \Delta(\Omega, \mathcal{F}) \subset \Delta(\Omega, \mathcal{T}). \quad (4.63)$$

We interpret the set  $\mathcal{M}$  as a set of marginal distributions. The set  $\mathcal{K}(\Omega, \mathcal{F})$  of Radon-Nikodim derivatives with respect to  $\mathcal{F}$  is defined by

$$\mathcal{K}(\Omega, \mathcal{F}) = \{ \mathbf{X} \in L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}}) \mid \mathbf{X} \geq 0, \mathbf{X} \leq 1, \mathbb{E}_{\mathbb{P}}[\mathbf{X} \mid \mathcal{F}] = \mathbf{1} \}, \quad (4.64)$$

where equalities and inequalities are taken  $\mathbb{P}$ -almost surely. We denote by  $k$  a generic element of  $\mathcal{K}(\Omega, \mathcal{F})$ . For convenience, we use the abuse of notation  $\mathcal{K}$ .

Proposition 4.3.3	Risk measures formalism
$\mathbb{X}$	$L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}})$
$\mathbb{Y}$	$\mathcal{M} = \Delta(\Omega, \mathcal{F})$
$\mathbb{Z}$	$\mathcal{K}$
$\mathbb{Y}'$	$L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \overline{\mathbb{R}})$
$g : \mathbb{Y} \rightarrow \overline{\mathbb{R}}$	$g : \mathcal{M} \rightarrow \overline{\mathbb{R}}$
$\varphi : \mathbb{Y}' \times \mathbb{Y} \rightarrow \overline{\mathbb{R}}$	$\varphi(\underbrace{\mathbf{X}}_{\mathcal{F}\text{-measurable}}, m) = \mathbb{E}_m[\mathbf{X}]$
$\theta_{\mathbb{X}} : \mathbb{X} \rightarrow \mathbb{Y}'$	$\theta_{\mathbb{X}}(\mathbf{X}) = \sup_{k \in \mathcal{K}} \mathbb{E}_k[\mathbf{X}] + \theta_{\mathbb{Z}}(k)$
$\theta_{\mathbb{Z}} : \mathbb{Z} \rightarrow \mathbb{Y}'$	$\theta_{\mathbb{Z}} : \mathcal{K} \rightarrow L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \overline{\mathbb{R}})$
$\theta_{\mathbb{X} \times \mathbb{Z}} : \mathbb{X} \times \mathbb{Z} \rightarrow \mathbb{Y}'$	$\theta_{\mathbb{X} \times \mathbb{Z}}(\mathbf{X}, k) = \mathbb{E}_k[\mathbf{X}]$
$\Phi : \mathbb{X} \times (\mathbb{Y} \times \mathbb{Z}) \rightarrow \overline{\mathbb{R}}$	$\Phi(\mathbf{X}, m, k) = \mathbb{E}_m[\mathbb{E}_k[x]]$
$G : \mathbb{Y} \times \mathbb{Z} \rightarrow \overline{\mathbb{R}}$	$G(m, k) = g(m) + \mathbb{E}_m[\theta_{\mathbb{Z}}(k)]$

Table 4.1.: Application of Proposition 4.3.3 in the framework of risk measures.

**Illustration.** As an application of the formalism developed in Sect. 4.3.2, we show how some results of Theorem 11.17 in Föllmer and Schied (2016) can be interpreted in the framework of Proposition 4.3.3.

Example 4.3.6 provides Time Consistency and Nested Formula when convex risk measures are given with a “robust” formulation and uses the notion of normal integrand whose definition can be found in Rockafellar and Wets (1998) (Definition 14.27). Before entering the example, we denote by  $\leq$  the order defined on  $L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}})$  by

$$\mathbf{X} \leq \mathbf{Y} \Leftrightarrow \mathbf{X}(\omega) \leq_{\mathbb{P}\text{-a.s.}} \mathbf{Y}(\omega), \quad \mathbf{X}, \mathbf{Y} \in L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}})^2. \quad (4.65)$$

**Example 4.3.6.** Let a mapping  $g : \mathcal{M} \rightarrow [-\infty, +\infty]$  and a normal integrand  $\tilde{\theta}_{\mathbb{Z}} : \Delta(\omega) \times \Omega \rightarrow \overline{\mathbb{R}}$  be given. Let  $\theta_{\mathbb{Z}} : \mathcal{K} \rightarrow L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \overline{\mathbb{R}})$  be defined by

$$\theta_{\mathbb{Z}}(k)(\omega) = \tilde{\theta}_{\mathbb{Z}}(k(\omega), \omega), \quad \forall \omega \in \Omega. \quad (4.66)$$

Let  $(A, F)$  be the couple of mappings defined by

$$A : L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}}) \rightarrow [-\infty, +\infty],$$

$$\mathbf{X} \mapsto \sup_{m, k \in \mathcal{M} \times \mathcal{K}} \left\{ \mathbb{E}_m[\mathbb{E}_k[\mathbf{X}]] - g(m) - \mathbb{E}_m[\theta_{\mathbb{Z}}(k)] \right\}, \quad (4.67a)$$

$$F : L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}}) \rightarrow L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \overline{\mathbb{R}}),$$

$$\mathbf{X} \mapsto \sup_{k \in \mathcal{K}} \left\{ \mathbb{E}_k[\mathbf{X}] - \theta_{\mathbb{Z}}(k) \right\}, \quad (4.67b)$$

where the second supremum is taken with respect to the order  $\leq$  defined in (4.65) over the complete sup-semilattice  $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \overline{\mathbb{R}})$ . Then we have the Nested Formula

$$A = S^{A, F} \circ F, \quad (4.68)$$

with  $S^{A, F} : L^\infty(\Omega, \mathcal{T}, \mathbb{P}; \overline{\mathbb{R}}) \rightarrow \overline{\mathbb{R}}$  given by

$$S^{A, F}(\mathbf{X}) = \sup_{m \in \mathcal{M}} \left\{ \mathbb{E}_m[\mathbf{X}] - g(m) \right\}. \quad (4.69)$$



#### 4. Three classes of time consistent mappings

A particular case arises when the mappings  $G$ ,  $g$  and  $\theta_{\mathbb{Z}}$  in Table 4.1 are indicators mappings as in 4.3.2. In that case, the mappings  $A$ ,  $F$  and  $S^{A,F}$  are support functions. We refer to Proposition 5.1 in [Ruszczynski and Shapiro \(2006a\)](#) for an illustration.

*Proof.* We invite the reader to refer to Table 4.1 for the definition of the different elements. We insist on two points. First the mapping  $\theta_{\mathbb{X}}$  given by

$$\theta_{\mathbb{X}}(\mathbf{X}) = \sup_{k \in \mathcal{K}} \{ \mathbb{E}_k[\mathbf{X}] + \theta_{\mathbb{Z}}(k) \} \quad (4.70)$$

is well defined as the supremum is taken over the complete inf-semilattice  $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \bar{\mathbb{R}})$ .

Second, as the mapping  $\theta_{\mathbb{Z}}$  is a normal integrand, we can apply Theorem 14.60 in [Rockafellar and Wets \(1998\)](#) and Proposition 5.1 in [Ruszczynski and Shapiro \(2006a\)](#). Equation (4.41) is hence satisfied, i.e.

$$\sup_{k \in \mathcal{K}} \left\{ \mathbb{E}_m[\mathbb{E}_k[\mathbf{X}]] + \mathbb{E}_m[\theta_{\mathbb{Z}}(k)] \right\} = \mathbb{E}_m \left[ \sup_{k \in \mathcal{K}(\omega)} \{ \mathbb{E}_k[\mathbf{X}] + \tilde{\theta}_{\mathbb{Z}}(k, \omega) \} \right]. \quad (4.71)$$

We can now apply Proposition 4.3.3. This ends the proof.  $\square$

## 4.4. Time consistent supremum mappings

We investigate in this section how the property of time consistency between mappings as defined in Chap. 3 remains stable by supremum, i.e. we look for the conditions such that mappings that are defined as a supremum form a time consistent couple. To speak of supremum mappings, we use lattice theory.

We first provide basic tools of the Scott's topology, then focus on time consistency for mappings defined as a supremum.

### 4.4.1. Background on lattices and their topologies

We first enunciate two definitions and a lemma that will be used further. Here is the definition of a complete inf-semilattice.

**Definition 4.4.1.** A *complete inf-semilattice* is a preordered set such that there exists a greatest lower bound for each (finite or infinite) subset. We denote the preorder by  $\leq$  and the greatest lower bound by  $\bigwedge$ .

We now present continuity in Scott's topology.

**Definition 4.4.2.** Let  $\mathbb{X}$  and  $\mathbb{Y}$  be two complete inf-semilattices. A mapping  $f : \mathbb{X} \rightarrow \mathbb{Y}$  is said to be

- *lower semi continuous* (l.s.c.) if

$$f\left(\bigwedge_{x \in X} x\right) \leq \bigwedge_{x \in X} f(x), \quad \forall X \subset \mathbb{X}, \quad (4.72)$$

- *$\wedge$ -increasing* if  $x \leq x' \Rightarrow f(x) \leq f(x')$  or, equivalently, if

$$f(x \wedge x') \leq f(x) \wedge f(x'), \quad \forall (x, x') \in \mathbb{X}^2, \quad (4.73)$$

- *$\wedge$ -Scott continuous* if

$$f\left(\bigwedge_{x \in X} x\right) = \bigwedge_{x \in X} f(x), \quad \forall X \subset \mathbb{X}, \quad (4.74)$$

- *weakly  $\wedge$ -Scott continuous* if

$$f\left(\bigwedge_{x \in X} x\right) = \bigwedge_{x \in X} f(x), \quad \forall X \subset \mathbb{X}, \quad X \text{ complete inf semi-lattice.} \quad (4.75)$$

Reverse definitions of upper semi-continuity,  $\vee$ -increasing,  $\vee$ -Scott continuity and weak  $\vee$ -Scott continuity follow naturally.

If the previous statements are only valid on a subset  $X \subset \mathbb{X}$ , we speak of continuity on the set  $X$ .

The next lemma, whose proof is left to the reader, details implications between the different notions of continuity.

#### 4. Three classes of time consistent mappings

**Lemma 4.4.3.** *Let  $\mathbb{X}$  and  $\mathbb{Y}$  be two complete inf-semilattices and let  $f : \mathbb{X} \rightarrow \mathbb{Y}$  be a mapping. Then,*

- *if  $f$  is weakly  $\wedge$ -Scott continuous, then  $f$  is  $\wedge$ -Scott continuous,*
- *if  $f$  is  $\wedge$ -continuous,  $f$  is  $\wedge$ -increasing,*
- *if  $f$  is  $\wedge$ -increasing,  $f$  is lower semi continuous.*

In Sect. 4.4.2, we will study mappings defined as supremum. The next proposition reads as an interchange result for mappings that are continuous and defined as supremum.

**Proposition 4.4.4.** *Let  $\mathbb{X}$ ,  $\mathbb{Y}$  and  $\mathbb{Z}$  be three complete inf-semilattices and let  $M : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{Z}$  be a mapping. We define the marginal mapping  $M_{\mathbb{Y}} : \mathbb{X} \rightarrow \mathbb{Z}$  by*

$$M_{\mathbb{Y}}(x) = \bigvee_{y \in \mathbb{Y}} M(x, y) . \quad (4.76)$$

*If the mapping  $M_{\mathbb{Y}}$  is  $\wedge$ -Scott continuous, then we have the min-max interchange equation*

$$\bigwedge_{x \in \mathbb{X}} \bigvee_{y \in \mathbb{Y}} M(x, y) = \bigvee_{y \in \mathbb{Y}} M\left(\bigwedge_{x \in \mathbb{X}} x, y\right) . \quad (4.77)$$

*Proof.* If the mapping  $M_{\mathbb{Y}}$  is  $\wedge$ -Scott continuous, then we have by Definition 4.4.2 that

$$\bigwedge_{x \in \mathbb{X}} M_{\mathbb{Y}}(x) = M_{\mathbb{Y}}\left(\bigwedge_{x \in \mathbb{X}} x\right) . \quad (4.78)$$

Then by definition (4.76) of the mapping  $M_{\mathbb{Y}}$ , we obtain Equation (4.77). This ends the proof.  $\square$

We give an example of application of Proposition 4.4.4, that is, an example of mappings that are continuous and defined as supremum. This example relies on [Ruszczynski and Shapiro \(2006b\)](#).

**Example 4.4.5.** Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a finite probability space such that  $|\Omega| = n$  and let  $\mathcal{X} = L^\infty(\Omega, \mathcal{T}, \mathbb{P})$  be the space of bounded random variables on  $\Omega$ . We denote by  $\Delta(\Omega)$  the set of all probability distributions. Let  $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$  be a numerical function and let  $F : \mathbb{R}^n \rightarrow \mathcal{X}$  be a mapping. For a subset  $S \subset \mathbb{R}$  of real numbers and a subset  $\mathcal{M} \subset \mathcal{X}$  of random variables, we introduce the set  $\mathcal{M}_S$  defined by

$$\mathcal{M}_S = \{\mathbf{Z} \in \mathcal{M} \mid \mathbf{Z}(\omega) \in S, \forall \omega \in \Omega\} . \quad (4.79)$$

Under technical assumptions detailed in Theorem 7.1 of [Ruszczynski and Shapiro \(2006b\)](#), we have the following result. If  $\rho$  is a monotonous risk function <sup>1</sup>, then

$$\rho\left(\inf_{z \in S} F(z)\right) = \inf_{\mathbf{Z} \in \mathcal{M}_S} \rho(F(\mathbf{Z})) . \quad (4.80)$$

---

<sup>1</sup>See [Ruszczynski and Shapiro \(2006b\)](#) for details on what is a monotonous risk function

#### 4.4. Time consistent supremum mappings

In particular, Equation (4.80) is satisfied for functions  $\rho$  that are given by

$$\rho(\mathbf{Z}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \langle \mathbb{Q}, \mathbf{Z} \rangle, \quad \forall \mathbf{Z} \in \mathcal{M}, \quad (4.81)$$

where  $\mathcal{Q} \subset \Delta(\Omega)$  is a subset of probability measures.

In Equation (4.80), the inner infimum to the left hand side is taken component per component. The space  $\mathcal{X} = L^\infty(\Omega, \mathcal{T}, \mathbb{P})$  is then naturally endowed with a complete inf-semilattice structure as an infinite product space. Following the notations of Proposition 4.4.4 we hence rewrite Equation (4.80)

$$\rho\left(\bigwedge_{\mathbf{Z} \in \mathcal{M}_S} F(\mathbf{Z})\right) = \bigwedge_{\mathbf{Z} \in \mathcal{M}_S} \rho(F(\mathbf{Z})). \quad (4.82)$$

#### 4.4.2. Usual time consistency with mappings defined with supremum

Let  $\mathbb{H}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{A}, \mathbb{F}$  be six sets, and  $A, F$  be two mappings:

$$A : \mathbb{H} \times \mathbb{T} \times \mathbb{U} \times \mathbb{V} \rightarrow \mathbb{A}, \quad (4.83a)$$

$$F : \mathbb{T} \times \mathbb{V} \rightarrow \mathbb{F}. \quad (4.83b)$$

We want to study how time consistency is inherited after a partial supremum operation leading to marginal mappings. For this purpose, we first characterize time consistency for the couple of mappings  $(A, F)$  given in (4.83) in terms of the subaggregator  $S^{A,F}$  defined in (3.2).

**Proposition 4.4.6.** *The couple  $(A, F)$  in (4.83) is Weak Time Consistent if and only if the set-valued mapping  $S^{A,F}$  defined by*

$$S^{A,F}(h, u, f) = \{A(h, t, u, v) \text{ such that } F(t, v) = f\}, \quad (4.84)$$

*is a mapping increasing in its third argument. We then have the nested formula*

$$A(h, t, u, v) = S^{A,F}(h, u, F(t, v)). \quad (4.85)$$

Applying the result of Proposition 3.3.3 replacing  $\mathbb{H}$  by  $\mathbb{H} \times \mathbb{U}$  and  $\mathbb{T}$  by  $\mathbb{T} \times \mathbb{V}$ , we get that

We now assume that the sets  $\mathbb{A}$  and  $\mathbb{F}$  are equipped with orders denoted by  $\leq$  and are complete sup-semilattices. We define the mappings  $A_{\mathbb{W}} : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$  and  $F_{\mathbb{V}} : \mathbb{T} \rightarrow \mathbb{F}$  by

$$F_{\mathbb{V}}(t) = \bigvee_{v \in \mathbb{V}} F(t, v), \quad (4.86a)$$

$$A_{\mathbb{W}}(h, t) = \bigvee_{(u,v) \in \mathbb{W}} A(h, t, u, v), \quad (4.86b)$$

where  $\mathbb{W}$  is a subset of the product set  $\mathbb{U} \times \mathbb{V}$ , i.e.

$$\mathbb{W} \subset \mathbb{U} \times \mathbb{V}. \quad (4.86c)$$

#### 4. Three classes of time consistent mappings

**Proposition 4.4.7.** *Let us assume that the couple  $(A, F)$  is Usual Time Consistent, that is,*

$$\begin{aligned} F(t, v) \leq F(t', v') &\Rightarrow A(h, t, u, v) \leq A(h, t', u, v'), \\ \forall (h, u) \in \mathbb{H} \times \mathbb{U}, \quad \forall (t, t', v, v') &\in \mathbb{T}^2 \times \mathbb{V}^2. \end{aligned} \quad (4.87)$$

Let us assume that the set  $\mathbb{W}$  in (4.86c) is such that

$$\begin{aligned} \bigvee_{(u,v) \in \mathbb{W}} A(h, t, u, v) &= \bigvee_{u \in \mathbb{U}} \bigvee_{v \in \mathbb{V}} A(h, t, u, v), \\ \forall (h, t) \in \mathbb{H} \times \mathbb{T} \end{aligned} \quad (4.88)$$

Let also assume that one of the following conditions is satisfied

- either the mappings  $A(h, t, u, \cdot) : \mathbb{V} \rightarrow \mathbb{A}$  and  $F(t, \cdot) : \mathbb{V} \rightarrow \mathbb{F}$  are  $\vee$ -Scott continuous on the set  $\mathbb{V}$  for all  $(h, t, u) \in \mathbb{H} \times \mathbb{T} \times \mathbb{U}$ ,
- or, for all  $t \in \mathbb{T}$ , there exists  $v_t^* \in \mathbb{V}$  such that

$$\bigvee_{v \in \mathbb{V}} F(t, v) = F(t, v_t^*). \quad (4.89)$$

Then the couple  $(A_{\mathbb{W}}, F_{\mathbb{V}})$  in (4.86) is Usual Time Consistent and the subaggregator introduced in (3.2) is given by

$$S^{A_{\mathbb{W}}, F_{\mathbb{V}}}(h, f) = \bigvee_{u \in \mathbb{U}} S^{A, F}(h, u, f). \quad (4.90)$$

Before stating the proof, we notice that Equation (4.88) is satisfied when  $\mathbb{W} = \mathbb{U} \times \mathbb{V}$ .

*Proof.* The proof goes as follows:

1. we first show that the couple  $(A_{\mathbb{W}}, F_{\mathbb{V}})$  in (4.86) is Usual Time Consistent,
2. then we explicit the form of the subaggregator  $S^{A_{\mathbb{W}}, F_{\mathbb{V}}}$  in (4.90).

Here is the proof.

1. We first assume that the mappings  $A(h, t, u, \cdot) : \mathbb{V} \rightarrow \mathbb{A}$  and  $F(t, \cdot) : \mathbb{V} \rightarrow \mathbb{F}$  are  $\vee$ -Scott continuous on the set  $\mathbb{V}$  for all  $(h, t, u) \in \mathbb{H} \times \mathbb{T} \times \mathbb{U}$ . Let  $(t, t') \in \mathbb{T}^2$ . Then we have the successive equivalence and implications

$$\begin{aligned} F_{\mathbb{V}}(t) \geq F_{\mathbb{V}}(t') &\Leftrightarrow \bigvee_{v \in \mathbb{V}} F(t, v) \geq \bigvee_{v \in \mathbb{V}} F(t', v), \quad (\text{by definition (4.86) of } F_{\mathbb{V}}) \\ &\Leftrightarrow F(t, \bigvee_{v \in \mathbb{V}} v) \geq F(t', \bigvee_{v \in \mathbb{V}} v), \end{aligned}$$

by  $\vee$ -Scott continuity of the mapping  $F(t, \cdot) : \mathbb{V} \rightarrow \mathbb{F}$  for all  $t \in \mathbb{T}$ ,

$$\Rightarrow A(h, t, u, \bigvee_{v \in \mathbb{V}} v) \geq A(h, t', u, \bigvee_{v \in \mathbb{V}} v), \quad \forall (h, u) \in \mathbb{H} \times \mathbb{U},$$

#### 4.4. Time consistent supremum mappings

by UTC of the couple  $(A, F)$  in (4.83),

$$\Leftrightarrow \bigvee_{v \in \mathbb{V}} A(h, t, u, v) \geq \bigvee_{v \in \mathbb{V}} A(h, t', u, v), \quad \forall (h, u) \in \mathbb{H} \times \mathbb{U},$$

by  $\vee$ -Scott continuity of the mapping  $A(h, t, u, \cdot) : \mathbb{V} \rightarrow \mathbb{A}$  for all  $(h, t, u) \in \mathbb{H} \times \mathbb{T} \times \mathbb{U}$ .

We now assume that Equation (4.89) is satisfied. In that case, we have the successive equivalence and implications

$$\begin{aligned} F_{\mathbb{V}}(t) \geq F_{\mathbb{V}}(t') &\Leftrightarrow \bigvee_{v \in \mathbb{V}} F(t, v) \geq \bigvee_{v \in \mathbb{V}} F(t', v), && \text{(by definition (4.86) of } F_{\mathbb{V}}) \\ &\Rightarrow \bigvee_{v \in \mathbb{V}} F(t, v) \geq F(t', \tilde{v}), \quad \forall \tilde{v} \in \mathbb{V}, \end{aligned}$$

by definition of the supremum of  $F(t', v)$

$$\Rightarrow F(t, v_t^*) \geq F(t', \tilde{v}), \quad \forall \tilde{v} \in \mathbb{V},$$

by (4.89) that states that the supremum is attained,

$$\Rightarrow A(h, t, u, v_t^*) \geq A(h, t', u, \tilde{v}), \quad \forall \tilde{v} \in \mathbb{V}, \quad \forall (h, u) \in \mathbb{H} \times \mathbb{U},$$

by UTC of the couple  $(A, F)$  in (4.83),

$$\Rightarrow \bigvee_{v \in \mathbb{V}} A(h, t, u, v) \geq A(h, t, u, v_t^*) \geq A(h, t', u, \tilde{v}), \quad \forall \tilde{v} \in \mathbb{V}, \quad \forall (h, u) \in \mathbb{H} \times \mathbb{U},$$

by definition of the supremum on  $\mathbb{V}$ .

With one or the other assumption ( $\vee$ -Scott continuity of mappings  $A|_{\mathbb{V}}$  and  $F|_{\mathbb{V}}$  of Equation (4.89) is satisfied), we have that

$$\bigvee_{v \in \mathbb{V}} A(h, t, \tilde{u}, v) \geq A(h, t', \tilde{u}, \tilde{v}) \Rightarrow \bigvee_{u \in \mathbb{V}} \bigvee_{v \in \mathbb{V}} A(h, t, u, v) \geq A(h, t', \tilde{u}, \tilde{v}),$$

for all  $\tilde{v} \in \mathbb{V}$ , by definition of the supremum over  $\mathbb{U}$ ,

$$\Rightarrow \bigvee_{u \in \mathbb{V}} \bigvee_{v \in \mathbb{V}} A(h, t, u, v) \geq \bigvee_{u \in \mathbb{V}} \bigvee_{v \in \mathbb{V}} A(h, t', u, v), \quad (4.91)$$

by definition of the supremum over  $\mathbb{U}$ ,

$$\begin{aligned} &\Rightarrow \bigvee_{(u,v) \in \mathbb{W}} A(h, t, u, v) \geq \bigvee_{(u,v) \in \mathbb{W}} A(h, t', u, v), \\ &\Rightarrow A_{\mathbb{W}}(h, t) \geq A_{\mathbb{W}}(h, t'), \end{aligned} \quad \text{(by Eq. (4.88))} \quad (4.92)$$

by definition (4.88) of  $A_{\mathbb{W}}$  This shows that the couple  $(A, F)$  in (4.83) is Usual Time Consistent.

#### 4. Three classes of time consistent mappings

2. As the couple  $(A, F)$  is Usual Time Consistent, we know from Proposition 3.3.6 that

$$A(h, t, u, v) = S^{A,F}(h, u, F(t, v)) , \quad (4.93)$$

where the mapping  $S^{A,F}$  given in (4.84) is increasing in its third argument.

If Equation (4.89) is satisfied, we have for all  $(h, t, v) \in \mathbb{H} \times \mathbb{T} \times \mathbb{V}$  that

$$\begin{aligned} F(t, v^*) &\geq F(t, v) , && \text{(by definition of the supremum)} \\ S^{A,F}(h, F(t, v^*)) &\geq S^{A,F}(h, F(t, v)) , && \text{(as the mapping } S^{A,F} \text{ is increasing)} \\ S^{A,F}(h, F(t, v^*)) &\geq \bigvee_{v \in \mathbb{V}} S^{A,F}(h, F(t, v)) , && \text{(by definition of the supremum.)} \end{aligned}$$

As  $v^* \in \mathbb{V}$ , we have by definition of the supremum over  $\mathbb{V}$  the reverse inequality  $\bigvee_{v \in \mathbb{V}} S^{A,F}(h, F(t, v)) \geq S^{A,F}(h, F(t, v^*))$ . Hence, we have the equality

$$\bigvee_{v \in \mathbb{V}} S^{A,F}(h, F(t, v)) = S^{A,F}(h, F(t, v^*)) . \quad (4.94)$$

Then we have the following equalities

$$\begin{aligned} A_{\mathbb{W}}(h, t) &= \bigvee_{(u,v) \in \mathbb{W}} A(h, t, u, v) , && \text{(by Equation (4.86))} \\ &= \bigvee_{u \in \mathbb{U}} \bigvee_{v \in \mathbb{V}} A(h, t, u, v) , && \text{(by Equation (4.88))} \\ &= \bigvee_{u \in \mathbb{U}} \bigvee_{v \in \mathbb{V}} S^{A,F}(h, u, F(t, v)) , && \text{(by Equation (4.85))} \\ &= \bigvee_{u \in \mathbb{U}} S^{A,F}(h, u, \bigvee_{v \in \mathbb{V}} F(t, v)) , && \text{(by Equation (4.94))} \\ &= \bigvee_{u \in \mathbb{U}} S^{A,F}(h, u, F_{\mathbb{V}}(t)) , && \text{(by Equation (4.86))} \\ &= S^{A_{\mathbb{W}}, F_{\mathbb{V}}}(h, F_{\mathbb{V}}(t)) . && \text{(by definition (4.90))} \end{aligned}$$

This ends the proof.  $\square$

Proposition (4.4.7) covers the framework of coherent risk measures. Indeed, let  $(\Omega, \mathcal{T}, \mathbb{P})$  be a probability space and let  $L^\infty(\Omega, \mathcal{T}, \mathbb{P})$  be the space of bounded random variables. Let also  $\Delta(\Omega, \mathcal{T})$  be the set of probability distributions. Then a coherent risk measure  $\rho : L^\infty(\Omega, \mathcal{T}, \mathbb{P}) \rightarrow \mathbb{R}$  has the form

$$\rho(\mathbf{X}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] , \quad (4.95)$$

where the  $\mathcal{Q} \subset \Delta(\Omega, \mathcal{T})$  is a set of probability distributions.

## 4.5. Conclusion and perspectives

We have seen in Chap. 3 that Time Consistency and Nested Formula are equivalent. In this Chap. 4, we have studied three classes of mappings and described how time consistency is characterized or inherited for these classes.

- First, we have characterized time consistency in terms of acceptance sets for translation invariant mappings. This characterization, which involves two unknown mappings  $\rho$  and  $F$ , is an implicit equation in  $\rho$ . Hence it remains difficult to apply in practice to find interesting results.
- Second, we have given conditions for mappings that are Fenchel-Moreau transform to be time consistent. These Fenchel-Moreau mappings are defined with respect to couplings and are hence generalized convex mappings. When couplings are duality pairings between vector spaces, the Fenchel-Moreau's transform provides convex mappings. We have applied our result to convex risk measures. However, the conditions that we obtain for time consistency for generalized convex mappings are strong. It is not an easy task to exhibit generalized convex mappings that are time consistent (and not convex risk measures).
- Third, we have studied how time consistency is inherited when mappings are defined as supremum.

Time consistency for mappings remains a world to explore. Without being exhaustive, we detail work that could be done.

- Looking at the conditions that we have obtained on generalized convex mappings to obtain a nested formula, we see that these conditions are difficult to satisfy. Are there mappings that are not convex risk measures but satisfy these conditions ? The possibility of mixing translation invariant mappings, Fenchel-Moreau transform, supremum to construct time consistent mappings has not been investigated yet.
- We did not have the time to discuss the connection between the inheritance of time consistency presented in Sect. 4.4 and the recent work of Dentcheva and Ruszczyński on risk forms [Dentcheva and Ruszczyński \(2018\)](#).
- We were not able to exhibit examples for which the characterization in terms of acceptance sets could give some explicit results.





# 5. Player consistency and decomposition of optimization problems

## Contents

---

<b>5.1. Introduction</b>	<b>90</b>
<b>5.2. From time consistency to player consistency</b>	<b>90</b>
5.2.1. Consistency with binary relations	90
5.2.2. Player consistency for mappings	91
5.2.3. Time consistency as a special case of player consistency	93
5.2.4. Application of player consistency to translation invariant mappings	95
<b>5.3. Application of player consistency to decomposition for optimization</b>	<b>96</b>
5.3.1. Sequential decomposition with time consistency	96
5.3.2. Parallel optimization with player consistency	99
5.3.3. Mixing parallel and sequential optimization	100
<b>5.4. Conclusion</b>	<b>102</b>

---

## 5.1. Introduction

We have seen in Chap. 3 that Time Consistency and Nested Formula are equivalent. In Chap. 4, we have studied time consistency for three classes of mappings. We have discussed translation invariant mappings, mappings that are Fenchel-Moreau's transform and mappings that are defined as supremum. In this Chapter 5, we first generalize time consistency to player consistency. Then we will sketch how player consistency can be used to decompose optimization problems.

## 5.2. From time consistency to player consistency

For authors who study risk and optimization, time consistency is an ingredient to decompose optimization problems using a nested formula. We recall the example given in introduction in the risk neutral case as an illustration

$$\min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}, \mathbf{X}_2 \preceq_{\mathcal{F}}} \mathbb{E}_{\mathbb{P}}[c(\mathbf{X}_1, \mathbf{X}_2)] = \min_{\mathbf{X}_1 \preceq_{\mathcal{F}_1}} \mathbb{E}_{\mathbb{P}} \left[ \min_{x_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}}[c(\mathbf{X}_1, x_2) \mid \mathcal{F}] \right], \quad (5.1)$$

where  $\mathcal{F}_1 \subset \mathcal{F}_2$ . Time consistency can also be seen as a natural property for an optimal control as “stick to one's plan” moto. In this section, we introduce the notion of consistency without relying on the notion of time.

### 5.2.1. Consistency with binary relations

In Chap. 3, we defined and characterized time consistency for couples of mappings. We now explore a general framework for consistency with binary relations and show that it includes the case of time consistency. We will first provide definitions and then recover results of Chap. 3.

We define consistency over product sets, so we invite the reader to refer to Appendix 9.1 for the definition of products of binary relations.

**Definition 5.2.1.** Let  $P$  be a finite set (of players) and let  $(\mathbb{S}_p, \mathcal{R}_p)_{p \in P}$  be a collection of pairs of sets and relations. Let  $\mathcal{R}$  be a relation on the set  $\prod_{p \in P} \mathbb{S}_p$ . We hence have

$$\mathbb{S}_p \subset \mathbb{S}_p^2 \text{ and } \mathcal{R} \subset \left( \prod_{p \in P} \mathbb{S}_p \right)^2. \quad (5.2)$$

The tuple of relations  $(\mathcal{R}_p)_{p \in P}$  is said to be  $\mathcal{R}$ -consistent if

$$\prod_{p \in P} \mathbb{S}_p \subset \mathcal{R}. \quad (5.3)$$

When the tuple  $(\mathcal{R}_p)_{p \in P}$  is  $\mathcal{R}$ -consistent, we say that the relations  $(\mathbb{S}_p)_{p \in P}$  are *factors* of the relation  $\mathcal{R}$ .

We discuss relation of player consistency with the framework of time consistency presented in Chap. 3 in “Discussion on time consistency” below.

### 5.2.2. Player consistency for mappings

In what follows, we use the abuse of notation

$$(s_p)_{p \in P} = s_P, \quad \prod_{p \in P} \mathbb{S}_p = \mathbb{S}_P. \quad (5.4)$$

**Definition 5.2.2.** Let  $P$  be a finite set (of players), and  $\mathbb{A}$  be a set. We consider two collections of sets  $(\mathbb{S}_p)_{p \in P}$  and  $(\mathbb{F}_p)_{p \in P}$  and a collection of mappings  $(A, (F_p)_{p \in P})$  such that

$$A : \prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A}, \quad (5.5a)$$

$$F_p : \mathbb{S}_p \rightarrow \mathbb{F}_p, \quad \forall p \in P. \quad (5.5b)$$

The tuple  $(A, (F_p)_{p \in P})$  of mappings is said to be *Weakly Player Consistent* (WPC) if

$$\left( \forall p \in P, F_p(s_p) = F_p(s'_p) \right) \Rightarrow A(s_P) = A(s'_P). \quad (5.6)$$

and is said to be *Strongly Monotone Player Consistent* (SMPC) if

$$\left( \forall p \in P, F_p(s_p) \leq F_p(s'_p) \right) \Rightarrow A(s_P) \leq A(s'_P). \quad (5.7)$$

**Lemma 5.2.3.** *Weak Player Consistency and Strong Monotone Player Consistency (Definition 5.2.2) are particular cases of  $\mathcal{R}$ -consistency (Definition 5.2.1).*

*Proof.* Let us introduce the relation  $\mathcal{R}$  and the collection of relations  $\mathbb{S}_P$  defined by

$$s_P \mathcal{R} s'_P \Leftrightarrow A(s_P) = A(s'_P), \quad (5.8a)$$

$$s_p \mathbb{S}_p s'_p \Leftrightarrow F_p(s_p) = F_p(s'_p), \quad \forall p \in P. \quad (5.8b)$$

If we replace equalities  $=$  by inequalities  $\leq$  in Equation (5.8), we obtain the definition of Strongly Monotone Player Consistency.  $\square$

As for time consistency discussed in Chap. 3, we introduce a subaggregator.

**Definition 5.2.4.** Let  $P$  be a set (of players), and  $\mathbb{A}$  be a set. We consider two collections  $(\mathbb{S}_p)_{p \in P}$  of sets and  $(\mathbb{F}_p)_{p \in P}$  and a collection  $(A, (F_p)_{p \in P})$  of mappings such that

$$A : \prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A}, \quad (5.9a)$$

$$F_p : \mathbb{S}_p \rightarrow \mathbb{F}_p, \quad \forall p \in P. \quad (5.9b)$$

With the tuple  $(A, (F_p)_{p \in P})$ , we associate the set-valued mapping

$$\begin{aligned} S^{A,F} : \prod_{p \in P} \text{Im}(F_p) &\rightrightarrows \mathbb{A} \\ (f)_{p \in P} = f_P &\mapsto \left\{ A(s_P) \mid s_p \in F_p^{-1}(f_p), \forall p \in P \right\}. \end{aligned} \quad (5.10)$$

If the tuple  $(A, (F_p)_{p \in P})$  is WPC as in Definition 5.2.2, then  $S^{A,F}$  is called a *subaggregator*.

## 5. Player consistency and decomposition of optimization problems

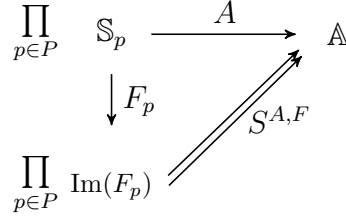


Figure 5.1.: Representation of connections between the mappings  $A$ ,  $S^{A,F}$  and  $((F_p)_{p \in P})$  of Definition 5.2.4.

To fix ideas, we illustrate the relations between the different sets and mappings in Fig. 5.1.

The next Theorem characterizes Weak Player Consistency in Definition 5.2.2 in terms of subaggregator. The proof is left to the reader as a generalization of Theorem 3.3.3.

**Theorem 5.2.5.** *[Nested decomposition of WPC mappings] The tuple  $(A, (F_p)_{p \in P})$  is WPC if and only if the set valued mapping  $S^{A,F}$  in (5.10) is a mapping. In that case, we have the nested formula between mappings.*

$$A(s_P) = S^{A,F} \left( (F_p(s_p))_{p \in P} \right), \quad \forall s_P \in \mathbb{S}_P. \quad (5.11)$$

This kind of nested formula will prove useful when we discuss decomposition in 5.3.

**Mixing weak and strong monotone player consistency** As a step towards decomposition, we refine the notion of Weak Player Consistency and Strong Monotone Player Consistency.

**Definition 5.2.6.** Let  $P$  be a set (of players), and  $\mathbb{A}$  be a set. We consider two collections of sets  $(\mathbb{S}_p)_{p \in P}$  and  $(\mathbb{F}_p)_{p \in P}$  and a collection of mappings  $(A, (F_p)_{p \in P})$  such that

$$A : \prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A}, \quad (5.12a)$$

$$F_p : \mathbb{S}_p \rightarrow \mathbb{F}_p, \quad \forall p \in P. \quad (5.12b)$$

Let  $(P^=, P^{\leq})$  be a partition of the set  $P$ . The tuple  $(A, (F_p)_{p \in P})$  is said to be  $(P^=, P^{\leq})$ -Mixed Player Consistent if have

$$\left. \begin{array}{l} \forall p \in P^=, F_p(s_p) = F_p(s'_p) \\ \forall p \in P^{\leq}, F_p(s_p) \leq F_p(s'_p) \end{array} \right\} \Rightarrow A(s_P) \leq A(s'_P). \quad (5.13)$$

**Proposition 5.2.7.** *Let  $(A, (F_p)_{p \in P})$  be a tuple of mappings as in (5.12). The tuple  $(A, (F_p)_{p \in P})$  is Mixed Player Consistent if and only if the subaggregator  $S^{A,F}$  defined in (5.10) is a mapping that is increasing over the set  $P^{\leq}$ .*

This property will prove to be useful for optimization. We finish this section 5.2 by showing how we can extend results on time consistency and translation invariant mappings to the case of player consistency.

### 5.2.3. Time consistency as a special case of player consistency

We now illustrate how time consistency presented in Chap. 3 can be seen as a particular case of player consistency as defined in Definition 5.2.2.

We consider four sets  $\mathbb{H}$ ,  $\mathbb{T}$ ,  $\mathbb{A}$  and  $\mathbb{F}$  and two mappings  $A$  and  $F$  given by

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A} , \quad (5.14)$$

$$F : \mathbb{T} \rightarrow \mathbb{F} . \quad (5.15)$$

Let  $P$ ,  $\mathbb{S}_1$ ,  $\mathbb{S}_2$ ,  $\mathbb{F}_1$  and  $\mathbb{F}_2$  be sets defined by

$$P = \{1, 2\} , \quad \mathbb{S}_1 = \mathbb{H} , \quad \mathbb{S}_2 = \mathbb{T} , \quad \mathbb{F}_1 = \mathbb{H} , \quad \mathbb{F}_2 = \mathbb{F} . \quad (5.16)$$

We define the mapping  $F_1$  by

$$F_1 = \text{Id} . \quad (5.17)$$

Then studying time consistency of a couple of mappings  $(A, F)$  is equivalent to study player consistency of the triplet  $(A, (F_1, F))$  defined in (5.14) and in (5.17). Indeed, time consistency in this example reads

$$F(t) = F(t') \Rightarrow A(h, t) = A(h, t') , \quad \forall h \in \mathbb{H} , \quad \forall (t, t') \in \mathbb{T}^2 , \quad (5.18)$$

and player consistency reads

$$\left. \begin{array}{l} \text{Id}(h) = \text{Id}(h') \\ F(t) = F(t') \end{array} \right\} \Rightarrow A(h, t) = A(h', t') , \quad \forall (h, h', t, t') \in \mathbb{H}^2 \times \mathbb{T}^2 . \quad (5.19)$$

All cases and results treated in Chap. 3 and Sect. 5.2.2 are summarized in Table 5.1

	$(\mathbb{S}_p)_{p \in P}$	$\prod_{p \in P} \mathbb{S}_p$	$A$	$F_p$	$\mathbb{S}_p$	$\mathcal{R}$	Characterization in terms of subaggregator	Characterization in terms of partitions or level sets
WTC	$\mathbb{H}, \mathbb{T}$	$\mathbb{H} \times \mathbb{T}$	$\mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$	$\mathbb{T} \rightarrow \mathbb{F}$	$F(t) = F(t')$	$A(h, t) = A(h, t')$	mapping	$\bigvee_{h \in \mathbb{H}} \mathbb{T}/A_h$ $\wr$ $\mathbb{T}/F$
UTC	$\mathbb{H}, \mathbb{T}$	$\mathbb{H} \times \mathbb{T}$	$\mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$	$\mathbb{T} \rightarrow \mathbb{F}$	$F(t) \leq F(t')$	$A(h, t) \leq A(h, t')$	mapping increasing in second argument	$F_{\leq F(t')}$ $\cap$ $\bigcap_{h \in \mathbb{H}} A_{h \leq A_h(t')}$
STC	$\mathbb{H}, \mathbb{T}$	$\mathbb{H} \times \mathbb{T}$	$\mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$	$\mathbb{T} \rightarrow \mathbb{F}$	$h \leq h'$ $F(t) \leq F(t')$	$A(h, t) \leq A(h', t')$	mapping increasing in both ar- guments	$F_{\leq F(t')}$ $\cap$ $\bigcap_{h \leq h'} A_{h \leq A_{h'}(t')}$
WPC	$\mathbb{S}_p$	$\prod_{p \in P} \mathbb{S}_p$	$\prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A}$	$\mathbb{S}_p \rightarrow \mathbb{F}_p$	$F_p(s_p) = F_p(s'_p)$	$A((s_p)_{p \in P})$ $=$ $A((s'_p)_{p \in P})$	mapping	$(\prod_{p \in P} \mathbb{S}_p) / A$ $\wr$ $(\prod_{p \in P} \mathbb{S}_p / F_p)$
SAC	$\mathbb{S}_p$	$\prod_{p \in P} \mathbb{S}_p$	$\prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A}$	$\mathbb{S}_p \rightarrow \mathbb{F}_p$	$F_p(s_p) \leq F_p(s'_p)$	$A((s_p)_{p \in P})$ $\wedge$ $A((s'_p)_{p \in P})$	mapping increasing in all argu- ments	$\prod_{p \in P} F_{p \leq F_p(s_p)}$ $\cap$ $A_{\leq A((s_p)_{p \in P})}$
Consistency	$\mathbb{S}_p$	$\prod_{p \in P} \mathbb{S}_p$	$\emptyset$	$\emptyset$	$\mathbb{S}_p$	$\mathcal{R}$	$\emptyset$	$\prod_{p \in P} \mathbb{S}_p$ $\cap$ $\mathcal{R}$

Table 5.1.: Recovering Time Consistency and Player Consistency with  $\mathcal{R}$ -consistency.

### 5.2.4. Application of player consistency to translation invariant mappings

We observe that time consistency results extend quite naturally to player consistency. The next proposition is left to the reader as generalization of Proposition 4.2.4.

**Proposition 5.2.8.** *Let  $(\mathbb{H}, \oplus)$  and  $(\mathbb{T}, \oplus)$  be two commutative groups. Let  $(\mathbb{A}, \oplus)$  be a group such that*

$$(\mathbb{A}, \oplus) \subset (\mathbb{H}, \oplus) \times (\mathbb{T}, \oplus) , \quad (5.20)$$

and let

$$A : \mathbb{H} \times \mathbb{T} \rightarrow (\mathbb{A}, \oplus, \leq) , \quad (5.21)$$

$$(h, t) \mapsto A(h, t) \quad (5.22)$$

be a  $(\mathbb{H} \times \mathbb{T}, \mathbb{A})$ -translation invariant mapping. We consider two subgroups

$$(\mathbb{G}, \oplus) \subset (\mathbb{H}, \oplus) \text{ and } (\mathbb{F}, \oplus) \subset (\mathbb{T}, \oplus) . \quad (5.23)$$

Let  $G : \mathbb{H} \rightarrow (\mathbb{G}, \oplus, \leq)$  be a  $(\mathbb{H}, \mathbb{G})$ -translation invariant mapping and  $F : \mathbb{T} \rightarrow (\mathbb{F}, \oplus, \leq)$  be a  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping. Then,

1. the triplet  $(A, (G, F))$  is Player Consistent if and only if

$$(\mathcal{A}_G \oplus G(h)) \times (\mathcal{A}_F \oplus F(t)) \subset \mathcal{A}_\rho \oplus \rho(h, t) , \quad \forall (h \times t) \in \mathbb{H} \times \mathbb{T} . \quad (5.24)$$

2. If, in addition, we have that

- $(\mathbb{A}, \oplus) \subset (\mathbb{G} \times \mathbb{F}, \oplus) \subset (\mathbb{H} \times \mathbb{T}, \oplus)$ ,
- the  $(\mathbb{T}, \mathbb{A})$ -translation invariant mapping  $\rho$  is increasing,
- the  $(\mathbb{H}, \mathbb{G})$ -translation invariant mapping  $G$  satisfies  $G(0) = 0$  (where 0 is the neutral element of  $(\mathbb{H}, \oplus)$ ),
- the  $(\mathbb{T}, \mathbb{F})$ -translation invariant mapping  $F$  satisfies  $F(0) = 0$  (where 0 is the neutral element of  $(\mathbb{T}, \oplus)$ ),

then the tuple  $(A, (G, F))$  of mappings is Usual Time Consistent if and only if

$$(\mathcal{A}_G \times \mathcal{A}_F) \oplus \mathcal{A}_{\rho|_{\mathbb{G} \times \mathbb{F}}} = \mathcal{A}_\rho . \quad (5.25)$$

After having generalized some results from time consistency to player consistency, we now outline how they can be used in optimization, especially for decomposition purpose.



### 5.3. Application of player consistency to decomposition for optimization

In Chap. 3, we have defined the notion of Time Consistency for mappings. We have then discussed in Chap. 4 the notion of Time Consistency for classes of mappings. Finally, in Sect. 5.2 we have highlighted the fact that Time Consistency is a particular case of Player Consistency. Now that we have presented frameworks of Time Consistency and Player Consistency, we want to highlight how these tools can be used in optimization.

In Chap. 3, we have shown an equivalence between Nested Formula and Time Consistency. In Chapter 5.2, we have extended this equivalence between a Nested Formula and Player Consistency. We explore in § 5.3.1 and in § 5.3.2 what minimal assumptions are required to decompose an optimization problem thanks to a Nested Formula.

As we will use notion on lattices, we refer the reader to § 4.4.1 for background on complete inf-semilattices (Definitions 4.4.1) and Scott-continuity (Definition 4.4.2).

First, we show how time consistency is adapted to sequential decomposition, then how player consistency is adapted to parallel decomposition. Second, we detail how mixed player consistency offers the most general perspective for sequential and parallel decomposition.

#### 5.3.1. Sequential decomposition with time consistency

We start to study an optimization problem under a Time Consistency property and we show how we can decompose the problem recursively. We first recall the main ingredients. Let  $\mathbb{H}$ ,  $\mathbb{T}$ ,  $\mathbb{A}$  and  $\mathbb{F}$  be four sets, and let  $A$  and  $F$  be two mappings as follows:

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A} , \quad F : \mathbb{T} \rightarrow \mathbb{F} . \quad (5.26)$$

The following proposition states how we can operate optimization sequentially.

**Theorem 5.3.1** (Sequential optimization). *Assume that the sets  $\mathbb{A}$  and  $\mathbb{F}$  in (5.26) are equipped with orders denoted by  $\leq$  and are complete inf-semilattices. Let the couple  $(A, F)$  in (5.26) be Usual Time Consistent, so that, by Proposition 3.3.6, there exists a mapping  $S^{A,F} : \mathbb{H} \times \text{Im}(F) \rightarrow \mathbb{A}$  increasing in its second argument such that*

$$A(h, t) = S^{A,F}(h, F(t)) , \quad \forall (h, t) \in \mathbb{H} \times \mathbb{T} . \quad (5.27)$$

Then, we have the “sequential programming inequality”

$$\bigwedge_{h \in \mathbb{H}} S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) \leq \bigwedge_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h, t) . \quad (5.28)$$

If, in addition, one of the following conditions is satisfied

- i)  $\text{Im}(F)$  is a complete inf-semilattice and the mapping  $S^{A,F}$  is  $\wedge$ -Scott continuous,

### 5.3. Application of player consistency to decomposition for optimization

ii) there exists  $t^* \in \mathbb{T}$  such that

$$\bigwedge_{t \in \mathbb{T}} F(t) = F(t^*) , \quad (5.29)$$

then, we have the “sequential programming equation”

$$\bigwedge_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h,t) = \bigwedge_{h \in \mathbb{H}} S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) . \quad (5.30)$$

Note that in assumption [i](#)) we assume that  $\text{Id}(F)$  is a complete inf-semilattice. This implies that  $\bigwedge_{t \in T \subset \mathbb{T}} F(t) \in \text{Im}(F)$  for all subset  $T \subset \mathbb{T}$ . Assumption [i](#)) is hence stronger than assumption [ii](#)).

*Proof.* Since the couple  $(A, F)$  in [\(5.26\)](#) is Usual Time Consistent, we have that

$$A(h,t) = S^{A,F}(h, F(t)) , \quad \text{by Proposition } \a href="#">3.3.6 , \quad (5.31)$$

and the mapping  $S^{A,F}$  is increasing in its second argument. As the mapping  $S^{A,F}$  is increasing, it is lower semi-continuous by [Lemma 4.4.3](#). Using the fact that item [i](#)) or item [ii](#)) implies

$$\bigwedge_{t \in \mathbb{T}} F(t) \in \text{Im}(F) , \quad (5.32)$$

we have the following inequalities

$$S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) \leq \bigwedge_{t \in \mathbb{T}} S^{A,F}(h, F(t)) , \quad (5.33)$$

by [Definition 4.4.2](#) of lower semi-continuity of  $S^{A,F}$

$$\begin{aligned} \bigwedge_{h \in \mathbb{H}} S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) &\leq \bigwedge_{h \in \mathbb{H}} \bigwedge_{t \in \mathbb{T}} S^{A,F}(h, F(t)) , && \text{(by taking the infimum over } \mathbb{H}) \\ \bigwedge_{h \in \mathbb{H}} S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) &\leq \bigwedge_{h \in \mathbb{H}} \bigwedge_{t \in \mathbb{T}} A(h,t) , && \text{(by the Nested Formula } \a href="#">(5.31).) \end{aligned}$$

We hence have shown that [Equation \(5.28\)](#) is satisfied.

Now we turn to the equality case. By [Definition 4.4.2](#) of  $\wedge$ -Scott continuity, [Equation \(5.33\)](#) is an equality if the mapping  $S^{A,F}$  is  $\wedge$ -Scott continuous. We focus on the second condition. Assume that there exists  $t^* \in \mathbb{T}$  such that

$$F(t^*) = \bigwedge_{t \in \mathbb{T}} F(t) . \quad (5.34)$$

Then we have that

$$S^{A,F}(h, F(t^*)) = S^{A,F}(h, \bigwedge_{t \in \mathbb{T}} F(t)) , \quad \forall h \in \mathbb{H} ,$$

by [Equation \(5.29\)](#) that states that infimum is achieved,

$$\bigwedge_{t \in \mathbb{T}} S^{A,F}(h, F(t)) \leq S^{A,F}(h, F(t^*)) = \bigwedge_{t \in \mathbb{T}} S^{A,F}(h, F(t)) , \quad \forall h \in \mathbb{H} ,$$

by definition of the infimum over the set  $\mathbb{T}$  and because  $t^*$  belongs to the set  $\mathbb{T}$ . This provides the reverse inequality in [\(5.33\)](#). This ends the proof.  $\square$

## 5. Player consistency and decomposition of optimization problems

We call Equations (5.28) and (5.30) sequential because the optimization problem  $\bigwedge_{(h,t) \in \mathbb{H} \times \mathbb{T}} A(h,t)$  can be solved sequentially: we first optimize over  $\mathbb{T}$  with  $h$  fixed playing the role of a parameter, then we optimize over  $\mathbb{H}$ .

As optimization is done sequentially, we can allow the set  $\mathbb{T}$  to depend parametrically of elements  $h \in \mathbb{H}$ . The next corollary focuses on this case.

**Corollary 5.3.2.** *Assume that the sets  $\mathbb{A}$  and  $\mathbb{F}$  in (5.26) are equipped with orders denoted respectively by  $\leq$  and that the set  $\mathbb{A}$  is a complete inf-semilattice and the set  $\mathbb{F}$  is a complete lattice. Let the couple  $(A, F)$  in (5.26) be Usual Time Consistent, so that, by Proposition 3.3.6, there exists a mapping  $S^{A,F} : \mathbb{H} \times \text{Im}(F)$  increasing in its second argument such that*

$$A(h, t) = S^{A,F}(h, F(t)) . \quad (5.35)$$

Let  $S \subset \mathbb{H} \times \mathbb{T}$  be a set and let the set-valued mapping  $\mathcal{T} : \mathbb{H} \rightrightarrows \mathbb{T}$  be defined by

$$\mathcal{T}(h) = \{t \in \mathbb{T} \mid (h, t) \in S\} , \quad \forall h \in \mathbb{H} . \quad (5.36)$$

Then, we have the “sequential programming inequality”

$$\bigwedge_{h \in \text{dom}(\mathcal{T})} S^{A,F}(h, \bigwedge_{t \in \mathcal{T}(h)} F(t)) \leq \bigwedge_{(h,t) \in S} A(h, t) , \quad (5.37)$$

where the set  $\text{dom}(\mathcal{T})$  is given by

$$\text{dom}(\mathcal{T}) = \{h \in \mathbb{H} \mid \mathcal{T}(h) \neq \emptyset\} . \quad (5.38)$$

If, in addition, one of the following conditions is satisfied

- i) either  $\text{Im}(F)$  is a complete inf-semilattice and the mapping  $S^{A,F}$  is  $\wedge$ -Scott continuous,
- ii) for all  $h \in \mathbb{H}$ , there exists  $t_h^* \in \mathcal{T}(h)$  such that

$$\bigwedge_{t \in \mathcal{T}(h)} F(t) = F(t_h^*) , \quad (5.39)$$

then, we have the “sequential programming equation”

$$\bigwedge_{(h,t) \in S} A(h, t) = \bigwedge_{h \in \text{dom}(\mathcal{T})} S^{A,F}(h, \bigwedge_{t \in \mathcal{T}(h)} F(t)) . \quad (5.40)$$

Notice that if the set  $\mathbb{F}$  is a complete lattice, then the supremum over  $h$  in (5.37) and (5.40) can be taken over the entire set  $\mathbb{H}$ . Indeed, in this case, if  $\mathcal{T}(h)$  is empty we obtain

$$\bigwedge_{t \in \emptyset} F(t) = \top , \quad (5.41)$$

where  $\top$  denotes the top element of the complete lattice  $\mathbb{F}$ .

### 5.3.2. Parallel optimization with player consistency

The next proposition is the counterpart of Theorem 5.3.1 in the framework of Strong Monotone Player Consistency developed in Sect. 5.2.1. We hence consider a set  $P$  (of players), a set  $\mathbb{A}$  and two collections of sets  $(\mathbb{S}_p)_{p \in P}$  and  $(\mathbb{F}_p)_{p \in P}$ . We consider a collection of mappings  $(A, (F_p)_{p \in P})$

$$A : \prod_{p \in P} \rightarrow \mathbb{A} \quad (5.42a)$$

$$F_p : \mathbb{S}_p \rightarrow \mathbb{F}_p, \quad \forall p \in P. \quad (5.42b)$$

For clarity, we use the abuse of notation

$$\bigwedge_{(s_p)_{p \in P} \in \prod_{p \in P} \mathbb{S}_p} = \bigwedge_{s_P \in \mathbb{S}_P}. \quad (5.43)$$

**Proposition 5.3.3.** *Assume that the image sets  $\mathbb{A}$  and  $(\mathbb{F}_p)_{p \in P}$  in (5.42) are equipped with orders denoted respectively by  $\leq$  and are complete inf-semilattices. Let the tuple  $(A, (F_p)_{p \in P})$  in (5.42) be Strong Monotone Player Consistent, so that, by Theorem 5.2.5, there exist a mapping  $S^{A,F} : \prod_{p \in P} \text{Im}(F_p) \rightarrow \mathbb{A}$  increasing in all arguments such that*

$$A(s_P) = S^{A,F} \left( (F_p(s_p))_{p \in P} \right), \quad \forall s_P \in \mathbb{S}_P. \quad (5.44)$$

Then, we have the “parallel programming inequality”

$$S^{A,F} \left( \left( \bigwedge_{s_p \in \mathbb{S}_p} F_p(s_p) \right)_{p \in P} \right) \leq \bigwedge_{(s_p)_{p \in P}} A((s_p)_{p \in P}). \quad (5.45)$$

If, in addition, one of the following conditions is satisfied

- either  $\text{Im}(F_p)$  is a complete inf-semilattice for all  $p$  and the mapping  $S^{A,F}$  is  $\wedge$ -Scott continuous,
- or for each  $p \in P$  there exist  $s_p^* \in \mathbb{S}_p$  such that

$$\bigwedge_{s_p \in \mathbb{S}_p} F_p(s_p) = F_p(s_p^*), \quad \forall p \in P, \quad (5.46)$$

then, we have the “parallel programming equation”

$$\bigwedge_{(s_p)_{p \in P}} A((s_p)_{p \in P}) = S^{A,F} \left( \left( \bigwedge_{s_p \in \mathbb{S}_p} F_p(s_p) \right)_{p \in P} \right). \quad (5.47)$$

Equations (5.45) and (5.47) are called parallel because the infimum to the left hand side is done player by player in the right hand side.

### 5.3.3. Mixing parallel and sequential optimization

We introduce a notion of decomposability related to mixed player consistency. With this tool, we show how to handle both sequential and parallel decompositions of an optimization problem. Finally, we sketch a recursive procedure to go deeper and deeper in decomposition. For the sake of clarity, we use the abuses of notation

$$\prod_{p \in P} \mathbb{S}_p = \mathbb{S}_P \text{ and } \bigwedge_{(s_p)_{p \in P} \in \prod_{p \in P} \mathbb{S}_p} = \bigwedge_{s_P \in \mathbb{S}_P} . \quad (5.48)$$

**Definition 5.3.4.** Let  $P$  be a set of players and  $\mathbb{A}$  be a set. Let  $(\mathbb{S}_p)_{p \in P}$  be a collection of sets and  $A_P$  be a mapping given by

$$A_P : \mathbb{S}_P \rightarrow \mathbb{A} . \quad (5.49)$$

The set  $P$  of players is said to be  $(P_=, P_{\leq})$ -decomposable with respect to  $A_P$  if

- $(P_=, P_{\leq})$  is a partition of  $P$ ,
- there exists two sets  $\mathbb{F}_{P_=}$  and  $\mathbb{F}_{P_{\leq}}$  and two mappings

$$F_{P_=} : \mathbb{S}_{P_=} \rightarrow \mathbb{F}_{P_=} \text{ and } F_{P_{\leq}} : \mathbb{S}_{P_{\leq}} \rightarrow \mathbb{F}_{P_{\leq}} \quad (5.50)$$

such that the triplet  $(A, (F_{P_=}, F_{P_{\leq}}))$  of mappings in (5.49) and (5.50) is  $(P_=, P_{\leq})$ -Mixed Player Consistent as in Definition 5.2.6.

In that case, we know from Theorem 5.2.5 that the subaggregator denoted by  $S$  and defined in (5.10) is a mapping and that we have the nested formula

$$A_P(s_P) = S(F_{P_=}(s_{P_=}), F_{P_{\leq}}(s_{P_{\leq}})) . \quad (5.51)$$

The previous definition is closed to the definition 5.2.6 of Mixed Player Consistency. The difference is that only  $P$  and  $A_P$  are given in Definition 5.3.4. In particular, if a tuple  $(A, F_{P_=}, F_{P_{\leq}})$  is mixed player consistent, then the set  $P$  is  $(P_1, P_2)$ -decomposable with respect to  $A$ .

**Proposition 5.3.5.** Let  $P$  be a set of players and  $\mathbb{A}$  be a set. Let  $(\mathbb{S}_p)_{p \in P}$  be a collection of sets and  $A_P$  be a mapping given by

$$A_P : \mathbb{S}_P \rightarrow \mathbb{A} , \quad (5.52)$$

and consider the problem

$$\bigwedge_{s_P \in \mathbb{S}_P} A(s_P) . \quad (5.53)$$

Assume that the set  $P$  is  $(P_=, P_{\leq})$ -decomposable, and let  $F_{P_=}$  and  $F_{P_{\leq}}$  be mappings as in (5.50). If the mapping  $S$  in (5.51) is  $\wedge$ -Scott continuous in its second argument, then problem (5.53) can be decomposed sequentially and in parallel, that is,

$$\bigwedge_P A(s_P) = \bigwedge_{P_=} S\left(F_{P_=}(s_{P_=}), \bigwedge_{P_{\leq}} F_{P_{\leq}}(s_{P_{\leq}})\right) , \quad (5.54)$$

where the infimum over the set  $P_{\leq}$  is done in parallel.

### 5.3. Application of player consistency to decomposition for optimization

We now introduce the notion of reduced system to be able to apply decomposition scheme recursively.

**Definition 5.3.6.** Let  $P$  be a set (of players), and  $\mathbb{A}$  be a set. We consider two collections of sets  $(\mathbb{S}_p)_{p \in P}$  and  $(\mathbb{F}_p)_{p \in P}$  and a collection of mappings  $(A, (F_p)_{p \in P})$  such that

$$A : \prod_{p \in P} \mathbb{S}_p \rightarrow \mathbb{A} , \quad (5.55a)$$

$$F_p : \mathbb{S}_p \rightarrow \mathbb{F}_p , \quad \forall p \in P . \quad (5.55b)$$

Let  $P_r \subset P$  be a non empty set of players, and let  $(\bar{s}_p)_{p \notin P_r} \in \prod_{p \notin P_r} \mathbb{S}_p$  be fixed elements. We introduce the mapping  $A_r$  defined by

$$A_r : \prod_{p \in P_r} \mathbb{S}_p \rightarrow \mathbb{A} \quad (5.56)$$

$$(s_{P_r}, \bar{s}_{P \setminus P_r}) \mapsto A(s_{P_r}, \bar{s}_{P \setminus P_r}) . \quad (5.57)$$

The tuple of mappings  $(A_r, F_{P_r})$  is said to be a  $(P_r, \bar{s}_{P \setminus P_r})$ -reduced system.

Let us detail how to apply Proposition 5.3.5 recursively using the notion of reduced system.

- Assume that the triplet  $(A, (F_{P_=}, F_{P_{\leq}}))$  of mappings in (5.49) and (5.50) is  $(P_=, P_{\leq})$ -Mixed Player Consistent as in Definition 5.2.6
- We first consider the  $(P_{\leq}, s_{P_=})$ -reduced system  $(A_{\leq}, F_{P_{\leq}})$  as in Definition 5.3.6 parametrized in  $s_{P_=}$ . We hence focus on the optimization problem  $\bigwedge_{P_{\leq}} F_{P_{\leq}}(s_{P_{\leq}})$  that can be solved in parallel.
- We denote by  $s_{P_{\leq}}^*$  an optimal solution of  $\bigwedge_{P_{\leq}} F_{P_{\leq}}(s_{P_{\leq}})$ .
- We secondly consider the  $(P_=, s_{P_{\leq}}^*)$ -reduced system  $(A_=, F_{P_=})$  as in Definition 5.3.6. If  $P_=$  est  $(P_=, P_{\leq})$ -decomposable with respect to  $A_=$ , we apply recursively Proposition 5.3.5 to this  $(P_=, s_{P_{\leq}}^*)$ -reduced system  $(A_=, F_{P_=})$ .

The next pseudo code sums up the recursion.

**Data:** A set  $P$  of players, a mapping  $A_P$   
**Result:** A sequential-parallel decomposition scheme

- 1 **while** If  $P$  is  $(P_=, P_{\leq})$ -decomposable with respect to  $A_P$  as in Definition 5.3.4. **do**
- 2     For each  $s_{P_=}$ , solve in parallel the problem  $\bigwedge_{P_{\leq}} F_{P_{\leq}}(s_{P_{\leq}})$ . and denote the solution  $s_{P_{\leq}}^*$  ;
- 3     consider the  $(P_=, s_{P_{\leq}}^*)$ -reduced system  $(A_=, F_{P_=})$  as in Definition 5.3.6. Apply algorithm to the couple  $(P, A_=)$  ;
- 4 **end**

**Algorithm 1:** Recursive decomposition scheme algorithm

## 5.4. Conclusion

In Sect. 5.2, we have generalized the notion of time consistency to the one of player consistency where there is no notion of time. We have described how player consistency and time consistency can be written in a framework of  $\mathcal{R}$ -consistency for binary relations.

In Sect. 5.3, we have shown how a Nested Formula is adapted for Nested Optimization and can be used to decompose an optimization problem. We have described two processes of decomposition: a parallel one and a sequential one. Then, we have examined when we can mix sequential and parallel decomposition and how to apply them recursively.

We discuss here perspectives that remain open for us.

- Further work needs to be done to account for dependencies between players, that is, when there are joint constraints on the players decision sets.
- In machine learning, influence diagrams capture dynamic interactions between players. They may provide an application for the above decomposition method.

## Conclusion of Part II

We started with the idea of adapting decomposition methods to deal with risk measures. As a well-known decomposition scheme relies on dynamic programming and on the notion of time consistency, we first studied time consistency. Our survey of the literature in the field has revealed that it was very disparate, mixing peripheral assumptions with the ideas of time consistency and nested formula. Our first work has been to clarify the contributions of each author building an abstract framework for time consistency. This framework highlights the fact that the notion of time is not essential. We hence have been able to generalize the notion of time consistency for mappings to the one of player consistency for mappings. Then we have shown that time consistency and player consistency can be incorporated into a common framework of  $\mathcal{R}$ -consistency for binary relations. Definition, theorems and characterizations for player consistency are naturally deduced from time consistency. Our framework is incomplete since we lack examples; however, we have clearly identified the difficulties.

To sum up the works in this part, we connect the frameworks of Chap. 3, Chap. 4 and Chap. 5.2 through Table 5.2. We recall that time consistency appears as a particular case of player consistency, both being an application of consistency for binary relations. We have chosen to focus on usual time consistency in Table 5.2 when dealing with time consistency. The final Chap. 5.3 illustrates how the results of Table 5.2 can be used to deal with optimization problems.

We detail now the different perspectives of work that we have identified.

- One could look for application of the characterization of time consistency in terms of acceptance sets, either to demonstrate that some mappings are not time consistent, or to build mappings that are time consistent. In particular, we have provided a characterization of player consistency in terms of acceptance set.
- The conditions to be satisfied by generalized convex mappings in order to obtain a nested formula are difficult to check. Finding mappings that are not convex risk measures which satisfy these conditions is an open question. We should investigate what happens when we mix translation invariant mappings, Fenchel-Moreau transform, supremum to construct time consistent mappings.
- We did not have the time to discuss the connection between the inheritance of time consistency presented in Sect. 4.4 and the recent work of Dentcheva and Ruszczyński on risk forms [Dentcheva and Ruszczyński \(2018\)](#).
- Further work needs to be done to account for dependencies between players, that is, when there are joint constraints on the players decision sets.



5. Player consistency and decomposition of optimization problems

Consistency	Chap. 3	Sect. 4.2	Sect. 4.3	Sect. 5.2.1
Sets	$\mathbb{H}, \mathbb{T}, \mathbb{A}, \mathbb{F}$	$\mathbb{H}, \mathbb{T}, \mathbb{A}, \mathbb{F}$	$\mathbb{X}, \mathbb{Y}, \mathbb{Z}, \mathbb{Y}'$	$(\mathbb{S}_p)_{p \in P}$
Property of sets	$\mathbb{A}$ and $\mathbb{F}$ equipped with orders denoted by $\leq$	$(\mathbb{H}, \oplus), (\mathbb{A}, \oplus), (\mathbb{F}, \oplus)$ subgroups of $(\mathbb{T}, \oplus)$	$\emptyset$	$\emptyset$
Mappings	$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$ $F : \mathbb{T} \rightarrow \mathbb{F}$	$F : \mathbb{T} \rightarrow \mathbb{F}$ $\rho : \mathbb{T} \rightarrow \mathbb{A}$ $A_\rho : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{A}$	$\theta_{\mathbb{X} \times \mathbb{Z}} : \mathbb{X} \times \mathbb{Z} \rightarrow \mathbb{Y}'$ $\theta_{\mathbb{Z}} : \mathbb{Z} \rightarrow \mathbb{Y}'$ $\theta_{\mathbb{X}} : \mathbb{X} \rightarrow \mathbb{Y}'$ $\varphi : \mathbb{Y}' \times \mathbb{Y} \rightarrow \overline{\mathbb{R}}$ $g : \mathbb{Y} \rightarrow \overline{\mathbb{R}}$ $\Phi : \mathbb{X} \times (\mathbb{Y} \times \mathbb{Z}) \rightarrow \overline{\mathbb{R}}$ $G : \mathbb{Y} \times \mathbb{Z} \rightarrow \overline{\mathbb{R}}$	$\emptyset$
Property of mappings	$\emptyset$	$F$ ( $\mathbb{T}, \mathbb{F}$ )-trans. inv. <sup>2</sup> $\rho$ ( $\mathbb{T}, \mathbb{A}$ )-trans. inv. <sup>2</sup>	$\varphi, \Phi$ couplings	$\emptyset$
Relations	$\emptyset$	$\emptyset$	$\emptyset$	$\mathcal{R}$ on $\prod_{p \in P} \mathbb{S}_p$ $\mathbb{S}_p$ on $\mathbb{S}_p$
Expression of Consistency	$F(t) \geq F(t')$ $\downarrow$ $A(h, t) \geq A(h, t')$	$\mathcal{A}_F \oplus F(t) \subset \mathcal{A}_\rho \ast \mathbb{C}_\rho(t)$ $\mathcal{A}_F + \mathcal{A}_{\rho _{\mathbb{F}}} = \mathcal{A}_\rho$	$G^\Phi = g^\varphi \circ \theta_{\mathbb{X}}$	$\prod_{p \in P} \mathbb{S}_p \subset \mathcal{R}$
Results	Proposition 3.3.6 and Proposition 3.5.6	Proposition 4.2.3 and Proposition 4.2.4	Proposition 4.3.3	$\emptyset$

Table 5.2.: Comparison of different framework for Consistency.

1: cis. = complete inf semilattice

2: trans. inv. = translation invariant

- In machine learning, influence diagrams capture dynamic interactions between players. They may provide an application for the above decomposition method.

## **Part III.**

### **Risk averse competitive equilibrium**



# 6. On risk averse competitive equilibrium

This chapter provides the content of [Gérard, Leclère, and Philpott \(2018\)](#). Some notations have changed to preserve coherence with the rest of the document.

## Contents

---

<b>6.1. Introduction</b>	<b>108</b>
6.1.1. Notation	109
<b>6.2. Statement of problem</b>	<b>109</b>
6.2.1. Social planner problem	109
6.2.2. Equilibrium problem	111
6.2.3. Trading risk with Arrow-Debreu securities	112
<b>6.3. Some equivalences between social planner and equilibrium problems</b>	<b>113</b>
6.3.1. Equivalence in the risk neutral case	113
6.3.2. Equivalence in the risk-averse case	113
<b>6.4. Multiple risk averse equilibrium</b>	<b>116</b>
6.4.1. Multiple equilibrium	116
6.4.2. Analytical results	117
<b>6.5. Analytical results</b>	<b>119</b>
6.5.1. Parametric solution with respect to $\pi$	119
6.5.2. Finding price equilibrium	122
<b>6.6. Unstability of equilibrium</b>	<b>123</b>
<b>6.7. Conclusion</b>	<b>123</b>

---

## 6. On risk averse competitive equilibrium

We discuss risked competitive partial equilibrium in a setting in which agents are endowed with coherent risk measures. In contrast to social planning models, we show by example that risked equilibria are not unique, even when agents' objective functions are strictly concave. We also show that standard computational methods find only a subset of the equilibria, even with multiple starting points.

### 6.1. Introduction

Most industrialised regions of the world have over the last thirty years established wholesale electricity markets that take the form of an auction that matches supply and demand. The exact form of these auction mechanisms vary by jurisdiction, but they typically require offers of energy from suppliers at costs they are willing to supply, and clear a market by dispatching these offers in order of increasing cost. Day-ahead markets such as those implemented in many North American electricity systems, seek to arrange supply well in advance of its demand, so that thermal units can be prepared in time. Since the demand cannot be predicted with absolute certainty, day-ahead markets must be accompanied by a separate balancing market to deal with the variation in load and generator availability in real time. These are often called *two-settlement* markets. The market mechanisms are designed to be as efficient as possible in the sense that they should aim to maximize the total welfare of producers and consumers.

In response to pressure to reduce  $CO_2$  emissions and increase the penetration of renewables, electricity pool markets are procuring increasing amounts of electricity from intermittent sources such as wind and solar. If probability distributions for intermittent supply are known for these systems then it makes sense to maximize the expected total welfare of producers and consumers in each dispatch. Then many repetitions of this will yield a long run total benefit that is maximized. Maximizing expected welfare can be modeled as a two-stage stochastic program. Methods for computing prices and single-settlement payment mechanisms for such a *stochastic market clearing* mechanism are described in a number of papers (see [Pritchard et al. \(2010\)](#), [Wong and Fuller \(2007\)](#) and [Zakeri et al. \(2016\)](#)). When evaluated using the assumed probability distribution on supply, stochastic market clearing can be shown to be more efficient than two-settlement systems.

If agents in these systems are risk averse then one might also seek to maximize some risk-adjusted social welfare. In this setting the computation of prices and payments to the agents becomes more complicated. If agents use coherent risk measures then it is possible to define a complete market for risk in a precise sense. If the market is complete then a perfectly competitive partial equilibrium will also maximize risk-adjusted social welfare, i.e. it is efficient. On the other hand if the market for risk is not complete, then perfectly competitive partial equilibrium can be inefficient. This has been explored in a number of papers (see e.g. [de Maere d'Aertrycke et al. \(2017\)](#), [Ehrenmann and Smeers \(2011\)](#) and [Ralph and Smeers \(2015\)](#)).

In this paper we study a class of stochastic dispatch and pricing mechanisms under the assumption that agents will attempt to maximize their risk-adjusted welfare at these prices. Agents have coherent risk measures and are assumed to behave as price takers in the energy and risk markets. We aim at enlightening some difficulties

that arise when risk markets are not complete. We describe a simple instance of a stochastic market that has three different equilibria. Two of these points are stable in the sense of Samuelson (1941) and are attractors of tatonnement algorithms. The third equilibrium is unstable, yet is the solution yielded by the well-known PATH solver in GAMS (See Ferris and Munson (2000)). Our example illustrates the delicacy of seeking numerical solutions for equilibria in incomplete markets. Since these are used for justifying decisions, the nonuniqueness of solutions in this setting is undesirable.

The paper is laid out as follow. In Section 6.2 we present the equilibrium and optimization models we are going to study. In Section 6.3 we give links between equilibrium and optimization problems in the risk neutral and complete risk-averse cases. Finally, in Section 6.4 we showcase a simple example with multiple equilibria in the incomplete risk-averse case.

### 6.1.1. Notation

We use the following notation throughout the paper:  $\llbracket a; b \rrbracket$  is the set of integers between  $a$  and  $b$  (included), random variables are denoted in bold,  $\Omega$  is a finite sample space,  $\mathbb{P}$  is a probability distribution over  $\Omega$ ,  $\mathbb{E}_{\mathbb{P}}$  is used to refer to expectation with respect to  $\mathbb{P}$ ,  $\rho$  is used to refer to a risk measure. We denote by  $x \perp y$  the complementarity condition  $x^T y = 0$ .

## 6.2. Statement of problem

Consider a two time-step single-settlement market for one good. In a single settlement market, the producer can arrange in advance for a production of  $x$  at a marginal cost  $cx$  as a first-step decision, and choose the value of a recourse variable  $\mathbf{x}_r$  incurring an uncertain marginal cost  $\mathbf{c}_r \mathbf{x}_r$ . We assume that there are a finite number of scenarios  $\omega \in \Omega$  determining the coefficient  $\mathbf{c}_r(\omega)$ .

The product is purchased in the second step by a consumer with a utility function  $\mathbf{V}(\omega)\mathbf{y}(\omega) - \frac{1}{2}\mathbf{r}(\omega)\mathbf{y}^2(\omega)$ . The consumer has no first-stage decision, and the amount purchased  $\mathbf{y}(\omega)$  depends on the scenario.

### 6.2.1. Social planner problem

Decisions  $x$ ,  $\mathbf{x}_r(\omega)$  and  $\mathbf{y}(\omega)$  can be made to maximize a social objective. We denote by

$$\mathbf{W}_p(\omega) = -\frac{1}{2}cx^2 - \frac{1}{2}\mathbf{c}_r(\omega)\mathbf{x}_r(\omega)^2, \quad (6.1a)$$

the welfare of the producer, and by

$$\mathbf{W}_c(\omega) = \mathbf{V}(\omega)\mathbf{y}(\omega) - \frac{1}{2}\mathbf{r}(\omega)\mathbf{y}(\omega)^2, \quad (6.1b)$$

the welfare of the consumer where both these expressions ignore the price paid for the good in scenario  $\omega$ . Then the welfare of the social planner can be defined by  $\mathbf{W}_{sp} = \mathbf{W}_p + \mathbf{W}_c$ .

## 6. On risk averse competitive equilibrium

Optimization of the social objective requires us to aggregate the uncertain outcomes from the scenarios. This can be done by taking expectations with respect to an underlying probability measure  $\mathbb{P}$  or using a more general risk measure.

### Risk neutral social planner problem

Endow the set of scenario  $\Omega$  with a probability  $\mathbb{P}$ , then a risk-neutral social planner might seek to maximize the expected total social welfare under the constraint that supply equals demand. This problem is denoted by  $\text{RnSp}(\mathbb{P})$  and reads

$$\begin{aligned} & \text{RnSp}(\mathbb{P}) : \\ & \max_{x, \mathbf{x}_r, \mathbf{y}} \quad \mathbb{E}_{\mathbb{P}}[\mathbf{W}_{sp}] , \end{aligned} \tag{6.2a}$$

$$\text{s.t.} \quad x + \mathbf{x}_r(\omega) \geq \mathbf{y}(\omega) , \quad \forall \omega \in \Omega . \tag{6.2b}$$

### Risk averse social planner problem

Choosing expectation  $\mathbb{E}_{\mathbb{P}}$ , assumes a risk-neutral point of view, where two random losses with same expectation but different variances are deemed equivalent. In practice a number of agents are risk averse. To model risk aversion we generally use a *risk measure*  $\rho$ , that is a functional that associates to a random welfare its deterministic equivalent, i.e. the deterministic welfare deemed as equivalent to the random loss.

A risk-averse planner solves a maximization problem  $\text{RaSp}(\rho)$  defined by

$$\begin{aligned} & \text{RaSp}(\rho) : \\ & \max_{x, \mathbf{x}_r, \mathbf{y}} \quad \rho(\mathbf{W}_{sp}) , \end{aligned} \tag{6.3a}$$

$$\text{s.t.} \quad x + \mathbf{x}_r(\omega) \geq \mathbf{y}(\omega) , \quad \forall \omega \in \Omega . \tag{6.3b}$$

A risk measure  $\rho$  is said to be coherent (see [Artzner et al. \(1999\)](#)) if it satisfies four natural properties: monotonicity (if  $\mathbf{X} \geq \mathbf{Y}$  then  $\rho[\mathbf{X}] \geq \rho[\mathbf{Y}]$ ), concavity ( $\rho$  is concave), translation-equivariance ( $\rho[\mathbf{X} + c] = \rho[\mathbf{X}] + c$  with  $c \in \mathbb{R}$ ) and positive homogeneity ( $\rho[\lambda \mathbf{X}] = \lambda \rho[\mathbf{X}]$ , with  $\lambda \geq 0$ ). By convex duality theory (see [Shapiro et al. \(2014\)](#)), a lower-semicontinuous coherent risk measure can be written  $\rho(\mathbf{Z}) = \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{Z}]$ , where  $\mathcal{Q}$  is a closed, convex, non-empty set of probability distributions over  $\Omega$ . If  $\mathcal{Q}$  is a polyhedron defined by  $K$  extreme points  $(\mathbb{Q}_k)_{k \in [1; K]}$ , then the risk measure is denoted  $\check{\rho}$  and said to be *polyhedral*, with  $\check{\rho}[\mathbf{Z}] = \min_{\mathbb{Q}_1, \dots, \mathbb{Q}_K} \mathbb{E}_{\mathbb{Q}_k}[\mathbf{Z}]$ .

Problem  $\text{RaSp}(\check{\rho})$  can be written as follows

$$\begin{aligned} & \text{RaSp}(\check{\rho}) : \\ & \max_{\theta, x, \mathbf{x}_r, \mathbf{y}} \quad \theta \end{aligned} \tag{6.4a}$$

$$\text{s.t.} \quad \theta \leq \mathbb{E}_{\mathbb{Q}_k}[\mathbf{W}_{sp}] , \quad \forall k \in [1; K] , \tag{6.4b}$$

$$x + \mathbf{x}_r(\omega) \geq \mathbf{y}(\omega) , \quad \forall \omega \in \Omega . \tag{6.4c}$$

In what follows we assume that all risk measures are coherent.

**Remark on non linearity of risk averse objective function**

By linearity of expectation we have  $\mathbb{E}_{\mathbb{P}}[\mathbf{W}_{sp}] = \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p] + \mathbb{E}_{\mathbb{P}}[\mathbf{W}_c]$  hence the criterion of the social planner is natural, which is not the case anymore with risk-aversion. The social planner criterion could be either  $\rho(\mathbf{W}_{sp})$  or  $\rho(\mathbf{W}_p) + \rho(\mathbf{W}_c)$ . Furthermore, by concavity and positive homogeneity, we have  $\rho(\mathbf{W}_p + \mathbf{W}_c) \geq \rho(\mathbf{W}_p) + \rho(\mathbf{W}_c)$ .

**6.2.2. Equilibrium problem**

We now define a competitive partial equilibrium for our model. This competitive equilibrium can be risk neutral or risk averse. Definitions come from general equilibrium theory (See [Arrow and Debreu \(1954\)](#) or [Uzawa \(1960\)](#)).

**Risk neutral equilibrium**

Given a probability  $\mathbb{P}$  on  $\Omega$ , a *risk-neutral equilibrium*  $\text{RnEq}(\mathbb{P})$  is a set of prices  $\{\boldsymbol{\pi}(\omega), \omega \in \Omega\}$  such that there exists a solution to the system

$$\text{RnEq}(\mathbb{P}) :$$

$$\max_{x, \mathbf{x}_r} \mathbb{E}_{\mathbb{P}} \left[ \mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r) \right], \quad (6.5a)$$

$$\max_{\mathbf{y}} \mathbb{E}_{\mathbb{P}} \left[ \mathbf{W}_c - \boldsymbol{\pi}\mathbf{y} \right], \quad (6.5b)$$

$$0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \quad (6.5c)$$

Here, the producer maximizes its expected profit (6.5a), the consumer maximizes its expected utility (6.5b) and the market clears with (6.5c) (which means that either prices are null or supply equals demand). As the consumer has no first stage decision, she can optimize each scenario independently and so problem (6.5b) can be replaced by

$$\max_{\mathbf{y}(\omega)} \mathbf{W}_c(\omega) - \boldsymbol{\pi}(\omega)\mathbf{y}(\omega), \quad \forall \omega \in \Omega.$$

**Risk averse equilibrium**

Given two risk measures  $\rho_p$  and  $\rho_c$  over  $\Omega$ , a *risk-averse equilibrium*  $\text{RaEq}(\rho_p, \rho_c)$  is a set of prices  $\{\boldsymbol{\pi}(\omega) : \omega \in \Omega\}$  such that there exists a solution to the following system

$$\text{RaEq}(\rho_p, \rho_c) :$$

$$\max_{x, \mathbf{x}_r} \rho_p \left( \mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r) \right), \quad (6.6a)$$

$$\max_{\mathbf{y}} \rho_c \left( \mathbf{W}_c - \boldsymbol{\pi}\mathbf{y} \right), \quad (6.6b)$$

$$0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0, \quad \forall \omega \in \Omega. \quad (6.6c)$$

Since the coherent risk measure  $\rho_c$  of the consumer is monotonic, and noting that she has no first-stage decision, she can optimize scenario per scenario. Thus, she is



## 6. On risk averse competitive equilibrium

*insensitive to risk* as any monotonic risk measure will lead to the same action (although not the same welfare). Since  $\rho_p$  is also monotonic, we can endow both agents with the same risk measure. In that case, we denote problem (6.6) by  $\text{RaEq}(\rho)$ .

We now consider polyhedral risk measure  $\check{\rho}$ , using formulation (6.4), the equilibrium problem (6.6) reads

$$\begin{aligned} & \text{RaEq}(\check{\rho}) : \\ & \max_{\theta, x, \mathbf{x}_r} \quad \theta \end{aligned} \tag{6.7a}$$

$$\begin{aligned} & \text{s.t.} \quad \theta \leq \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r)] , \quad \forall k \in \llbracket 1; K \rrbracket , \\ & \max_{\mathbf{y}(\omega)} \quad \mathbf{W}_c(\omega) - \boldsymbol{\pi} \mathbf{y}(\omega) , \quad \forall \omega \in \Omega , \end{aligned} \tag{6.7b}$$

$$0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0 , \quad \forall \omega \in \Omega . \tag{6.7c}$$

### 6.2.3. Trading risk with Arrow-Debreu securities

Until now, we have considered equilibrium problems in an incomplete market. Following the path of [Philpott et al. \(2016\)](#), we complete the market using Arrow-Debreu securities.

**Definition 6.2.1.** An *Arrow-Debreu security* for node  $\omega \in \Omega$  is a contract that charges a price  $\mu(\omega)$  in the first stage, to receive a payment of 1 in scenario  $\omega$ .

The consumer now has a first-stage decision which is the number of contracts she buys, so the choice of the consumer risk measure  $\rho_c$  has now consequences. For convenience, this risk measure  $\rho_c$  is chosen to be the same as that of the producer  $\rho_p$  and will be denoted by  $\rho$ . Unless stated otherwise, from now on we use polyhedral risk measures.

Denote  $\mathbf{a}(\omega)$  (resp.  $\mathbf{b}(\omega)$ ) the number of Arrow-Debreu securities bought by the producer (resp. the consumer). We denote by  $\boldsymbol{\mu}(\omega)$  the price of the Arrow-Debreu securities associated with scenario  $\omega$ . In this case the producer pays  $\sum_{\omega \in \Omega} \boldsymbol{\mu}(\omega) \mathbf{a}(\omega)$  in the first stage, in order to receive  $\mathbf{a}(\omega)$  in scenario  $\omega$ . As  $\mathbf{a}(\omega) + \mathbf{b}(\omega)$  represents excess demand, requiring that supply is greater than demand consists in requiring  $\mathbf{a}(\omega) + \mathbf{b}(\omega) \leq 0$ . Prices  $\{\boldsymbol{\pi}(\omega), \boldsymbol{\mu}(\omega)\}_{\omega \in \Omega}$  form a *risk-trading equilibrium* if there exists a solution to:

$$\begin{aligned} & \text{RaEq-AD}(\check{\rho}) : \\ & \max_{\theta, x, \mathbf{x}_r, \mathbf{a}} \quad \theta - \sum_{\omega \in \Omega} \boldsymbol{\mu}(\omega) \mathbf{a}(\omega) \end{aligned} \tag{6.8a}$$

$$\text{s.t.} \quad \theta \leq \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r) + \mathbf{a}] , \quad \forall k \in \llbracket 1; K \rrbracket , \tag{6.8b}$$

$$\max_{\phi, \mathbf{y}, \mathbf{b}} \quad \phi - \sum_{\omega \in \Omega} \boldsymbol{\mu}(\omega) \mathbf{b}(\omega) \tag{6.8c}$$

$$\text{s.t.} \quad \phi \leq \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_c - \boldsymbol{\pi} \mathbf{y} + \mathbf{b}] , \quad \forall k \in \llbracket 1; K \rrbracket , \tag{6.8d}$$

$$0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0 , \quad \forall \omega \in \Omega , \tag{6.8e}$$

$$0 \leq -\mathbf{a}(\omega) - \mathbf{b}(\omega) \perp \boldsymbol{\mu}(\omega) \geq 0 , \quad \forall \omega \in \Omega . \tag{6.8f}$$

and the price  $\mu$  is zero. Given prices  $\boldsymbol{\pi}(\omega)$  and  $\mu$  the risk-trading

## 6.3. Some equivalences between social planner and equilibrium problems

We recall a trivial equivalence between problem  $\text{RnSp}(\mathbb{P})$  and problem  $\text{RnEq}(\mathbb{P})$  before showing an equivalence between problem  $\text{RaSp}(\check{\rho})$  and problem  $\text{RaEq-AD}(\check{\rho})$ .

### 6.3.1. Equivalence in the risk neutral case

**Proposition 6.3.1.** *Let  $\mathbb{P}$  be a probability measure over  $\Omega$ . The elements  $x^\sharp$ ,  $\mathbf{x}_r^\sharp$  and  $\mathbf{y}^\sharp$  are optimal solutions to  $\text{RnSp}(\mathbb{P})$  if and only if there exist equilibrium prices  $\boldsymbol{\pi}^\sharp$  for  $\text{RnEq}(\mathbb{P})$  with associated optimal decisions  $x^\sharp$ ,  $\mathbf{x}_r^\sharp$  and  $\mathbf{y}^\sharp$ .*

*Proof.* As the producer and the consumer optimize over different uncoupled variables, it is equivalent to optimize their objectives separately or jointly. Problem (6.5) is thus equivalent to

$$\begin{aligned} \max_{x, \mathbf{x}_r, \mathbf{y}} \quad & \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r)] + \mathbb{E}_{\mathbb{P}}[\mathbf{W}_c - \boldsymbol{\pi}\mathbf{y}] , \\ & 0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0 , \quad \forall \omega \in \Omega , \end{aligned}$$

which by linearity of the expectation is equivalent to

$$\begin{aligned} \max_{x, \mathbf{x}_r, \mathbf{y}} \quad & \mathbb{E}_{\mathbb{P}}[\mathbf{W}_{sp} + \boldsymbol{\pi}(x + \mathbf{x}_r - \mathbf{y})] , \\ & 0 \leq x + \mathbf{x}_r(\omega) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0 , \quad \forall \omega \in \Omega . \end{aligned}$$

This is equivalent to the optimality conditions for problem (6.2a). Convexity and linearity of constraints ends the proof.  $\square$

**Corollary 6.3.2.** *If both the producer's and the consumer's criterion are strictly concave and if  $\mathbb{P}$  charges all  $\omega$ , then the problem  $\text{RnSp}(\mathbb{P})$  admits a unique solution and the problem  $\text{RnEq}(\mathbb{P})$  admits a unique equilibrium.*

*Proof.* The probability distribution  $\mathbb{P}$  charges all  $\omega$ . Then by strict concavity,  $\text{RnSp}(\mathbb{P})$  has a unique solution. If  $\text{RnEq}(\mathbb{P})$  has two different solutions  $(x^1, \mathbf{x}_r^1, \mathbf{y}^1)$  and  $(x^2, \mathbf{x}_r^2, \mathbf{y}^2)$  with  $\boldsymbol{\pi}^1$  and  $\boldsymbol{\pi}^2$  respectively then, by Proposition 6.3.1,  $x^1 = x^2$ ,  $\mathbf{x}_r^1 = \mathbf{x}_r^2$ , and  $\mathbf{y}^1 = \mathbf{y}^2$ . Since (6.5b) implies  $\boldsymbol{\pi}^1(\omega) = \mathbf{V}(\omega) - \mathbf{r}(\omega)\mathbf{y}^1(\omega)$ , we have  $\boldsymbol{\pi}^1 = \boldsymbol{\pi}^2$  which gives the result.  $\square$

### 6.3.2. Equivalence in the risk-averse case

The following proposition is an extension of Theorem 7 of [Ralph and Smeers \(2015\)](#), to a model with producers and consumers, in the special case of a finite number of scenarios with polyhedral risk measures.

**Proposition 6.3.3.** *Let  $\boldsymbol{\pi}$  and  $\boldsymbol{\mu}$  be equilibrium prices such that  $(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp, \mathbf{a}, \mathbf{b}, \theta, \varphi)$  solves  $\text{RaEq-AD}(\check{\rho})$ . Then*

(i)  *$\boldsymbol{\mu}$  is a probability measure, and  $x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp$  solves the risk-neutral social planning problem when evaluated using probability  $\boldsymbol{\mu}$ ,  $\text{RnSp}(\boldsymbol{\mu})$ .*

(ii)  *$x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp$  solves the risk-averse social planning problem,  $\text{RaEq-AD}(\check{\rho})$  with worst case measure  $\boldsymbol{\mu}$ .*

## 6. On risk averse competitive equilibrium

*Proof.* (i) Each agent problem is convex with linear constraints. Hence the optimal solution satisfies for each problem the Karush-Kuhn-Tucker (KKT) conditions. The Lagrangian of the producer problem reads

$$\mathcal{L}_p = \theta - \sum_{\omega \in \Omega} \mu(\omega) \mathbf{a}(\omega) + \sum_k \lambda_k \left( \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r) + \mathbf{a}] - \theta \right),$$

where  $\lambda_k$  is the multiplier associated to constraint (6.8b). Then, the KKT conditions imply that  $\sum_k \lambda_k = 1$ , and  $\boldsymbol{\mu} = \sum_k \lambda_k \mathbb{Q}_k$ . In particular,  $\boldsymbol{\mu}$  is a probability measure in  $\mathcal{Q}$ . Furthermore  $(x^\sharp, \mathbf{x}_r^\sharp)$  maximizes  $\sum_{\omega \in \Omega} \boldsymbol{\mu}(\omega) (\mathbf{W}_p(\omega) - \boldsymbol{\pi}(\omega)(x + \mathbf{x}_r(\omega)))$  which is the risk-neutral producer objective evaluated with measure  $\boldsymbol{\mu}$ .

Similarly, looking at the consumer problem with multiplier  $\sigma_k$  associated to constraint (6.8d), we obtain  $\sum_k \sigma_k = 1$  and  $\boldsymbol{\mu} = \sum_k \sigma_k \mathbb{Q}_k$ . Hence, the consumer maximizes her risk-neutral objective under the same probability  $\boldsymbol{\mu}$  as the producer.

Since by hypothesis the solutions satisfy (6.8e) we have that  $(x^\sharp, \mathbf{x}_r^\sharp(\omega), \mathbf{y}^\sharp(\omega))$  solves  $\text{RnSp}(\boldsymbol{\mu})$ .

(ii) Observe that complementary slackness gives

$$\begin{aligned} \lambda_k \left( \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_p^\sharp + \boldsymbol{\pi}(x^\sharp + \mathbf{x}_r^\sharp) + \bar{\mathbf{a}}] - \bar{\theta} \right) &= 0, \\ \sigma_k \left( \mathbb{E}_{\mathbb{Q}_k} [\mathbf{W}_c^\sharp - \boldsymbol{\pi} \mathbf{y}^\sharp + \bar{\mathbf{b}}] - \bar{\varphi} \right) &= 0, \end{aligned}$$

where  $\mathbf{W}_p^\sharp$  and  $\mathbf{W}_c^\sharp$  are defined by (6.1) in terms of  $x^\sharp$ ,  $\mathbf{x}_r^\sharp$  and  $\mathbf{y}^\sharp$ . Summing over  $k$ , and leveraging (6.8f) gives

$$\begin{aligned} \bar{\theta} + \bar{\varphi} &= \mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_p^\sharp + \boldsymbol{\pi}(x^\sharp + \mathbf{x}_r^\sharp) + \bar{\mathbf{a}}] + \mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_c^\sharp - \boldsymbol{\pi} \bar{\mathbf{y}} + \bar{\mathbf{b}}], \\ &= \mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp]. \end{aligned} \quad (6.11)$$

However as

$$\begin{aligned} \bar{\theta} + \bar{\varphi} &= \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [\mathbf{W}_p^\sharp + \boldsymbol{\pi}(x^\sharp + \mathbf{x}_r^\sharp) + \bar{\mathbf{a}}] \\ &\quad + \min_{\mathbb{Q}' \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}'} [\mathbf{W}_c^\sharp - \boldsymbol{\pi} \mathbf{y}^\sharp + \bar{\mathbf{b}}], \\ &\leq \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp + \bar{\mathbf{a}} + \bar{\mathbf{b}}], \\ &\leq \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp]. \end{aligned} \quad (6.12)$$

Combining (6.11) and (6.12) and observing that  $\boldsymbol{\mu} \in \mathcal{Q}$ , we have

$$\mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp] = \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp]. \quad (6.13)$$

To complete the proof, consider any feasible  $x, \mathbf{x}_r(\omega), \mathbf{y}(\omega)$ . By part (i) and  $\boldsymbol{\mu} \in \mathcal{Q}$ , we have

$$\mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_p^\sharp + \mathbf{W}_c^\sharp] \geq \mathbb{E}_{\boldsymbol{\mu}} [\mathbf{W}_p + \mathbf{W}_c] \geq \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} [\mathbf{W}_p + \mathbf{W}_c],$$

### 6.3. Some equivalences between social planner and equilibrium problems

where  $\mathbf{W}_p$  and  $\mathbf{W}_c$  are defined by (6.1). Thus (6.13) gives

$$\min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{W}_p^{\#} + \mathbf{W}_c^{\#}] \geq \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{W}_p + \mathbf{W}_c].$$

This shows that

$$(x^{\#}, \mathbf{x}_r^{\#}, \mathbf{y}^{\#}) \in \arg \max_{x, \mathbf{x}_r, \mathbf{y}} \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{W}_p + \mathbf{W}_c],$$

as required.  $\square$

*Remark 6.3.4.* Note that an equilibrium of RaEq-AD( $\check{\rho}$ ) consists of a price vector  $\boldsymbol{\pi}$ , giving one price per scenario, and a probability  $\boldsymbol{\mu}$  that is seen by both the producer and the consumer as a worst-case probability for the welfare plus trade evaluation.

*Remark 6.3.5.* In Section 6.4 we give an example of three risked equilibrium without Arrow-Debreu securities, each corresponding to a risk-neutral equilibrium with different measure  $\boldsymbol{\mu}(\omega)$ . However if Arrow-Debreu securities are included then two of these equilibria are no longer equilibria in a risk-averse setting. The risk-averse consumer, who without Arrow-Debreu securities had no mechanism to alter his outcomes will trade these securities to improve their risk-adjusted payoff.

*Remark 6.3.6.* Consider a set of prices  $\boldsymbol{\pi}$  that gives a risked equilibrium in which agent  $i$  has payoff  $\mathbf{W}_i(\boldsymbol{\pi})$  and risked payoff  $\rho_i(\mathbf{W}_i(\boldsymbol{\pi}))$ . Suppose that there exists a probability measure  $\mathbb{Q}^*$  such that  $\rho_i(\mathbf{W}_i(\boldsymbol{\pi})) = \mathbb{E}_{\mathbb{Q}^*}[\mathbf{W}_i(\boldsymbol{\pi})]$ . Observe that this does not imply that choosing actions  $x$  to maximize  $\mathbb{E}_{\mathbb{Q}^*}[\mathbf{W}_i(\boldsymbol{\pi})]$  will give  $\max_x \rho_i(\mathbf{W}_i(\boldsymbol{\pi}))$ . This is because  $x^*$  solves

$$\max_x \rho_i(\mathbf{W}_i(\boldsymbol{\pi})) = \max_x \min_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\mathbf{W}_i(\boldsymbol{\pi})],$$

and not

$$\max_x \mathbb{E}_{\mathbb{Q}^*}[f_i(x, \boldsymbol{\pi})],$$

since  $\mathbb{Q}^*$  depends on  $x$ .

*Remark 6.3.7.* Proposition 6.3.3 is easily extended to the case where the agents have different risk measures  $\rho_p$  and  $\rho_c$  with non-disjoint risk set. In this case, (6.12) becomes

$$\begin{aligned} \bar{\theta} + \bar{\varphi} &= \min_{\mathbb{Q}_p \in \mathcal{Q}_p} \mathbb{E}_{\mathbb{Q}_p}[\boldsymbol{\pi}(x^{\#} + \mathbf{x}_r^{\#}) + \mathbf{W}_p^{\#} + \bar{\mathbf{a}}] \\ &\quad + \min_{\mathbb{Q}_c \in \mathcal{Q}_c} \mathbb{E}_{\mathbb{Q}_c}[\mathbf{W}_c^{\#} - \boldsymbol{\pi} \mathbf{y}^{\#} + \bar{\mathbf{b}}], \\ &\leq \min_{\mathbb{Q} \in \mathcal{Q}_p \cap \mathcal{Q}_c} \mathbb{E}_{\mathbb{Q}}[\mathbf{W}_c^{\#} + \mathbf{W}_p^{\#}], \end{aligned} \tag{6.14}$$

and the social planner uses a risk measure with  $\mathcal{Q} = \mathcal{Q}_p \cap \mathcal{Q}_c$ .

The following proposition (Theorem 11 [Philpott et al. \(2016\)](#)) stands as a reverse statement for Proposition 6.3.3.

**Proposition 6.3.8.** *Let the elements  $x^{\#}$ ,  $\mathbf{x}_r^{\#}$  and  $\mathbf{y}_r^{\#}$  be optimal solutions to RaSp( $\check{\rho}$ ), with associated worst case probability measure  $\boldsymbol{\mu}$ . Then there exists prices  $\boldsymbol{\pi}$  such that the couple  $(\boldsymbol{\pi}, \boldsymbol{\mu})$  forms a risk trading equilibrium for RaEq-AD( $\check{\rho}$ ) with associated optimal solutions  $(x^{\#}, \mathbf{x}_r^{\#}, \mathbf{y}^{\#})$ .*

## 6. On risk averse competitive equilibrium

Combining Proposition 6.3.3 and Proposition 6.3.8, we are able to state the following result of uniqueness of equilibrium.

**Corollary 6.3.9.** *If both the producer's and consumer's criterion are strictly concave, and if each of the extreme points  $\mathbb{Q}_k$  charges all  $\omega$ , then  $\text{RaSp}(\check{\rho})$  admits a unique solution  $(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp)$ . Furthermore  $\text{RaEq-AD}(\check{\rho})$  admits unique optimal decisions  $(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp)$ . If, in addition, solving  $\text{RaSp}(\check{\rho})$  admit a unique worst case probability measure  $\boldsymbol{\mu}$ , then equilibrium prices  $(\boldsymbol{\pi}, \boldsymbol{\mu})$  are unique.*

*Proof.* As each of the extreme points  $\mathbb{Q}_k$  charges all  $\omega$ , the risk averse social planner problem is strictly convex with linear constraints. Thus there exists a unique solution  $(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp)$  attained for a worst case probability  $\boldsymbol{\mu}$ . Applying Proposition 6.3.8, we know that there exists  $\boldsymbol{\pi}$  such that  $(\boldsymbol{\pi}, \boldsymbol{\mu})$  forms a risk trading equilibrium. Suppose now that there exists two risk-trading equilibria  $(\boldsymbol{\pi}^1, \boldsymbol{\mu}^1, x^1, \mathbf{x}_r^1, \mathbf{y}^1)$  and  $(\boldsymbol{\pi}^2, \boldsymbol{\mu}^2, x^2, \mathbf{x}_r^2, \mathbf{y}^2)$ . Then, by Proposition 6.3.3, they both solve  $\text{RaSp}(\check{\rho})$  which admits a unique solution. Consequently, we have  $x^\sharp = x^1 = x^2$ ,  $\mathbf{x}_r^\sharp = \mathbf{x}_r^1 = \mathbf{x}_r^2$  and  $\mathbf{y}^\sharp = \mathbf{y}^1 = \mathbf{y}^2$ .

If in addition,  $\boldsymbol{\mu}^1 = \boldsymbol{\mu}^2$ , then by Corollary 6.3.2, we deduce that  $\boldsymbol{\pi}^1 = \boldsymbol{\pi}^2$  which ends the proof.  $\square$

We have shown a first equivalence between  $\text{RnSp}(\mathbb{P})$  and  $\text{RnEq}(\mathbb{P})$  and a second one between  $\text{RaSp}(\check{\rho})$  and  $\text{RaEq-AD}(\check{\rho})$ . These equivalences lead to uniqueness of equilibrium if there is uniqueness of the solution of the social planner. A natural question arises: if  $\text{RaSp}(\check{\rho})$  has a unique solution, is there a unique equilibrium for  $\text{RaEq}(\check{\rho})$ ? The next section provides a simple counterexample.

## 6.4. Multiple risk averse equilibrium

In this section, we present a toy problem where  $\text{RaSp}(\check{\rho})$  has a unique optimum but there are three different equilibria for  $\text{RaEq}(\check{\rho})$ . They are first found numerically using classical methods (PATH solver and a tâtonnement algorithm), then derived analytically. An interesting point is that the equilibrium found by PATH is unstable.

Let  $\Omega = \{1, 2\}$  and  $\mathcal{Q} = \text{conv}\left\{\left(\frac{1}{4}, \frac{3}{4}\right), \left(\frac{3}{4}, \frac{1}{4}\right)\right\}$ . For simplicity of notation index by  $i \in \{1, 2\}$  the realization of each random variable. We choose the following parameters:  $V_1 = 4$ ,  $V_2 = \frac{48}{5}$ ,  $c = \frac{23}{2}$ ,  $c_1 = 1$ ,  $c_2 = \frac{7}{2}$ ,  $r_1 = 2$ ,  $r_2 = 10$ .

### 6.4.1. Multiple equilibrium

#### PATH solver

First we look for equilibrium using GAMS with the solver PATH in the EMP framework (See Brook et al. (1988), Ferris et al. (2009) and Ferris and Munson (2000)). We have run GAMS from different starting points defined by a grid  $100 \times 100$  over the square  $[1.220; 1.255] \times [2.05; 2.18]$ . We always find an equilibrium defined by

$$\boldsymbol{\pi} = (\pi_1, \pi_2) = (1.23578; 2.10953),$$

leading to risked adjusted welfare (0.821; 2.134) for producer and consumer respectively.

### Walras tâtonnement

We now compute equilibria using a tâtonnement algorithm (See [Uzawa \(1960\)](#)).

```

Data: MAX-ITER,  $(\pi_1^0, \pi_2^0)$ ,  $\tau$ 
1 for  $k$  from 0 to MAX-ITER do
2   Compute an optimal decision for each player given a price :
3    $x, x_1, x_2 \in \arg \max \rho(\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r))$ ;
4    $y_1, y_2 \in \arg \max \rho(\mathbf{W}_c - \boldsymbol{\pi}\mathbf{y})$ ;
5   Update the price :
6    $\pi_1 = \pi_1 - \tau \max \{0, y_1 - (x + x_1)\}$ ;
7    $\pi_2 = \pi_2 - \tau \max \{0, y_2 - (x + x_2)\}$ ;
8 end
9 return  $(\pi_1, \pi_2)$ 

```

**Algorithm 2:** Walras tâtonnement

Running algorithm 2 starting from (1.25; 2.06), respectively from (1.22; 2.18), with 100 iterations and a step size of 0.1, we find two new equilibria:

$$\boldsymbol{\pi} = (1.2256; 2.0698) \text{ and } \boldsymbol{\pi} = (1.2478; 2.1564) ,$$

leading to risked-adjusted welfare for producer and consumer respectively

$$(0.797; 2.152) \text{ and } (0.845; 2.113) .$$

Notice that neither equilibrium dominates the other.

An alternative tâtonnement method called FastMarket (see [Facchinei and Kanzow \(2007\)](#)) finds the same equilibrium.

#### 6.4.2. Analytical results

We now compute the three equilibrium analytically. Details of the computation are in 6.5.

Consider two probabilities  $(\underline{p}, 1 - \underline{p})$  and  $(\bar{p}, 1 - \bar{p})$ . Given prices  $0 < \pi_1 < \pi_2$ , we solve the producer (resp. consumer) optimization problem. Optimal decisions are derived in 6.5.1 and summed up in Table 6.1 where  $x_c$  is given by

$$x_c(\boldsymbol{\pi}) = \frac{1}{2(\pi_1 - \pi_2)} \left( \frac{\pi_2^2}{c_2} - \frac{\pi_1^2}{c_1} \right) .$$

We see that there are three regimes, depending only on the prices  $(\pi_1, \pi_2)$ , of optimal first stage solutions. Case a) (resp. case c)), corresponds to a set of prices such that  $\mathbb{E}_{\bar{p}}[\mathbf{W}_p] < \mathbb{E}_{\underline{p}}[\mathbf{W}_p]$  (resp.  $\mathbb{E}_{\bar{p}}[\mathbf{W}_p] > \mathbb{E}_{\underline{p}}[\mathbf{W}_p]$ ), and the optimal decision corresponds to an optimal risk-neutral decision with respect to one of the two extreme points of  $\mathcal{Q}$ . On the other hand, case b) corresponds to a set of prices such that the expected welfare is equivalent for all probability in  $\mathcal{Q}$ , i.e.  $\mathbb{E}_{\bar{p}}[\mathbf{W}_p] = \mathbb{E}_{\underline{p}}[\mathbf{W}_p]$ . In Figure 6.1, the red area corresponds to case a), the blue to case b) and the red

## 6. On risk averse competitive equilibrium

	condition	$x^\sharp$	$x_i^\sharp$	$y_i^\sharp$
case a)	$x_c \leq \frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}$	$\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$
case b)	$\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c} \leq x_c \leq \frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}$	$x_c$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$
case c)	$\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c} \leq x_c$	$\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$

Table 6.1.: Optimal control for producer and consumer problems

to case c), separated by black lines of equations  $\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c} = x_c(\boldsymbol{\pi})$  and  $\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c} = x_c(\boldsymbol{\pi})$  respectively.

We are now looking for prices  $(\pi_1, \pi_2)$  such that the complementarity constraints are satisfied. For strictly positive prices, these constraints can be summed up as

$$z_i(\boldsymbol{\pi}) = x^\sharp(\boldsymbol{\pi}) + x_i^\sharp(\boldsymbol{\pi}) - y_i^\sharp(\boldsymbol{\pi}) = 0, \quad i \in \{1, 2\}.$$

Accordingly we define excess supply functions  $z_i^l$  for case  $l \in \{a, b, c\}$ , and  $i \in \{1, 2\}$ . The red, blue and green lines corresponds to manifolds of null excess supply function for scenario  $i$ , that is of prices such that  $z_i^l(\pi_1, \pi_2) = 0$ . When the lines cross we have  $z_i^l = z_i^2 = 0$ , and thus we have candidate equilibrium. If the lines cross in the area of the same color we have an equilibrium. This is the case with the parameters chosen, and equilibrium can be derived in exact arithmetic.

We end with a few remarks derived from this example.

*Remark 6.4.1.* The PATH solver finds the blue equilibrium, Algorithm 2 finds the green and the red equilibrium as illustrated by Figure 6.2. Interestingly it can be shown that the blue equilibrium is unstable in the sense that the dynamical system driven by  $\pi' = z(\pi)$  is not locally stable (see Samuelson (1941)) around the blue equilibrium (see 6.6).

*Remark 6.4.2.* No equilibrium dominates another: if going from one equilibrium to another increases the (risk-adjusted) welfare of one agent, then it decreases the (risk-adjusted) welfare of the other.

*Remark 6.4.3.* Using the analytical results we check that there exists a set of non-zero Lebesgue measure of parameters  $V_1, V_2, c, c_1, c_2, r_1$ , and  $r_2$  (albeit small), that have three distinct equilibria with the same properties.

*Remark 6.4.4.* We can show that the blue equilibrium is a convex combination of red and green equilibrium, illustrated on Figure 6.1 by the dashed blue line.

## Acknowledgments

The first-named author want to thank French embassy of New-Zealand for their administrative help and for the financial support thanks to France–New-Zealand friendship fund.

The authors want to thank PGMO programs for their financial support.

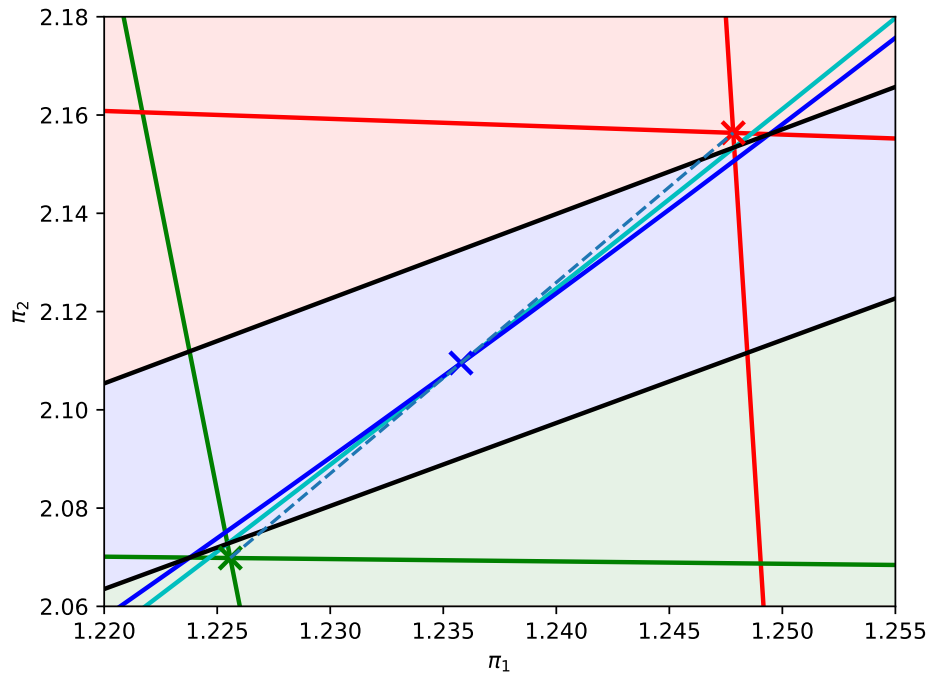


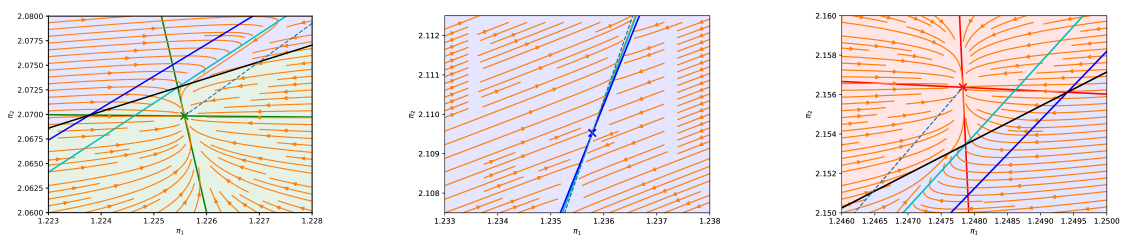
Figure 6.1.: Null excess function per scenario manifold for  $V_1 = 4$ ,  $V_2 = \frac{48}{5}$ ,  $c = \frac{23}{2}$ ,  $c_1 = 1$ ,  $c_2 = \frac{7}{2}$ ,  $r_1 = 2$ ,  $r_2 = 10$ .

## 6.5. Analytical results

We first analyse the best responses of the producer and the consumer given a price  $\pi$ . Then, we deduce conditions on the price and find equilibrium prices.

### 6.5.1. Parametric solution with respect to $\pi$

Assume without loss of generality that  $0 < \pi_1 < \pi_2$ .



(a) around green equilibrium.      (b) around blue equilibrium.      (c) around red equilibrium.

Figure 6.2.: Representation of vector field  $\pi' = z(\pi)$



## 6. On risk averse competitive equilibrium

### Statement of consumer's problem

The consumer solves one problem per scenario  $\omega_i$ ,  $i = 1, 2$ . Let  $V_1$ ,  $V_2$ ,  $r_1$  and  $r_2$  be strictly positive constants. The consumer problem for  $\omega_i$  is

$$\min_{y_i} \quad \pi_i y_i - V_i y_i + \frac{1}{2} r_i y_i^2 .$$

### Statement of producer's problem

The risk aversion of the producer is represented by a coherent risk measure  $\rho$  with risk set  $\mathcal{Q}$ . Then the producer problem reads

$$\min_{x \geq 0, \mathbf{x}_r \geq 0} \quad \rho \left[ C(x) + \mathbf{C}_r(\mathbf{x}_r) - \boldsymbol{\pi}(x + \mathbf{x}_r) \right] .$$

Note that in the case of two outcomes the probability  $\mathbb{P}$  measure can be defined by  $\mathbb{P}(\omega_1)$ , which we denote  $p$ . Hence the probability set  $\mathcal{P}$  can be described by an interval  $[\underline{p}, \bar{p}]$ .

Then the producer problem reads

$$\min_{x \geq 0, x_1 \geq 0, x_2 \geq 0} \quad \frac{1}{2} c x^2 + \max_{p \in [\underline{p}, \bar{p}]} \left\{ p \left( \frac{c_1 x_1^2}{2} - \pi_1(x + x_1) \right) \right. \quad (6.15)$$

$$\left. + (1 - p) \left( \frac{c_2 x_2^2}{2} - \pi_2(x + x_2) \right) \right\} \quad (6.16)$$

### Statement of complementary constraints

The complementary constraint states that a feasible solution is a solution where production is greater than demand for each scenario  $\omega \in \Omega$ . Moreover, we want equality between production and demand at equilibrium. These constraints are written

$$0 \leq (x + \mathbf{x}_r(\omega)) - \mathbf{y}(\omega) \perp \boldsymbol{\pi}(\omega) \geq 0 . \quad (6.17)$$

### Analytic solution of the producer's problem

Focusing on the second stage problem of (6.15) we have

$$Q^{(\boldsymbol{\pi})}(x) = \max_{p \in [\underline{p}, \bar{p}]} \quad p \min_{x_1 \geq 0} \left\{ \frac{c_1 x_1^2}{2} - \pi_1(x + x_1) \right\} \\ + (1 - p) \min_{x_2 \geq 0} \left\{ \frac{c_2 x_2^2}{2} - \pi_2(x + x_2) \right\} .$$

Note that for  $i \in \{1, 2\}$   $c_i > 0$ , hence we have  $x_i^\# = \frac{\pi_i}{c_i}$  which in turn gives

$$Q^{(\boldsymbol{\pi})}(x) = \max_{p \in [\underline{p}, \bar{p}]} \quad -p \left( \frac{\pi_1^2}{2c_1} + \pi_1 x \right) - (1 - p) \left( \frac{\pi_2^2}{2c_2} + \pi_2 x \right) \quad (6.19) \\ = \max_{p \in [\underline{p}, \bar{p}]} \quad p \left( \left( \frac{\pi_2^2}{2c_2} - \frac{\pi_1^2}{2c_1} \right) + (\pi_2 - \pi_1)x \right) - \left( \frac{\pi_2^2}{2c_2} + \pi_2 x \right) .$$

	condition	$x^\sharp$	$x_i^\sharp$	$y_i^\sharp$
case a)	$x_c \leq \frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}$	$\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$
case b)	$\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c} \leq x_c \leq \frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}$	$x_c$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$
case c)	$\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c} \leq x_c$	$\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}$	$\frac{\pi_i}{c_i}$	$\frac{V_i - \pi_i}{r_i}$

Table 6.2.: Optimal control for producer and consumer problems

Defining

$$x_c(\boldsymbol{\pi}) = \frac{-1}{\pi_2 - \pi_1} \left( \frac{\pi_2^2}{2c_2} - \frac{\pi_1^2}{2c_1} \right), \quad (6.20)$$

we see that the worst case probability is given by

$$p^\sharp(\boldsymbol{\pi}) = \begin{cases} \bar{p} & \text{if } x > x_c(\boldsymbol{\pi}), \\ \underline{p} & \text{if } x < x_c(\boldsymbol{\pi}), \\ \text{any } p \in [\underline{p}, \bar{p}] & \text{if } x = x_c(\boldsymbol{\pi}), \end{cases}$$

and thus Equation (6.19) yields

$$Q^{(\boldsymbol{\pi})}(x) = \begin{cases} -\mathbb{E}_{\bar{p}} \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x \right] & \text{if } x \geq x_c(\boldsymbol{\pi}), \\ -\mathbb{E}_p \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x \right] & \text{if } x < x_c(\boldsymbol{\pi}). \end{cases}$$

Now the first stage problem (Problem (6.15)) reads

$$\min_{x \geq 0} \frac{1}{2}cx^2 - \mathbb{E}_{\bar{p}} \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x \right] \mathbf{1}_{x \geq x_c} - \mathbb{E}_p \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x \right] \mathbf{1}_{x < x_c}.$$

We have

$$\min_{x \geq x_c} \frac{1}{2}cx^2 + Q^{(\boldsymbol{\pi})}(x) = \begin{cases} -\frac{1}{2c} \mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]^2 - \mathbb{E}_{\bar{p}} \left[ \frac{\boldsymbol{\pi}^2}{2c_r} \right] & \text{if } x_c \leq \frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}, \\ \frac{1}{2}cx_c^2 - \mathbb{E}_{\bar{p}} \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x_c \right] & \text{if } \frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c} \leq x_c, \end{cases}$$

attained at  $\frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c}$  and  $x_c$  respectively.

If  $x_c > 0$  we also have

$$\min_{0 \leq x \leq x_c} \frac{1}{2}cx^2 + Q^{(\boldsymbol{\pi})}(x) = \begin{cases} \frac{1}{2}cx_c^2 - \mathbb{E}_{\bar{p}} \left[ \frac{\boldsymbol{\pi}^2}{2c_r} + \boldsymbol{\pi}x_c \right] & \text{if } x_c \leq \frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}, \\ -\frac{1}{2c} \mathbb{E}_p[\boldsymbol{\pi}]^2 - \mathbb{E}_p \left[ \frac{\boldsymbol{\pi}^2}{2c_r} \right] & \text{if } \frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c} \leq x_c, \end{cases}$$

attained at  $x_c$  and  $\frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c}$  respectively. If  $x_c \leq 0$  the solution given earlier holds.

Recall that  $\mathbb{E}_p[\boldsymbol{\pi}] \leq \mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]$ , thus the optimal solution can be summed up in Table 6.2

## 6. On risk averse competitive equilibrium

### 6.5.2. Finding price equilibrium

Looking at Table 6.2 we see that there are three regimes, depending only on the prices  $(\pi_1, \pi_2)$  of optimal first stage solutions. We are now looking for prices  $(\pi_1, \pi_2)$  such that the complementarity constraint (6.17) is satisfied. For strictly positive prices, this constraint can be summed up as

$$z_i(\boldsymbol{\pi}) = x^\sharp(\boldsymbol{\pi}) + x_1^\sharp(\boldsymbol{\pi}) - y_i^\sharp(\boldsymbol{\pi}) = 0, \quad i \in \{1, 2\}. \quad (6.21)$$

To go further we are going to split cases by defining the auxiliary excess demand function

$$\begin{aligned} z_i^a(\boldsymbol{\pi}) &= \frac{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]}{c} + \frac{\pi_i}{c_i} - \frac{V_i - \pi_i}{r_i}, \\ z_i^b(\boldsymbol{\pi}) &= x_c(\boldsymbol{\pi}) + \frac{\pi_i}{c_i} - \frac{V_i - \pi_i}{r_i}, \\ z_i^c(\boldsymbol{\pi}) &= \frac{\mathbb{E}_p[\boldsymbol{\pi}]}{c} + \frac{\pi_i}{c_i} - \frac{V_i - \pi_i}{r_i}, \end{aligned}$$

such that we have

$$z = z^a \mathbb{1}_{cx_c(\boldsymbol{\pi}) \leq \mathbb{E}_{\bar{p}}[\boldsymbol{\pi}]} + z^b \mathbb{1}_{\mathbb{E}_{\bar{p}}[\boldsymbol{\pi}] \leq cx_c(\boldsymbol{\pi}) \leq \mathbb{E}_p[\boldsymbol{\pi}]} + z^c \mathbb{1}_{\mathbb{E}_p[\boldsymbol{\pi}] \leq cx_c(\boldsymbol{\pi})}.$$

#### Case a and c

The set of prices such that  $z_i^a(\boldsymbol{\pi}) = 0$  are lines given by

$$\begin{aligned} \pi_2 &= \frac{cc_1V_1 - (c_1r_1\bar{p} + c(r_1 + c_1))\pi_1}{c_1r_1(1 - \bar{p})}, \\ \pi_2 &= \frac{cc_2V_2 - c_2r_2\bar{p}\pi_1}{c_2r_2(1 - \bar{p}) + c(r_2 + c_2)}, \end{aligned}$$

and the equilibrium can be found by solving the linear system. Case c is similar, substituting  $\bar{p}$  by  $p$ .

#### Case b

The set of prices such that  $z_i^b(\boldsymbol{\pi}) = 0$  are an ellipsoid and an hyperbola given by

$$\begin{aligned} \frac{1}{\pi_1 - \pi_2} \left( \frac{\pi_2^2}{2c_2} - \frac{\pi_1^2}{2c_1} \right) + \frac{\pi_1}{c_1} - \frac{V_1 - \pi_1}{r_1} &= 0, \\ \frac{1}{\pi_1 - \pi_2} \left( \frac{\pi_2^2}{2c_2} - \frac{\pi_1^2}{2c_1} \right) + \frac{\pi_2}{c_2} - \frac{V_2 - \pi_2}{r_2} &= 0, \end{aligned}$$

whose affine equations read

$$\begin{aligned} \frac{\pi_2^2}{2c_2} - \left( \frac{1}{c_1} + \frac{1}{r_1} \right) \pi_1 \pi_2 + \left( \frac{1}{r_1} + \frac{1}{2c_1} \right) \pi_1^2 + (\pi_2 - \pi_1) \frac{V_1}{r_1} &= 0, \\ \left( \frac{1}{r_2} + \frac{1}{2c_2} \right) \pi_2^2 - \left( \frac{1}{c_2} + \frac{1}{r_2} \right) \pi_1 \pi_2 + \frac{1}{2c_1} \pi_1^2 - (\pi_2 - \pi_1) \frac{V_2}{r_2} &= 0. \end{aligned}$$

## 6.6. Unstability of equilibrium

**Definition 6.6.1.** Let  $\boldsymbol{\pi}(t)$  be the general solution of the differential equation

$$\boldsymbol{\pi}' = z(\boldsymbol{\pi}), \quad (6.23)$$

such that  $\boldsymbol{\pi}(0) = \boldsymbol{\pi}_0$ . An equilibrium  $\boldsymbol{\pi}^\sharp$  such that  $z(\boldsymbol{\pi}) = 0$  is said to be *locally stable* if for all  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$\|\boldsymbol{\pi}_0 - \boldsymbol{\pi}^\sharp\| < \delta \Rightarrow \|\boldsymbol{\pi}(t) - \boldsymbol{\pi}^\sharp\| < \epsilon, \quad \forall t > 0. \quad (6.24)$$

Using results from the field of Ordinary Differential Equations (see [Mattheij and Molenaar \(2002\)](#)), the local stability can be determined from studying the linearization of the system around the equilibrium point.

**Proposition 6.6.2.** *Let  $\boldsymbol{\pi}^\sharp$  be an equilibrium point. Let  $A$  be the Jacobian matrix of  $z(\boldsymbol{\pi})$  at point  $\boldsymbol{\pi}^\sharp$ . Then  $\boldsymbol{\pi}^\sharp$  is stable if and only both real parts of eigenvalues of  $A$  are strictly positive.*

Computing matrix  $A$  and its eigenvalues in exact arithmetic (using Maxima), we find that the blue equilibrium is unstable and that green and red equilibria are stable.

## 6.7. Conclusion

In Section 6.2 we have presented equilibrium and optimization models in the risk neutral case and the risk averse case. In Section 6.3 we have given links between equilibrium and optimization problems in the risk neutral and complete risk-averse cases. Finally, in Section 6.4 we have detailed a simple example that displays multiple equilibria in the incomplete risk-averse case.



# 7. Equilibrium and player consistency

## Contents

---

7.1. Introduction . . . . .	126
7.2. Player consistency and equilibrium with risk neutral players . . . . .	126
7.3. Player consistency and equilibrium with risk averse players . . . . .	128
7.4. Conclusion . . . . .	128

---

## 7.1. Introduction

We have seen in Chap. 6 that prices and Arrow-Debreu securities are suitable tools to get uniqueness of equilibrium. In this chapter, we revisit the toy problem presented in Chapter 6 at the light of the framework of Player Consistency presented in Sect. 5.2. We invite the reader to refer to Sect. 6.2 even if we recall notations to remain self-contained.

Consider a two time-step single-settlement market for one good. In a single settlement market, the producer can arrange in advance for a production of  $x$  at a marginal cost  $cx$  as a first-step decision, and choose the value of a recourse variable  $\mathbf{x}_r$  incurring an uncertain marginal cost  $\mathbf{c}_r \mathbf{x}_r$ . We assume that there are a finite number of scenarios  $\omega \in \Omega$  determining the coefficient  $\mathbf{c}_r(\omega)$ .

The product is purchased in the second step by a consumer with a utility function  $\mathbf{V}(\omega)\mathbf{y}(\omega) - \frac{1}{2}\mathbf{r}(\omega)\mathbf{y}^2(\omega)$ . The consumer has no first-stage decision, and the amount purchased  $\mathbf{y}(\omega)$  depends on the scenario.

We also recall that the vectorial order is the pointwise order. Decisions  $x$ ,  $\mathbf{x}_r(\omega)$  and  $\mathbf{y}(\omega)$  can be made to maximize a social objective. We denote by

$$\mathbf{W}_p(\omega) = -\frac{1}{2}cx^2 - \frac{1}{2}\mathbf{c}_r(\omega)\mathbf{x}_r(\omega)^2, \quad \forall \omega \in \Omega, \quad (7.1a)$$

the welfare of the producer, and by

$$\mathbf{W}_c(\omega) = \mathbf{V}(\omega)\mathbf{y}(\omega) - \frac{1}{2}\mathbf{r}(\omega)\mathbf{y}(\omega)^2, \quad \forall \omega \in \Omega, \quad (7.1b)$$

the welfare of the consumer. Then the welfare of the social planner can be defined by

$$\mathbf{W}_{sp} = \mathbf{W}_p + \mathbf{W}_c. \quad (7.2)$$

Following the idea of making connections with the framework of player consistency presented in Sect. 5.2, we introduce the sets

$$\mathbb{H} = \mathbb{R} \times \mathbb{R}^{|\Omega|} \text{ and } \mathbb{T} = \mathbb{R}^{|\Omega|}. \quad (7.3)$$

## 7.2. Player consistency and equilibrium with risk neutral players

As we focus on the question of time consistency, we introduce the mappings  $A$ ,  $F_{\mathbb{H}}$  and  $F_{\mathbb{T}}$  defined by

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{R}, \quad (7.4a)$$

$$((x, \mathbf{x}_r), \mathbf{y}) \mapsto \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p + \mathbf{W}_c], \quad (7.4b)$$

$$F_{\mathbb{H}}^{\pi} : \mathbb{H} \rightarrow \mathbb{R}, \quad (7.4c)$$

$$(x, \mathbf{x}_r) \mapsto \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p + \pi(x + \mathbf{x}_r)], \quad (7.4d)$$

$$F_{\mathbb{T}}^{\pi} : \mathbb{T} \rightarrow \mathbb{R}, \quad (7.4e)$$

$$\mathbf{y} \mapsto \mathbb{E}_{\mathbb{P}}[\mathbf{W}_c - \pi\mathbf{y}], \quad (7.4f)$$

## 7.2. Player consistency and equilibrium with risk neutral players

and we denote by  $(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp)$  the solutions of the optimization problem of the social planner

$$\max_{x, \mathbf{x}_r, \mathbf{y}} A((x, \mathbf{x}_r), \mathbf{y}) , \quad (7.5a)$$

$$\text{s.t. } x + \mathbf{x}_r = \mathbf{y} . \quad (7.5b)$$

Looking at Equation (7.4), it is not obvious if there exists a mapping  $S^{A,F}$  such that  $A = S^{A, F_{\mathbb{H}}, F_{\mathbb{T}}} \circ (F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi)$  or in other words, it is not clear if the triplet  $(A, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi)$  is Player Consistent.

We now interpret the prices  $\pi$  as a tool to build agent consistency. For this purpose, we introduce the mapping  $\tilde{A}^\pi$  by

$$\tilde{A}^\pi : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{R} , \quad (7.6a)$$

$$(x, \mathbf{x}_r, \mathbf{y}) \mapsto \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p + \mathbf{W}_c + \pi(x + \mathbf{x}_r - \mathbf{y})] . \quad (7.6b)$$

Let us notice two facts.

1. First, the triplet  $(\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi)$  is Player Consistent for any price  $\pi$ .
2. Second, there exists a system of prices  $\pi^\sharp$  such that, when trying to maximize welfares,  $A$  and  $\tilde{A}^{\pi^\sharp}$  coincides at optimum i.e.

$$\tilde{A}^{\pi^\sharp}(\tilde{x}^\sharp, \tilde{\mathbf{x}}_r^\sharp, \tilde{\mathbf{y}}^\sharp) = A(x^\sharp, \mathbf{x}_r^\sharp, \mathbf{y}^\sharp) . \quad (7.7)$$

We now state our proposition that links Player Consistency and Equilibrium in the risk neutral case

### Proposition 7.2.1.

1. For all prices  $\pi$ , the triplet  $(\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi)$  in (7.4) and in (7.6) is Strong Player Consistent and the subaggregator  $S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}$  has the following form

$$S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}(\alpha, \beta) = \alpha + \beta . \quad (7.8)$$

2. If there exists equilibrium prices  $\tilde{\pi}$ , the mapping  $A$  and  $\tilde{A}^{\tilde{\pi}}$  coincides at optimum, i.e. Equation (7.7) is satisfied.

*Proof.* By Definition 5.10, the subaggregator  $S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}$  is equal to

$$S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}(\alpha, \beta) = \{ \tilde{A}^\pi(x, \mathbf{x}_r, \mathbf{y}) \mid F_{\mathbb{H}}(x, \mathbf{x}_r) = \alpha , F_{\mathbb{T}}(\mathbf{y}) = \beta \} . \quad (7.9)$$

Using Equations (7.4) and (7.6) this rewrites

$$S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}(\alpha, \beta) = \alpha + \beta . \quad (7.10)$$

The subaggregator  $S^{\tilde{A}^\pi, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi}$  is hence an mapping increasing in both arguments. By Proposition 5.2.7, the triplet  $(A, F_{\mathbb{H}}^\pi, F_{\mathbb{T}}^\pi)$  is thus Strong Player Consistent.

Last assertion is a direct application of Proposition 6.3.1. This ends the proof.  $\square$



### 7.3. Player consistency and equilibrium with risk averse players

In the risk neutral case, linearity of the expectation plays a crucial role. We are now dealing with convex risk measures but the question remains the same. Assuming that each player and the social planner are endowed with their own risk measures, does there exist perturbations such that perturbed criterion of the social planner is an aggregation of perturbed criteria of each player ?

Let  $\rho_{sp}$ <sup>1</sup>,  $\rho_p$  and  $\rho_c$  be the risk measures of the social planner, the producer and the consumer respectively.

We introduce the mappings  $A, F_{\mathbb{H}}$  and  $F_{\mathbb{T}}$  defined by

$$A : \mathbb{H} \times \mathbb{T} \rightarrow \mathbb{R} , \quad (7.11a)$$

$$(x, \mathbf{x}_r, \mathbf{y}) \mapsto \rho_{sp}(\mathbf{W}_p + \mathbf{W}_c) , \quad (7.11b)$$

$$F_{\mathbb{H}}^{\pi} : \mathbb{H} \rightarrow \mathbb{R} , \quad (7.11c)$$

$$(x, \mathbf{x}_r) \mapsto \rho_p(\mathbf{W}_p + \boldsymbol{\pi}(x + \mathbf{x}_r)) , \quad (7.11d)$$

$$F_{\mathbb{T}}^{\pi} : \mathbb{T} \rightarrow \mathbb{R} , \quad (7.11e)$$

$$\mathbf{y} \mapsto \rho_c(\mathbf{W}_c - \boldsymbol{\pi}\mathbf{y}) , \quad (7.11f)$$

Looking at Equation (7.11), it is not obvious if there exists a mapping  $S^{A,F}$  such that  $A = S^{A,F_{\mathbb{H}},F_{\mathbb{T}}} \circ (F_{\mathbb{H}}^{\pi}, F_{\mathbb{T}}^{\pi})$ . Finding good properties on risk measures to get Consistency is still an open question.

### 7.4. Conclusion

In this section, we have made some connections between the framework of Player Consistency presented in Sect. 5.2 and Equilibrium problems presented in Chapter 6.

This allows us to interpret, in the risk neutral case, equilibrium prices, as tools to build player consistency.

---

<sup>1</sup> As said in Chapter 6, by linearity of expectation we have  $\mathbb{E}_{\mathbb{P}}[\mathbf{W}_p + \mathbf{W}_c] = \mathbb{E}_{\mathbb{P}}[\mathbf{W}_p] + \mathbb{E}_{\mathbb{P}}[\mathbf{W}_c]$  hence the criterion of the social planner is natural, which is not the case anymore with risk-aversion. The social planner criterion could be either  $\rho_{sp}(\mathbf{W}_p + \mathbf{W}_c)$  or  $\rho_{sp}(\mathbf{W}_p) + \rho_{sp}(\mathbf{W}_c)$ . The first is more natural as it makes economic sense to add welfares.

## **Part IV.**

### **Risk averse classification and estimation problems**



# 8. General risk measures for robust machine learning

## Contents

---

<b>8.1. Introduction</b>	<b>132</b>
<b>8.2. Problem statement</b>	<b>133</b>
<b>8.3. Convex formulation of robust inference problems using risk measures</b>	<b>134</b>
8.3.1. Definition and properties of a risk measure	134
8.3.2. Convex reformulation	136
<b>8.4. Examples of risks measures</b>	<b>138</b>
8.4.1. Perspective functions and divergences	138
8.4.2. Divergence penalty functions	141
8.4.3. Constrained formulations	142
<b>8.5. Numerical solution</b>	<b>146</b>
8.5.1. A unifying formulation	146
8.5.2. Description of the algorithm	146
<b>8.6. Simulation example</b>	<b>148</b>
<b>8.7. Conclusion</b>	<b>151</b>

---

## 8. General risk measures for robust machine learning

A wide array of machine learning problems are formulated as the minimization of the expectation of a convex loss function on some parameter space. Since the probability distribution of the data of interest is usually unknown, it is often estimated from training sets, which may lead to poor out-of-sample performance. In this work, we bring new insights in this problem by using the framework which has been developed in quantitative finance for risk measures. We show that the original min-max problem can be recast as a convex minimization problem under suitable assumptions. We discuss several important examples of robust formulations, in particular by defining ambiguity sets based on  $\varphi$ -divergences and the Wasserstein distance. We also propose an efficient algorithm for solving the corresponding convex optimization problems involving complex convex constraints. Through simulation examples, we demonstrate that this algorithm scales well on real data sets.

### 8.1. Introduction

In machine learning, the robustness of the solutions obtained for classification and prediction tasks remains a main issue. In [Papernot, McDaniel, and Goodfellow \(2016\)](#) and [Kurakin, Goodfellow, and Bengio \(2016\)](#), some examples are provided where small modifications of the input data can completely change the resulting solution.

This kind of problems also occurs in optimal control when there exist uncertainties on parameters. In [Ben-Tal and Nemirovski \(2000\)](#), the authors showed that a small perturbation on the parameters can turn a feasible solution into an infeasible one.

In this context, robust approaches appear as a way of controlling out-of-sample performance. There is an extensive literature dealing with robust problems and the reader is referred to [Ben-Tal, El Ghaoui, and Nemirovski \(2009\)](#) for a survey. One of the main approaches consists of introducing constraints on the probability distribution of the unknown data. Under some conditions, this approach is equivalent to deal with ambiguity sets or a modified loss function. The works in [Ben-Tal, Den Hertog, De Waegenare, Melenberg, and Rennen \(2013\)](#); [Hu and Hong \(2013\)](#); [Duchi, Glynn, and Namkoong \(2016\)](#); [Moghaddam and Mahlooji \(2016\)](#) and [Namkoong and Duchi \(2016\)](#) have brought more insight on ambiguity sets. In [Esfahani and Kuhn \(2015\)](#) and [Esfahani, Shafieezadeh-Abadeh, Hanasusanto, and Kuhn \(2017\)](#), the authors present a distributionally robust optimization framework based on the Wasserstein distance. A set of probability distributions is defined as a ball centered on the reference probability with respect to the Wasserstein distance, then the optimization is carried out for the worst cost over this probability set.

This idea of minimizing the worst cost over a given probability set is well-known in quantitative finance. The robust representation of risk measures provides a theoretical framework to do so. A good class of risk measures is the class of coherent ones which were introduced in the seminal paper by [Artzner, Delbaen, Eber, and Heath \(1999\)](#). In [Föllmer and Schied \(2016\)](#), a broader class of so-called convex risk measures was investigated, for which a large number of results were established.

In this paper, we follow the line of [Esfahani and Kuhn \(2015\)](#), which aims at reformulating robust problems using ambiguity sets as convex minimization problems. Our contribution is threefold. First we clarify the links existing between risk

measures and robust optimization. This allows us to transpose results from finance to machine learning. Second, we propose a unifying convex optimization setting for dealing with various risk measures, including those based on  $\varphi$ -divergences or the Wasserstein distance. Finally, we propose an accelerated algorithm grounded on the subgradient projection method proposed in [Combettes \(2003\)](#). We show that the proposed algorithm is able to solve efficiently large-scale robust problems.

The organization of the paper is as follows. In Section 8.2, we state the general problem we investigate in the context of machine learning. In Section 8.3, we first establish a parallel between this problem and convex monetary risk measures. We then provide a convex reformulation of the problem. In Section 8.4, we discuss some important classes of risk measures by revisiting some of the results in the literature. In Section 8.5, we describe our algorithm for solving convex formulations of robust problems. Then, in Section 8.6, we illustrate the good performance of the proposed algorithm through numerical experiments on real datasets. Finally, short conclusions are drawn in Section 8.7.

## 8.2. Problem statement

Let  $(\Omega, \mathcal{T}, \mathbb{P})$  be the underlying probability space where  $\Omega$  is a finite set of cardinal  $N$ ,  $\mathcal{T}$  is the  $\sigma$ -field generated by  $(\{\omega\})_{\omega \in \Omega}$ , and  $\mathbb{P}$  is a probability distribution that is assumed to charge all points. Let  $d$  be a nonzero integer and let  $\mathbf{z}: \Omega \rightarrow \mathbb{R}^d$  denote a general random variable. Note that function  $\mathbf{z}$  can be identified with a matrix in  $\mathbb{R}^{N \times d}$  where, for every  $i \in \llbracket 1, N \rrbracket$ ,  $z_i$  denote the  $i$ -th line of matrix  $\mathbf{z}$ . We denote by  $\mathcal{M}_1$  the set of probability distributions over  $(\Omega, \mathcal{T}, \mathbb{P})$ .

For every  $i \in \llbracket 1, N \rrbracket$ , let  $\ell(\cdot, z_i): \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a loss function which is assumed to be lower semicontinuous (lsc) and convex such that

$$\bigcap_{i=1}^N \text{dom}(\ell(\cdot, z_i)) \neq \emptyset. \quad (8.1)$$

In standard formulations of machine learning problems, one aims at finding an optimal regression vector  $\bar{\theta} \in \mathbb{R}^n$  such that

$$\bar{\theta} \in \arg \min_{\theta \in \mathbb{R}^n} \sum_{i=1}^N \mathbb{P}_i \ell(\theta, z_i), \quad (8.2)$$

using identifying the probability distribution  $\mathbb{P}$  with a vector of  $N$  components  $\mathbb{P} = (\mathbb{P}_i)_{1 \leq i \leq N}$ . Indeed, setting  $z_i = [x_i^\top \ y_i]^\top$  with  $n = d - 1$ ,  $x_i \in \mathbb{R}^n$ , and  $y_i \in \mathbb{R}$  allows us to recover a wide array of estimation and classification problems. For example, penalized least squares regression problems are obtained when

$$(\forall i \in \llbracket 1, N \rrbracket)(\forall \theta \in \mathbb{R}^n) \quad \ell(\theta, z_i) = \frac{1}{2} \|y_i - x_i^\top \theta\|^2 + \rho(\theta), \quad (8.3)$$

where  $\rho: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, lsc, convex penalty function. If the random variable  $\mathbf{y}$  is  $\{0, 1\}$ -valued, we can recover binary classification problems, for example by performing a logistic regression, i.e.

$$(\forall i \in \llbracket 1, N \rrbracket)(\forall \theta \in \mathbb{R}^n) \quad \ell(\theta, z_i) = \log \left( 1 + \exp(-y_i x_i^\top \theta) \right). \quad (8.4)$$

## 8. General risk measures for robust machine learning

One of the main limitations of this formulation is that it assumes that the true probability distribution of the data is perfectly known. In practice, this distribution is often estimated empirically from the available observations.

In this paper, we will focus on the following more general robust formulation to determine an optimal regression vector.

**Problem 1.** Let  $\alpha: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lsc convex penalty function whose domain is a nonempty subset of  $\mathcal{M}_1$ . We want to find

$$\bar{\theta} \in \arg \min_{\theta \in \mathbb{R}^N} \sup_{\mathbb{Q}=(\mathbb{Q}_i)_{1 \leq i \leq N} \in \mathcal{M}_1} \left( \sum_{i=1}^N q_i \ell(\theta, z_i) - \alpha(\mathbb{Q}) \right), \quad (8.5)$$

In this problem, if  $\alpha$  is the indicator function  $\iota_{\{\mathbb{P}\}}$ <sup>1</sup> of the singleton containing the probability distribution  $\mathbb{P}$ , then (8.2) is recovered. More generally, if  $\alpha$  is equal to the indicator function of a nonempty closed convex set  $\mathcal{Q} \subset \mathcal{M}_1$ , then the objective function in (8.5) reduces to

$$\sup_{q \in \mathcal{Q}} \sum_{i=1}^N q_i \ell(\theta, z_i) = \sigma_{\mathcal{Q}}(\ell(\theta, \mathbf{z})), \quad (8.6)$$

where  $\sigma_{\mathcal{Q}}$  is the support function of  $\mathcal{Q}$ . This corresponds to the well-known case of distributionally robust optimization using ambiguity sets .

### 8.3. Convex formulation of robust inference problems using risk measures

In this section, we address Problem 1 in the light of the financial framework for monetary risk measures. We first recall known properties of risk measures and then show how Problem 1 can be reformulated as a convex problem.

#### 8.3.1. Definition and properties of a risk measure

Let  $\mathbb{X}$  be the space of real-valued random variables defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We denote by  $\mathbf{X}$  a generic element of  $\mathbb{X}$  and we recall that  $\mathbb{P}$  is assumed to be a distribution that charges all points. The space  $\mathbb{X}$  is endowed with the pointwise order  $\leq$ , that is,

$$(\forall (\mathbf{X}, \mathbf{Y}) \in \mathbb{X}^2) \quad \mathbf{X} \leq \mathbf{Y} \Leftrightarrow (\forall \omega \in \Omega) \quad \mathbf{X}(\omega) \leq \mathbf{Y}(\omega). \quad (8.7)$$

A *risk measure*  $\rho$  is a real-valued function  $\rho: \mathbb{X} \rightarrow \mathbb{R}$ .

The next four properties of risk measures were first introduced in Artzner, Delbaen, Eber, and Heath (1999) to define the so called *coherent risk measures*. The interested reader can also refer to Föllmer and Schied (2016)[Part I, Chapter 4].

<sup>1</sup>The indicator mapping  $\iota_S: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  of a set  $S \subset \mathbb{R}^N$  is the mapping that takes the value 0 on  $S$  and the value  $+\infty$  otherwise.

### 8.3. Convex formulation of robust inference problems using risk measures

**Definition 8.3.1.** A risk measure  $\rho: \mathbb{X} \rightarrow \mathbb{R}$  is said to be

- *monotone*: if, for every  $(\mathbf{X}, \mathbf{Y}) \in \mathbb{X}^2$ ,  $\mathbf{X} \leq \mathbf{Y} \Rightarrow \rho(\mathbf{X}) \leq \rho(\mathbf{Y})$ ,
- *translation invariant*: if, for every  $\mathbf{X} \in \mathbb{X}$  and  $m \in \mathbb{R}$ ,  $\rho(\mathbf{X} + m) = \rho(\mathbf{X}) + m$ ,
- *convex*: if, for every  $(\mathbf{X}, \mathbf{Y}) \in \mathbb{X}^2$  and  $\lambda \in ]0, 1[$ ,  $\rho(\lambda\mathbf{X} + (1-\lambda)\mathbf{Y}) \leq \lambda\rho(\mathbf{X}) + (1-\lambda)\rho(\mathbf{Y})$ ,
- *positively homogeneous*: if, for every  $\mathbf{X} \in \mathbb{X}$  and  $\lambda \in [0, +\infty[$ ,  $\rho(\lambda\mathbf{X}) = \lambda\rho(\mathbf{X})$ .

A risk measure which satisfies the first two properties is called a *monetary risk measure*. A risk measure which satisfies the first three properties is called a *convex risk measure*. A risk measure which satisfies the four properties is called a *coherent risk measure*.

Depending on the author, the first axiom may also be expressed as: for every  $(\mathbf{X}, \mathbf{Y}) \in \mathbb{X}^2$ ,  $\mathbf{X} \leq \mathbf{Y} \Rightarrow \rho(\mathbf{X}) \geq \rho(\mathbf{Y})$  if the variables  $\mathbf{X}$  and  $\mathbf{Y}$  are interpreted as gains instead of losses, which is common in finance. For this reason, some sign differences may appear between results of various authors. We have chosen to follow the paths in [Rockafellar and Uryasev \(2000b\)](#); [Ruszczynski and Shapiro \(2006b,a\)](#) and interpret the random variable in argument as a loss. We however often refer to [Föllmer and Schied \(2016\)](#), providing a comprehensive view of risk measures, where the opposite convention has been adopted.

*Remark 8.3.2.*

1. It readily follows from the translation invariance property that a monetary risk measure  $\rho$  admits a primal form given by

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho(\mathbf{X}) = \inf_{s \in \mathbb{R}} \{s \mid \mathbf{X} - s \in \text{lev}_{\leq 0} \rho\}, \quad (8.8)$$

where  $\text{lev}_{\leq 0} \rho$  is the lower level set of  $\rho$  at height 0 given by

$$\text{lev}_{\leq 0} \rho = \{\mathbf{X} \in \mathbb{X} \mid \rho(\mathbf{X}) \leq 0\}. \quad (8.9)$$

2. A monetary risk measure  $\rho$  is 1-Lipschitz continuous with respect to the supremum norm  $\|\cdot\|_{\infty}$ . Indeed, for every  $(\mathbf{X}, \mathbf{Y}) \in \mathbb{X}^2$ , we have  $\mathbf{X} \leq \mathbf{Y} + \|\mathbf{X} - \mathbf{Y}\|$ . By monotonicity and translation invariance we obtain that  $\rho(\mathbf{X}) - \rho(\mathbf{Y}) \leq \|\mathbf{X} - \mathbf{Y}\|_{\infty}$ , which by symmetry implies that  $|\rho(\mathbf{X}) - \rho(\mathbf{Y})| \leq \|\mathbf{X} - \mathbf{Y}\|_{\infty}$ .

The class of convex risk measures includes a large number of useful functions. Without entering into details, we should mention: expectation, worst case, quantile, median, and average value at risk .



### 8.3.2. Convex reformulation

In this section, we will show that the “min-max” problem 1 admits a convex reformulation. We first gather in the following proposition some existing results in the literature.

**Proposition 8.3.3.**  *$\rho$  is a convex risk measure if and only if there exists a lsc and convex function  $\alpha: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  such that*

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho(\mathbf{X}) = \sup_{\mathbb{Q} \in \mathcal{M}_1} \left( \sum_{i=1}^N \mathbb{Q}_i x_i - \alpha(\mathbb{Q}) \right). \quad (8.10)$$

The function  $\alpha$  associated with  $\rho$  is uniquely defined as

$$(\forall \mathbb{Q} \in \mathbb{R}^N) \quad \alpha(\mathbb{Q}) = \begin{cases} \sup_{\mathbf{X} \in \text{lev}_{\leq 0} \rho} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] & \text{if } \mathbb{Q} \in \mathcal{M}_1, \\ +\infty & \text{otherwise.} \end{cases} \quad (8.11)$$

In addition,  $\rho$  is coherent if and only if its conjugate function  $\alpha$  is the indicator function of a nonempty closed convex subset of  $\mathcal{M}_1$ .

*Proof.*

1. We know from [Föllmer and Schied \(2016, Theorem 4.16 and Proposition 4.15\)](#) that any convex risk measure  $\rho$  on  $\mathbb{X}$  is of the form

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho(\mathbf{X}) = \sup_{\mathbb{Q} \in \mathcal{M}_1} (\mathbb{E}_{\mathbb{Q}}[\mathbf{X}] - \alpha(\mathbb{Q})), \quad (8.12)$$

where  $\alpha: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  is the lsc and convex function whose domain is a nonempty subset of  $\mathcal{M}_1$  given by

$$(\forall \mathbb{Q} \in \mathcal{M}_1) \quad \alpha(\mathbb{Q}) = \sup_{\mathbf{X} \in \mathbb{X}} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] - \rho(\mathbf{X}), \quad (8.13)$$

$$= \sup_{\mathbf{X} \in \text{lev}_{\leq 0} \rho} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] \quad (8.14)$$

(the second equality stems from [Remark 8.3.21](#)).

Conversely, one can associate to every lsc convex function  $\alpha: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  whose domain is a nonempty subset of  $\mathcal{M}_1$  a unique convex risk measure defined by [\(8.12\)](#).

2. It follows from [Föllmer and Schied \(2016, Proposition 4.15\)](#) that if, in addition, the risk measure  $\rho$  is coherent, then the function  $\alpha$  in [\(8.11\)](#) is the indicator function of a nonempty closed convex subset of  $\mathcal{M}_1$  and the converse property holds.

□

We now state the main result of this section.

### 8.3. Convex formulation of robust inference problems using risk measures

**Theorem 8.3.4.** *Let  $\alpha: \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lsc convex function whose domain is a nonempty subset of  $\mathcal{M}_1$ . Problem 1 is equivalent to find*

$$\bar{\theta} \in \arg \min_{\theta \in \mathbb{R}^N} \rho_\alpha(\ell(\theta, \mathbf{z})) , \quad (8.15)$$

where

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho_\alpha(\mathbf{X}) = \max_{\mathbb{Q} \in \mathcal{M}_1} \left( \sum_{i=1}^N \mathbb{Q}_i x_i - \alpha(\mathbb{Q}) \right) . \quad (8.16)$$

The function  $\rho_\alpha(\ell(\cdot, \mathbf{z}))$  is proper, lsc, and convex. In addition, this convex optimization problem admits a primal formulation which consists of finding

$$\inf_{(\theta, s) \in \mathcal{S}} s \quad (8.17)$$

where

$$\mathcal{S} = \{(\theta, s) \in \mathbb{R}^N \times \mathbb{R} \mid \ell(\theta, \mathbf{z}) - s \in \text{lev}_{\leq 0} \rho_\alpha\} \quad (8.18)$$

*Proof.* It follows from Proposition 8.3.3 that (8.5) is equivalent to (8.15) where  $\rho_\alpha$  is a convex risk measure. In addition, the sup in the definition of the risk measure is attained since  $\mathcal{M}_1$  is a compact set and  $q \mapsto \sum_{i=1}^N \mathbb{Q}_i x_i - \alpha(\mathbb{Q})$  is upper semicontinuous.

The function  $\ell(\cdot, Z)$  is lsc convex for every  $Z \in \mathbb{R}^d$ . Given a random variable  $\mathbf{z}$ , for every vectors  $\theta_1$  and  $\theta_2$  in  $\mathbb{R}^N$ , and scalar  $\lambda \in [0, 1]$ , the convexity of function  $\ell$  yields

$$(\forall \omega \in \Omega) \quad \ell(\lambda\theta_1 + (1 - \lambda)\theta_2, \mathbf{z}(\omega)) \leq \lambda\ell(\theta_1, \mathbf{z}(\omega)) + (1 - \lambda)\ell(\theta_2, \mathbf{z}(\omega)) . \quad (8.19)$$

Now, by using the fact that the risk measure  $\rho_\alpha$  is monotone and convex with respect to the ordering introduced in (8.7), we get

$$\rho_\alpha\left(\ell(\lambda\theta_1 + (1 - \lambda)\theta_2, \mathbf{z})\right) \leq \rho_\alpha\left(\lambda\ell(\theta_1, \mathbf{z}) + (1 - \lambda)\ell(\theta_2, \mathbf{z})\right) \quad (8.20)$$

$$\leq \lambda\rho_\alpha\left(\ell(\theta_1, \mathbf{z})\right) + (1 - \lambda)\rho_\alpha\left(\ell(\theta_2, \mathbf{z})\right) . \quad (8.21)$$

This shows that  $\rho_\alpha(\ell(\cdot, \mathbf{z}))$  is convex.

In addition, since  $\rho_\alpha$  is monotone and continuous (see Remark 8.3.21) and  $\ell(\cdot, \mathbf{z})$  is lsc,  $\rho_\alpha(\ell(\cdot, \mathbf{z}))$  is lsc. Because of (8.1),  $\rho_\alpha(\ell(\cdot, \mathbf{z}))$  is also proper.

Finally, formulation (8.17) is deduced from (8.8) and (8.15). □

The general convex reformulation (8.17) is not always easy to handle. In practical applications, the choice of the mapping  $\alpha$  plays a crucial role in this regard. We will see in the next section some useful examples of this function. In particular, some mappings  $\alpha$  lead to a formulation (8.17) that will be shown to be tractable numerically.

## 8.4. Examples of risks measures

By considering particular forms of function  $\alpha$  in Problem 1, we define three classes of interest for robust formulations. The first two ones are based on  $\varphi$ -divergences, while the third one is based on the Wasserstein metric.

### 8.4.1. Perspective functions and divergences

The notion of  $\varphi$ -divergence was first introduced independently by [Csiszár \(1964\)](#); [Morimoto \(1963\)](#) and [Ali and Silvey \(1966\)](#). For a more complete bibliography on the subject, we refer to [Basseville \(2013\)](#).

**Definition 8.4.1.** Let  $\varphi : \mathbb{R} \rightarrow ]-\infty, +\infty]$ . The *perspective function*  $f_\varphi$  of function  $\varphi$  is given by

$$f_\varphi : \mathbb{R} \times \mathbb{R} \rightarrow ]-\infty, +\infty]$$

$$(x, \xi) \mapsto \begin{cases} \xi \varphi\left(\frac{x}{\xi}\right) & \text{if } \xi > 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (8.22)$$

**Definition 8.4.2.** Let  $\varphi : \mathbb{R} \rightarrow [0, +\infty]$  be a lsc convex function with nonempty domain included in  $[0, +\infty[$  such that  $\varphi(1) = 0$ . The  $\varphi$ -divergence  $D_\varphi : \mathbb{R}^N \times \mathbb{R}^N \rightarrow [0, +\infty]$  is defined as

$$(\forall \mathbb{P} = (\mathbb{P}_i)_{1 \leq i \leq N} \in \mathbb{R}^N) (\forall \mathbb{Q} = (\mathbb{Q}_i)_{1 \leq i \leq N} \in \mathbb{R}^N) \quad D_\varphi(\mathbb{P}, \mathbb{Q}) = \sum_{i=1}^N \underline{f}_\varphi(\mathbb{P}_i, \mathbb{Q}_i), \quad (8.23)$$

where the function  $\underline{f}_\varphi$  is the lsc envelope of the mapping  $f_\varphi$ , that is

$$\underline{f}_\varphi : \mathbb{R} \times \mathbb{R} \rightarrow ]-\infty, +\infty]$$

$$(x, \xi) \mapsto \begin{cases} \xi \varphi\left(\frac{x}{\xi}\right) & \text{if } \xi > 0 \text{ and } x \geq 0, \\ x \lim_{t \rightarrow +\infty} \frac{\varphi(t)}{t} & \text{if } \xi = 0 \text{ and } x > 0, \\ 0 & \text{if } \xi = 0 \text{ and } x = 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (8.25)$$

We also recall the definitions of a conjugate function and an adjoint function.

**Definition 8.4.3.** Let  $\varphi : \mathbb{R} \rightarrow ]-\infty, +\infty]$ . The conjugate  $\varphi^*$  of function  $\varphi$  is defined by

$$(\forall s \in \mathbb{R}) \quad \varphi^*(s) = \sup_{t \in \mathbb{R}} (st - \varphi(t)), \quad (8.26)$$

and the so-called adjoint function of  $\varphi$  is defined by

$$(\forall t \in \mathbb{R}) \quad \tilde{\varphi}(t) = \begin{cases} t \varphi\left(\frac{1}{t}\right) & \text{if } t \geq 0, \\ \lim_{t \rightarrow +\infty} \frac{\varphi(t)}{t} & \text{if } t = 0. \end{cases} \quad (8.27)$$

Table 8.1 is an extension of the one in Ben-Tal, Den Hertog, De Waegenaere, Melenberg, and Rennen (2013) and provides the expressions of common  $\varphi$  functions, their conjugates, and the associated  $\varphi$ -divergence. It is well-known (Ben-Tal, Den Hertog, De Waegenaere, Melenberg, and Rennen, 2013; Combettes and Müller, 2018) that the adjoint  $\tilde{\varphi}$  of  $\varphi$  is such that

$$(\forall (\mathbb{P}, \mathbb{Q}) \in (\mathbb{R}^N)^2) \quad D_{\tilde{\varphi}}(\mathbb{P}, \mathbb{Q}) = D_{\varphi}(\mathbb{Q}, \mathbb{P}) \quad (8.28)$$

and the conjugate of function  $\lambda\varphi$  is

$$(\forall s \in \mathbb{R}) \quad (\lambda\varphi)^*(s) = \lambda\varphi^*\left(\frac{s}{\lambda}\right). \quad (8.29)$$

Divergence	$\varphi(t)$	$\varphi(t), t \geq 0$	$D_\varphi(\mathbb{P}, \mathbb{Q})$	$\varphi^*(s)$	$\tilde{\varphi}(t)$
Kullback-Leibler	$\varphi_{kl}(t)$	$t \log(t) - t + 1$	$\sum_{i=1}^N p_i \log\left(\frac{p_i}{q_i}\right)$	$e^s - 1$	$\varphi_b(t)$
Burg entropy	$\varphi_b(t)$	$-\log(t) + t - 1$	$\sum_{i=1}^N q_i \log\left(\frac{q_i}{p_i}\right)$	$-\log(1-s), s < 1$	$\varphi_{kl}(t)$
J-divergence	$\varphi_j(t)$	$(t-1) \log(t)$	$\sum_{i=1}^N (p_i - q_i) \log\left(\frac{p_i}{q_i}\right)$	no closed form	$\varphi_j(t)$
$\chi^2$ -distance	$\varphi_c(t)$	$\frac{1}{t}(t-1)^2$	$\sum_{i=1}^N \frac{p_i - q_i}{p_i}$	$2 - 2\sqrt{1-s}, s < 1$	$\varphi_{mc}(t)$
Modified $\chi^2$ -distance	$\varphi_{mc}(t)$	$(t-1)^2$	$\sum_{i=1}^N \frac{q_i - p_i}{q_i}$	$\begin{cases} -1, & s < -2 \\ s + s^2/4, & s \geq -2 \end{cases}$	$\varphi_c(t)$
Hellinger distance	$\varphi_h(t)$	$(\sqrt{t}-1)^2$	$\sum_{i=1}^N (\sqrt{p_i} - \sqrt{q_i})$	$\frac{s}{1-s}, s < 1$	$\varphi_h(t)$
$\chi$ -divergence of order $\theta, 1$	$\varphi_{ca}^\theta(t)$	$ t-1 ^\theta$	$\sum_{i=1}^N q_i \left 1 - \frac{p_i}{q_i}\right ^\theta$	$s + (\theta-1) \left(\frac{ s }{\theta}\right)^{\frac{\theta}{\theta-1}}$	$t^{1-\theta} \varphi_{ca}^\theta(t)$
Variation distance	$\varphi_v(t)$	$ t-1 $	$\sum_{i=1}^N  p_i - q_i $	$\begin{cases} -1, & s \leq -1 \\ s, & -1 \leq s \leq 1 \end{cases}$	$\varphi_v(t)$
Cressie and Read	$\varphi_{cr}^\theta(t)$	$\frac{1-\theta+\theta t-t^\theta}{\theta(1-\theta)}, \theta \notin \{0, 1\}$ <sup>2</sup>	$\frac{1}{\theta(1-\theta)} \left(1 - \sum_{i=1}^N p_i^\theta q_i^{1-\theta}\right)$	$\begin{cases} \frac{1}{\theta}(1-s(1-\theta))^{\frac{\theta}{\theta-1}} - \frac{1}{\theta} \\ s < \frac{1}{\theta-1} \end{cases}$	$\varphi_{cr}^{1-\theta}(t)$
Average Value at Risk of level $\beta$	$\varphi_{avar}^\beta(t)$	$\iota_{[0, \frac{1}{1-\beta}]}, \beta \in [0, 1]$	$\sum_{i=1}^N \iota_{[0, \frac{1}{1-\beta}]}\left(\frac{p_i}{q_i}\right)$	$\sigma_{[0, \frac{1}{1-\beta}]} = \begin{cases} \frac{1}{1-\beta}, & s \geq 0 \\ 0, & s < 0 \end{cases}$	$\iota_{[1-\beta, +\infty[}$

Table 8.1.: Common perspective functions and their conjugate used to define  $\varphi$ -divergences.

### 8.4.2. Divergence penalty functions

A first case of interest is when the penalty term  $\alpha(\mathbb{Q})$  in Problem 1 measures the “distance” between  $\mathbb{P}$  and  $\mathbb{Q}$  in the sense of a  $\varphi$ -divergence.

**Proposition 8.4.4.** *Let  $\varphi : \mathbb{R} \rightarrow [0, +\infty]$  be a lsc convex function with nonempty domain included in  $[0, +\infty[$  such that  $\varphi(1) = 0$ . Let  $\alpha$  be the function given by*

$$(\forall \mathbb{Q} \in \mathbb{R}^N) \quad \alpha(\mathbb{Q}) = \begin{cases} \lambda_0 D_\varphi(\mathbb{Q}, \mathbb{P}) & \text{if } \mathbb{Q} \in \mathcal{M}_1, \\ +\infty & \text{otherwise,} \end{cases} \quad (8.30)$$

with  $\lambda_0 \in ]0, +\infty[$ . Problem 1 is equivalent to find

$$\bar{\theta} = \arg \min_{\theta \in \mathbb{R}^N} \min_{\mu \in \mathbb{R}} g(\theta, \mu), \quad (8.31)$$

where  $g$  is the proper, lsc, convex function given by

$$(\forall (\theta \in \mathbb{R}^N)(\forall \mu \in \mathbb{R}) \quad g(\theta, \mu) = \mu + \sum_{i=1}^N \mathbb{P}_i \varphi^* \left( \frac{\ell(\theta, z_i)}{\lambda_0} - \mu \right). \quad (8.32)$$

*Proof.* We can reexpress (8.5) as

$$\bar{\theta} \in \arg \min_{\theta \in \mathbb{R}^N} \sup_{\mathbb{Q}=(\mathbb{Q}_i)_{1 \leq i \leq N} \in \mathcal{M}_1} \left( \sum_{i=1}^N q_i \frac{\ell(\theta, z_i)}{\lambda_0} - D_\varphi(\mathbb{P}, \mathbb{Q}) \right). \quad (8.33)$$

It follows from Föllmer and Schied (2016, Theorem 4.122) that

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho_{\frac{\alpha}{\lambda}}(\mathbf{X}) = \min_{\mu \in \mathbb{R}} \mu + \sum_{i=1}^N \mathbb{P}_i \varphi^*(x_i - \mu). \quad (8.34)$$

The equivalence between (8.33) and (8.31) then results from Theorem 8.3.4.

In addition, (8.32) yields

$$(\forall (\theta \in \mathbb{R}^N)(\forall \mu \in \mathbb{R}) \quad g(\theta, \mu) = \mu + \sup_{(t_i)_{i \in [1, n]} \in \mathbb{R}^N} \sum_{i=1}^N \mathbb{P}_i t_i \left( \frac{\ell(\theta, z_i)}{\lambda_0} - \mu \right) - \varphi(t_i). \quad (8.35)$$

For every  $(t_i)_{i \in [1, n]} \in \mathbb{R}^N$ ,

$$(\theta, \mu) \mapsto \sum_{i=1}^N \mathbb{P}_i t_i \left( \frac{\ell(\theta, z_i)}{\lambda_0} - \mu \right) - \varphi(t_i) \quad (8.36)$$

is a lsc convex function. Since convexity and lower semicontinuity are kept by the supremum operation,  $g$  is lsc and convex. By using (8.1), (8.32), and the fact that  $\varphi^*$  is proper, there exist  $\theta \in \mathbb{R}^N$  and  $\mu \in \mathbb{R}$ , such that  $g(\theta, \mu) < +\infty$ . □

### 8.4.3. Constrained formulations

We now investigate two particular cases when  $\alpha$  is the indicator function of a convex set  $\mathcal{Q}$  of probability distributions, so defining an ambiguity set.

#### Ball with respect to a divergence

A possibility is to introduce an upper bound on the divergence  $D_\varphi(\mathbb{Q}, \mathbb{P})$  between the sought distribution  $\mathbb{Q}$  and  $\mathbb{P}$  by considering the constraint set

$$\mathcal{Q} = \mathbb{B}_\epsilon^\varphi = \left\{ q \in \mathcal{M}_1 \mid D_\varphi(\mathbb{Q}, \mathbb{P}) \leq \epsilon \right\}, \quad (8.37)$$

where  $\epsilon \in ]0, +\infty[$ .

The following result generalizes both [Ben-Tal et al. \(2013\)](#) where the authors deal with linear costs under constraints and [Hu and Hong \(2013\)](#) where the authors focus on the Kullback-Leiber divergence.

**Proposition 8.4.5.** *Let  $\varphi : \mathbb{R} \rightarrow [0, +\infty]$  be a lsc convex function such that  $\text{dom}(\varphi) = ]0, +\infty[$  or  $\text{dom}(\varphi) = [0, +\infty[$ , and  $\varphi(1) = 0$ . Let  $\epsilon \in ]0, +\infty[$  and let  $\alpha = \iota_{\mathbb{B}_\epsilon^\varphi}$ . Problem 1 is equivalent to find*

$$\bar{\theta} = \arg \min_{\theta \in \mathbb{R}^N} \min_{(\lambda, \mu) \in \mathbb{R}^2} g(\theta, \lambda, \mu), \quad (8.38)$$

where  $g$  is the proper, lsc, convex function given by

$$(\forall \theta \in \mathbb{R}^N)(\forall (\lambda, \mu) \in \mathbb{R}^2) \quad g(\theta, \lambda, \mu) = \begin{cases} \lambda\epsilon + \mu + \sum_{i=1}^N p_i \lambda \varphi^* \left( \frac{\ell(\theta, z_i) - \mu}{\lambda} \right) & \text{if } \lambda \in [0, +\infty[, \\ +\infty & \text{otherwise,} \end{cases} \quad (8.39)$$

with the convention

$$0\varphi^* \left( \frac{\cdot}{0} \right) = \iota_{]-\infty, 0]}. \quad (8.40)$$

*Proof.* The risk function associated with  $\alpha = \iota_{\mathbb{B}_\epsilon^\varphi}$  is

$$(\forall \mathbf{X} \in \mathbb{X}) \quad \rho_\alpha(\mathbf{X}) = \sup_{q \in \mathbb{R}^N} \sum_{i=1}^N q_i x_i - \iota_{\mathcal{M}_1}(q), \quad (8.41a)$$

$$\text{s.t.} \quad \sum_{i=1}^N p_i \varphi \left( \frac{q_i}{p_i} \right) \leq \epsilon. \quad (8.41b)$$

Since 1 belongs to the interior of  $\text{dom}(\varphi)$  and

$$\sum_{i=1}^N p_i \varphi \left( \frac{p_i}{p_i} \right) = 0 < \epsilon, \quad (8.42)$$

Slater's condition holds for constraint (8.41b). Since the constraint is feasible and  $q \mapsto -\sum_{i=1}^N q_i x_i + \iota_{\mathcal{M}_1}(q)$  is lsc, convex, and coercive, there exists a solution  $\bar{q} \in \mathcal{M}_1$

to the above constrained maximization problem. It then follows from standard Lagrange duality for convex functions that there exists  $\bar{\lambda} \in [0, +\infty[$  such that  $(\bar{q}, \bar{\lambda})$  is a saddle point of the Lagrange function

$$(\forall q \in \mathcal{C})(\forall \lambda \in [0, +\infty[) \quad \Psi_{\mathbf{X}}(q, \lambda) = - \sum_{i=1}^N q_i x_i + \lambda \left( \sum_{i=1}^N p_i \varphi \left( \frac{q_i}{p_i} \right) - \epsilon \right), \quad (8.43)$$

where

$$\mathcal{C} = \left\{ q \in \mathcal{M}_1 \mid (\forall i \in \llbracket 1, n \rrbracket) \frac{q_i}{p_i} \in \text{dom } \varphi \right\}. \quad (8.44)$$

We have thus

$$\rho_{\alpha}(\mathbf{X}) = - \sup_{\lambda \in [0, +\infty[} \inf_{q \in \mathcal{C}} \Psi_{\mathbf{X}}(q, \lambda) = \min_{\lambda \in [0, +\infty[} G(\mathbf{X}, \lambda) = G(\mathbf{X}, \bar{\lambda}), \quad (8.45)$$

where, for every  $\lambda \in [0, +\infty[$ ,

$$G(\mathbf{X}, \lambda) = \lambda \epsilon + \sup_{q \in \mathcal{C}} \sum_{i=1}^N \left( q_i x_i - \lambda p_i \varphi \left( \frac{q_i}{p_i} \right) \right). \quad (8.46)$$

Two cases will be distinguished.

1. Case when  $\lambda = 0$ .

Then (8.46) reduces to

$$G(\mathbf{X}, 0) = \sup_{q \in \mathcal{C}} \sum_{i=1}^N q_i x_i \leq \sigma_{\mathcal{M}_1}(\mathbf{X}), \quad (8.47)$$

where

$$\sigma_{\mathcal{M}_1}(\mathbf{X}) = \sup_{q \in \mathcal{M}_1} \sum_{i=1}^N q_i x_i = \sup_{i \in \llbracket 1, n \rrbracket} x_i. \quad (8.48)$$

In addition, since  $]0, +\infty[ \in \text{dom}(\varphi)$ , the upper bound in (8.47) is attained, yielding

$$G(\mathbf{X}, 0) = \sup_{i \in \llbracket 1, n \rrbracket} x_i. \quad (8.49)$$

2. Case when  $\lambda > 0$ .

(8.46) can be reexpressed as

$$\begin{aligned} G(\mathbf{X}, \lambda) &= \lambda \epsilon + \sup_{q \in \mathcal{M}_1} \sum_{i=1}^N \left( q_i x_i - \lambda p_i \varphi \left( \frac{q_i}{p_i} \right) \right) \\ &= \lambda \epsilon + (\lambda \Phi + \iota_{\mathcal{M}_1})^*(\mathbf{X}), \end{aligned} \quad (8.50)$$

where

$$(\forall q \in \mathbb{R}^N) \quad \Phi(q) = \sum_{i=1}^N p_i \varphi \left( \frac{q_i}{p_i} \right). \quad (8.51)$$



## 8. General risk measures for robust machine learning

The conjugate of  $\Phi$  reads

$$\begin{aligned} (\forall \mathbf{Y} \in \mathbb{X}) \quad (\lambda\Phi)^*(\mathbf{Y}) &= \sup_{q \in \mathbb{R}^N} \sum_{i=1}^N q_i y_i - \lambda p_i \varphi\left(\frac{q_i}{p_i}\right) \\ &= \sum_{i=1}^N p_i (\lambda\varphi)^*(y_i), \end{aligned} \quad (8.52)$$

whereas the conjugate of  $\iota_{\mathcal{M}_1}$  is given by  $\sigma_{\mathcal{M}_1}$  in (8.48). Since  $\sigma_{\mathcal{M}_1}$  is finite valued, the conjugate of  $\Phi + \iota_{\mathcal{M}_1}$  is given by the following inf-convolution (Bauschke and Combettes, 2011, Theorem 15.3)

$$(\Phi + \iota_{\mathcal{M}_1})^*(\mathbf{X}) = \min_{\mathbf{Y} \in \mathbb{X}} \sigma_{\mathcal{M}_1}(\mathbf{Y}) + (\lambda\Phi)^*(\mathbf{X} - \mathbf{Y}), \quad (8.53)$$

which, by using (8.52), yields

$$G(\mathbf{X}, \lambda) = \lambda\epsilon + \min_{\substack{\mathbf{Y} \in \mathbb{X} \\ \sup_{i \in [1, n]} y_i = \mu}} \mu + \sum_{i=1}^N p_i (\lambda\varphi)^*(x_i - y_i). \quad (8.54)$$

Since  $\text{dom}(\lambda\varphi) \subset [0, +\infty[$ ,  $(\lambda\varphi)^*: \xi \mapsto \sup_{v \in [0, +\infty[} \xi v - \lambda\varphi(v)$  is an increasing function. This implies that

$$\begin{aligned} G(\mathbf{X}, \lambda) &= \lambda\epsilon + \min_{\mu \in \mathbb{R}} \mu + \sum_{i=1}^N p_i (\lambda\varphi)^*(x_i - \mu) \\ &= \lambda\epsilon + \min_{\mu \in \mathbb{R}} \mu + \sum_{i=1}^N p_i \lambda \varphi^*\left(\frac{x_i - \mu}{\lambda}\right). \end{aligned} \quad (8.55)$$

Note that the right-hand side in the previous formula when applied at  $\lambda = 0$  by using (8.40) and (8.48) reduces to

$$\begin{aligned} &\min_{\mu \in \mathbb{R}} \mu + \sum_{i=1}^N p_i \iota_{]-\infty, 0]}(x_i - \mu) \\ &= \min_{\substack{\mu \in \mathbb{R} \\ (\forall i \in [1, n]) x_i \leq \mu}} \mu \\ &= G(\mathbf{X}, 0). \end{aligned} \quad (8.56)$$

Consequently, (8.45) leads to

$$\rho_\alpha(\mathbf{X}) = \min_{\lambda \in [0, +\infty[, \mu \in \mathbb{R}} \lambda\epsilon + \mu + \sum_{i=1}^N p_i \lambda \varphi^*\left(\frac{x_i - \mu}{\lambda}\right), \quad (8.57)$$

and (8.38) follows from Theorem 8.3.4.

increasing function. This implies that 7.17(b)]RockyandWets that  $(\lambda_n \Phi + \iota_{\mathcal{M}_1})_{n \in \mathbb{N}}$  epi-converges to  $\iota_{\mathcal{M}_1}$ . By continuity of the Legendre-Fenchel conjugate, this implies that  $((\lambda_n \Phi + \iota_{\mathcal{M}_1})^*)_{n \in \mathbb{N}}$  epi-converges to  $\sigma_{\mathcal{M}_1}$ . Since

In addition, by using the expression of the conjugate,  $g$  can be reexpressed as

$$(\forall \theta \in \mathbb{R}^N)(\forall (\lambda, \mu) \in \mathbb{R}^2)$$

$$g(\theta, \lambda, \mu) = \sup_{(t_i)_{i \in \llbracket 1, n \rrbracket} \in \mathbb{R}^N} \lambda \epsilon + \mu + \sum_{i=1}^N p_i (\ell(\theta, z_i) - \mu) t_i - \lambda \varphi(t_i) + \iota_{[0, +\infty[}(\lambda). \quad (8.58)$$

As a supremum of lsc convex functions,  $g$  also is lsc convex. The fact that  $g$  is proper follows from arguments similar to those at the end of the proof of Proposition 8.4.4.  $\square$

*Remark 8.4.6.* The divergence risk measure in (8.34) is convex, whereas the risk measure in (8.57) is coherent, which means that the risk scales with the data in the latter case.

### Ball with respect to the Wasserstein metric

We now investigate Problem 1 when function  $\alpha$  is the indicator of a Wasserstein ball centered on  $\mathbb{P}$ . For this purpose, we first recall the notion of Wasserstein distance.

**Definition 8.4.7.** Let  $\mathcal{M}(\Xi^2)$  denote the set of probability distributions supported on  $\Xi^2$ . The Wasserstein distance between two distributions  $\mathbb{P}$  and  $\mathbb{Q}$  supported on  $\Xi$  is defined as

$$W(p, q) = \inf_{\Pi \in \mathcal{M}(\Xi^2)} \left\{ \int_{\Xi^2} d(\xi, \xi') \Pi(d\xi, d\xi') \mid \Pi(d\xi, \Xi) = q(d\xi), \Pi(\Xi, d\xi') = p(d\xi') \right\}, \quad (8.59)$$

where  $d$  is a metric on  $\Xi$ .

We now introduce the notion of Wasserstein ball. The considered constrained set is denoted by

$$\Omega = \mathbb{B}_\epsilon^{\mathbb{W}} = \left\{ q \in \mathcal{M}_1 \mid W(\mathbb{P}, \mathbb{Q}) \leq \epsilon \right\} \quad (8.60)$$

with  $\epsilon \in ]0, +\infty[$ .

The following convex reformulation of Problem 1 can be derived from (Esfahani and Kuhn, 2015, Theorem 4.2).

**Proposition 8.4.8.** *Let  $\epsilon \in ]0, +\infty[$  and let  $\alpha = \iota_{\mathbb{B}_\epsilon^{\mathbb{W}}}$ . Then, Problem 1 is equivalent to find*

$$\bar{\theta} = \arg \min_{\theta \in \mathbb{R}^N} \min_{\lambda \in \mathbb{R}, s \in \mathbb{R}^N} g(\theta, \lambda, s), \quad (8.61)$$

where  $g$  is the proper, lsc convex function given by

$$(\forall \theta \in \mathbb{R}^N)(\forall \lambda \in \mathbb{R})(\forall s = (s_j)_{1 \leq j \leq N} \in \mathbb{R}^N) \quad g(\theta, \lambda, s) = \lambda \epsilon + \sum_{j=1}^N p_j s_j + \iota_{\mathcal{W}}(\theta, \lambda, s), \quad (8.62)$$

where  $\mathcal{W}$  is the closed convex set defined as

$$\mathcal{W} = \{(\theta, \lambda, s) \in \mathbb{R}^n \times [0, +\infty[ \times \mathbb{R}^N \mid (\forall (i, j) \in \llbracket 1, N \rrbracket^2) \ell(\theta, z_i) - \lambda \|z_i - z_j\| \leq s_j\}. \quad (8.63)$$

## 8.5. Numerical solution

We will now propose an algorithm allowing us to solve numerically the three convex optimization problems in Propositions 8.4.4, 8.4.5, and 8.4.8. This algorithm applies to more general choices of function  $\alpha$  in Problem 1 where the constraint  $\mathcal{S}$  in (8.18) splits as an intersection of a finite number of convex constraints.

### 8.5.1. A unifying formulation

We first show that the convex optimization problems discussed in Section 8.4 can be reexpressed in a unifying manner.

**Proposition 8.5.1.** *The optimization problems in Propositions 8.4.4, 8.4.5, and 8.4.8 amount to finding*

$$(\bar{\theta}, \bar{\lambda}, \bar{\mu}, \bar{s}) \in \arg \min_{\theta \in \mathbb{R}^N, \lambda \in [0, +\infty[, \mu \in \mathbb{R}, s \in \mathbb{R}^N} \lambda \epsilon + \mu + \sum_{i=1}^N p_i s_i \quad (8.64a)$$

$$s.t. \quad (\forall k \in \llbracket 1, K \rrbracket) \quad f_k(\theta, \lambda, \mu, \mathbf{z}) \leq 0, \quad (8.64b)$$

where  $K \in \mathbb{N} \setminus \{0\}$  and the functions  $(f_k(\cdot, \mathbf{z}))_{k \in \llbracket 1, K \rrbracket}$  are proper, lsc, and convex. More precisely,

1. for divergence penalty functions,  $K = N$  and, for every  $k \in \llbracket 1, K \rrbracket$ ,

$$(\forall \theta \in \mathbb{R}^N)(\forall \lambda \in [0, +\infty[)(\forall \mu \in \mathbb{R})(\forall s \in \mathbb{R}^N) \\ f_k(\theta, \lambda, \mu, s, \mathbf{z}) = \varphi^* \left( \frac{\ell(\theta, z_k)}{\lambda} - \mu \right) + \iota_{\{\lambda_0\}}(\lambda) - s_k, \quad (8.65)$$

2. for divergence ball constraints,  $K = N$  and, for every  $k \in \llbracket 1, K \rrbracket$ ,

$$(\forall \theta \in \mathbb{R}^N)(\forall \lambda \in [0, +\infty[)(\forall \mu \in \mathbb{R})(\forall s \in \mathbb{R}^N) \\ f_k(\theta, \lambda, \mu, s, \mathbf{z}) = \lambda \varphi^* \left( \frac{\ell(\theta, z_k) - \mu}{\lambda} \right) - s_k, \quad (8.66)$$

3. for the Wassertein ball constraint,  $K = N^2$  and, for every  $k \in \llbracket 1, K \rrbracket$  and  $(i_k, j_k) \in \llbracket 1, N \rrbracket^2$  such that  $k = N(i_k - 1) + j_k$ ,

$$(\forall \theta \in \mathbb{R}^N)(\forall \lambda \in [0, +\infty[)(\forall \mu \in \mathbb{R})(\forall s \in \mathbb{R}^N) \\ f_k(\theta, \lambda, \mu, s, \mathbf{z}) = \ell(\theta, z_{i_k}) - \lambda \|z_{i_k} - z_{j_k}\| - s_{j_k}. \quad (8.67)$$

### 8.5.2. Description of the algorithm

In this section, we propose an accelerated projected gradient algorithm for solving Problem (8.64). One step of this proximal algorithm [Combettes and Pesquet \(2010\)](#) reads as a projection onto a set defined as an intersection of non trivial closed convex sets. To solve this projection problem, we use the subgradient projection algorithm in [Combettes \(2003\)](#), which is related to ideas introduced in [Haugazeau \(1968, Theorem 3-2\)](#). This algorithm allows the constraints to be activated individually in a flexible parallel manner. We will first recall the basic structure of our algorithm before describing in more details the subgradient projection step.

**Proximal algorithm** Let  $\mathcal{H} = \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N$  and let  $\|\cdot\|$  (resp.  $\langle \cdot, \cdot \rangle$ ) denote the standard norm (resp. the inner product) equipping this product space. By introducing the generic variable  $u = (\theta, \lambda, \mu, s) \in \mathcal{H}$ , (8.64) can be reexpressed more concisely as

$$\min_{u \in \mathcal{H}} \langle c, u \rangle + \iota_{\mathcal{C}}(u), \quad (8.68)$$

where  $c = (0, \epsilon, 1, p) \in \mathcal{H}$  and  $\mathcal{C} = \bigcap_{k=0}^K \mathcal{C}_k$  with

$$\mathcal{C}_0 = \{(\theta, \lambda, \mu, s) \in \mathbb{R}^n \times [0, +\infty[ \times \mathbb{R} \times \mathbb{R}^N\}, \quad (8.69)$$

$$(\forall k \in \llbracket 1, K \rrbracket) \quad \mathcal{C}_k = \text{lev}_{\leq 0} f_k(\cdot, \mathbf{z}). \quad (8.70)$$

To solve the above problem, we propose to employ a FISTA-like algorithm [Beck and Teboulle \(2009\)](#). Let  $n \in \mathbb{N} \setminus \{0\}$ . The  $n$ -th iteration of this algorithm reads

$$v^{(n)} = u^{(n)} + \frac{\tau^{(n)} - 1}{\tau^{(n+1)}} (u^{(n)} - u^{(n-1)}), \quad (8.71)$$

$$u^{(n+1)} = P_{\mathcal{C}}(v^{(n)} - \gamma c), \quad (8.72)$$

where  $\gamma \in ]0, +\infty[$  and  $P_{\mathcal{C}}: \mathcal{H} \rightarrow \mathcal{C}$  is the projection onto the closed convex set  $\mathcal{C}$ . It follows from ([Chambolle and Dossal, 2015](#), Theorem 3) that, if a solution to the minimization problem exists, and

$$\tau^{(n)} = \frac{n + a - 1}{a}, \quad a > 2, \quad (8.73)$$

then the convergence to a solution to the problem is guaranteed.

The main difficulty in the implementation of the algorithm lies in the computation of the projection onto  $\mathcal{C}$  that will be discussed next.

**Computation of the projection** Algorithm 3 presents our projection method inspired from [Combettes \(2003\)](#). At iteration  $\ell \in \mathbb{N}$ ,  $Q(p^{(0)}, p^{(\ell)}, r^{(\ell)})$  designates the projection of  $p^{(0)}$  onto the intersection of the 3 half-spaces  $\mathcal{C}_0$ ,  $H_{\ell}$ , and  $D_{\ell}$ , where

$$H_{\ell} = \{u \in \mathcal{H} \mid \langle u - r^{(\ell)}, p^{(\ell)} - r^{(\ell)} \rangle \leq 0\}, \quad (8.74)$$

$$D_{\ell} = \{u \in \mathcal{H} \mid \langle u - p^{(\ell)}, p^{(0)} - p^{(\ell)} \rangle \leq 0\}. \quad (8.75)$$

Since the projection onto  $H_{\ell} \cap D_{\ell}$  has an explicit form [Combettes \(2003\)](#), a dual forward-backward algorithm [Combettes et al. \(2010\)](#) allows us to compute in a fast manner the projection onto  $\mathcal{C}_0 \cap H_{\ell} \cap D_{\ell}$ . The algorithm has been initialized by setting  $p^{(0)} = P_{\mathcal{C}_0}(v^{(n)} - \gamma c)$ , taking into account the fact that  $P_{\mathcal{C}} = P_{\mathcal{C}} \circ P_{\mathcal{C}_0}$ . At each iteration  $\ell$ ,  $\mathbb{K}_{\ell}$  designates the set of indices of the constraints which are activated. When dealing with large-scale problems, it may be useful not to required all the constraints to be activated at each iteration. The convergence of the algorithm is guaranteed by the study in [Combettes \(2000\)](#), provided that, for every  $k \in \llbracket 1, K \rrbracket$ ,  $\mathcal{C}_0 \subset \text{dom}(\partial f_k)$  and there exists an integer  $M_k$  such that

$$(\forall \ell \in \mathbb{N}) \quad k \in \bigcup_{s=\ell}^{\ell+M_k} \mathbb{K}_s. \quad (8.76)$$

## 8. General risk measures for robust machine learning

The first assumptions on the domains of the subdifferentials of the functions  $(f_k)_{k \in \llbracket 1, K \rrbracket}$  is not however satisfied in (8.65). In this case, the direct simpler form of the algorithm in Combettes (2003) can be applied since the parameter  $\lambda$  is fixed.

<p><b>Data:</b> <math>u^{(n-1)} = u^{(n)} \in \mathcal{H}</math>, <math>\delta \in ]0, 1[</math></p> <p><b>Result:</b> Output of the accelerated projected gradient iteration (8.72)</p> <p>1 <math>v^{(n)} = u^{(n)} + \frac{\tau^{(n)} - 1}{\tau^{(n+1)}}(u^{(n)} - u^{(n-1)});</math></p> <p>2 <math>p^{(0)} = P_{\mathcal{C}_0}(v^{(n)} - \gamma c);</math></p> <p>3 Initialize <math>\ell = 0</math>, <b>while</b> <math>p^{(\ell)} \notin \mathcal{C}</math> <b>do</b></p> <p>4     Take a nonempty finite index set <math>\mathbb{K}_\ell \subset \llbracket 1, K \rrbracket;</math></p> <p>5     For every <math>k \in \mathbb{K}_\ell</math>,</p> <p style="padding-left: 20px;"> <math display="block">p_k^{(\ell)} = \begin{cases} p^{(\ell)} - \frac{f_k(p^{(\ell)})t_k^{(\ell)}}{\ t_k^{(\ell)}\ }, &amp; t_k^{(\ell)} \in \partial f_k(p^{(\ell)}), \text{ if } f_k(p^{(\ell)}) &gt; 0 \\ p^{(\ell)}, &amp; \text{if } f_k(p^{(\ell)}) \leq 0 \end{cases}</math> </p> <p>6     Choose <math>\{\omega_{k,\ell} \mid k \in \mathbb{K}_\ell\} \subset [\delta, 1]</math> such that <math>\sum_{k \in \mathbb{K}_\ell} \omega_{k,\ell} = 1</math></p> <p>7     <math>q^{(\ell)} = \sum_{k \in \mathbb{K}_\ell} \omega_{k,\ell} p_k^{(\ell)} - p^{(\ell)}</math></p> <p>8     <math>L_\ell = \begin{cases} \frac{\sum_{k \in \mathbb{K}_\ell} \omega_{k,\ell} \ p_k^{(\ell)} - p^{(\ell)}\ ^2}{\ q^{(\ell)}\ ^2}, &amp; \text{if } p^{(\ell)} \notin \bigcap_{k \in \mathbb{K}_\ell} \mathcal{C}_k \\ 1, &amp; \text{otherwise} \end{cases}</math></p> <p>9     <math>r^{(\ell)} = p^{(\ell)} - L_\ell q^{(\ell)};</math></p> <p>10    <math>p^{(\ell+1)} = Q(p^{(0)}, p^{(\ell)}, r^{(\ell)});</math></p> <p>11 <b>end</b></p> <p>12 <b>return</b> <math>u^{(n+1)} = p^{\text{end}}</math></p>
--

**Algorithm 3:** Projection algorithm.

## 8.6. Simulation example

Distributionally robust optimization aims at improving out-of-sample performance. In practice, “a little of robustness” typically improves a little bit the expected reward (around 1%) however, results in a larger reduction in terms of variance. We refer to Shafieezadeh-Abadeh, Esfahani, and Kuhn (2015) and to Gotoh, Kim, and Lim (2018) for a complete discussion about expected reward and variance reduction.

Our objective here is to demonstrate how our algorithm can tackle different ambiguity sets. We compare here the formulation in Problem 8.4.5 that uses ambiguity sets defined through the Kullback-Leiber divergence<sup>3</sup> with the formulation in Problem 8.4.8 that uses ambiguity sets defined through the Wasserstein distance. In all our simulations, the logistic regression loss in (8.4) is used for binary classification Briceno-Arias et al. (2017).

We first apply our algorithm to the “ionosphere” dataset<sup>4</sup> where each sample

<sup>3</sup>We perform an asymptotic expansion when  $\lambda$  is close to 0 in order to compute  $p_k^\ell$  in Algorithm 3.

<sup>4</sup>This dataset is accessible at <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html>.

provides  $n = 34$  features. We divide our dataset between a training set and a testing set where the training set contains 60% of the original dataset.

In a nutshell, the robust framework based on the Wasserstein distance provides a better expected reward but at the expense of a higher computational cost. The risk measure based on the Kullback-Leiber divergence is easily tractable, provides a reduction of the variance in out-of-samples results, but no significant increase in terms of expected reward (see [Gotoh, Kim, and Lim \(2018\)](#) for a more detailed theoretical analysis).

We first present convergence results in Figures 8.1 and 8.2 for both formulations. We see that, for both formulations, the proposed accelerated projected gradient algorithm converges and that it converges slightly slower for larger values of  $\epsilon$ . The convergence is faster in the case of the Kullback-Leiber divergence since the number of constraints grows linearly as a function of the number of observations, whereas the growth is quadratic in the case of the Wasserstein distance.

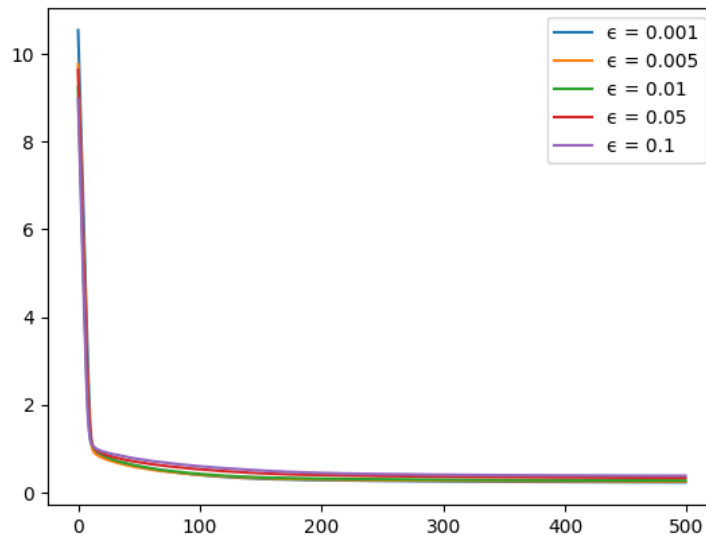


Figure 8.1.: Value of the objective function with respect to iteration number  $n$  for different values of  $\epsilon$  with Kullback-Leibler divergence.

We provide in Table 8.2 below, the value of the area under the ROC curve. We see that when  $\epsilon$  takes larger values, we first have an improved performance. Then the gain starts to decrease, and finally we observe a decay in performance. Indeed, when  $\epsilon$  tends to  $+\infty$ , we recover the worst case risk measure.

We now present in Table 8.3 the evolution of the Area Under the ROC curve for another dataset “colon-cancer” endowed with  $n = 2000$  features and only 64 observations <sup>5</sup>. We want to emphasize the power of robustness in the case when only a few observations are available. This case is particularly interesting because,

<sup>5</sup>This dataset is also accessible at <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html>

## 8. General risk measures for robust machine learning

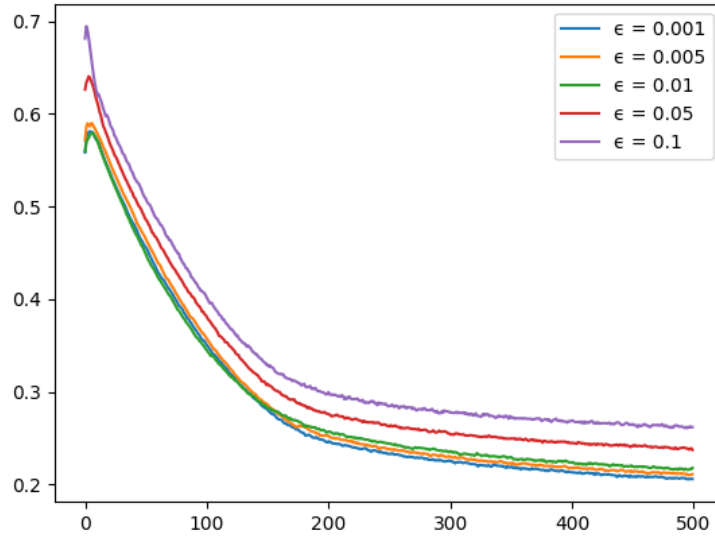


Figure 8.2.: Value of the objective function with respect to iteration number  $n$  for different values of  $\epsilon$  with Wasserstein distance.

Value of $\epsilon$	AUC with KL	AUC with Wasserstein
$\epsilon = 0$ (LR)	0.70	0.70
$\epsilon = 0.001$	<b>0.72</b>	0.77
$\epsilon = 0.005$	0.71	<b>0.77</b>
$\epsilon = 0.01$	0.71	0.77
$\epsilon = 0.05$	0.70	0.76
$\epsilon = 0.1$	0.69	0.76

Table 8.2.: Values of the area under ROC curve for different value of  $\epsilon$  for ionosphere dataset.

when the number of observations is small, the formulation in Problem 8.4.8 is not computationally expensive and provides an increase of 7%, which is significant.

Value of $\epsilon$	AUC with KL	AUC with Wasserstein
$\epsilon = 0$ (LR)	0.79	0.79
$\epsilon = 0.001$	0.73	0.84
$\epsilon = 0.005$	<b>0.84</b>	0.85
$\epsilon = 0.01$	0.77	0.81
$\epsilon = 0.05$	0.72	<b>0.86</b>
$\epsilon = 0.1$	0.81	0.79

Table 8.3.: Values of the area under ROC curve for different value of  $\epsilon$  for colon-cancer dataset.

## 8.7. Conclusion

We have highlighted that risk measures offer versatile tools for addressing robustly machine learning problems. By assuming that the loss function is convex, the related optimization problem has been recast as a convex one. We have shown that various classes of risk measures, e.g. those based on divergences or the Wasserstein distance, lead to a common convex formulation. In addition, an efficient convex optimization algorithm has been proposed to cope with the non trivial constrained problem resulting from this formulation.





# Conclusion

This manuscript is the result of three years of work at École des Ponts ParisTech, within the Optimization and Systems group whose seniors members are Pierre Carpentier, Jean-Philippe Chancelier, Michel De Lara and Vincent Leclère. I have also conducted work in the University of Auckland that is presented in Chap. 6 under the supervision of Andy Philpott. Chapter 8 is the result of a joint work with Jean-Christophe Pesquet and Émilie Chouzenoux.

## Contributions of this manuscript

This manuscript is a contribution to the domains of stochastic optimization, game theory and machine learning. We have presented links between risk averse optimization and decomposition methods.

- In Chap. 3, we have presented an abstract setting for time consistency after having revisited the literature. Our main contribution is to state an equivalence statement between time consistency and nested formula.
- In Chap. 4, we have studied three classes of times consistent mapping: translation invariant mappings, Fenchel-Moreau transforms and mappings defined as supremum. For each class we have given characterizations of time consistency or conditions to obtain time consistency.
- In Chap. 5, we have extended the framework of time consistency to the one of consistency for binary relations. Hence, we have been able to derive a notion of player consistency. Then we have detailed how the notion of player consistency can be used for sequential and parallel decomposition in optimization.
- In Chap. 6, we have studied risk averse equilibrium in risk averse markets. We have exhibited an example that displays multiple equilibria even with nice assumptions of strict concavity. The content of Chap. 6 has been published in *Operation Research Letters* (up to minor modifications).
- In Chap. 7, we have given an interpretation of market prices as tools to build agent consistent mappings.
- In Chap. 8, we have provided links between risk averse optimization and distributionally robust optimization in machine learning. We have proposed an algorithm that solves several different classical problems and we have shown that sub-differential projection is a tool that let us gain parallelism in the algorithm.

## **Perspective of future work**

Without being exhaustive, we recap here work that could be done and questions that remain open.

- One could look for application of the characterization of time consistency in terms of acceptance sets, either to demonstrate that some mappings are not time consistent, or to build mappings that are time consistent. In particular, we have provided a characterization of player consistency in terms of acceptance set.
- The conditions to be satisfied by generalized convex mappings in order to obtain a nested formula are difficult to check. Finding mappings that are not convex risk measures which satisfy these conditions is an open question. We should investigate what happens when we mix translation invariant mappings, Fenchel-Moreau transform, supremum to construct time consistent mappings.
- Further work needs to be done to account for dependencies between players, that is, when there are joint constraints on the players decision sets.
- In machine learning, influence diagrams capture dynamic interactions between players. They may provide an application for the decomposition method.
- In extension of Chap. 6, we could investigate tâtonnement algorithms on large scale equilibrium problem using classical approximation algorithms.
- In extension of Chap. 7, we could explain how completing the market is related to building player consistency.
- In extension of Chap. 8, we could benchmark our algorithm with other existing methods to understand the pros and cons of what we have proposed.

**Part V.**  
**Appendix**



# 9. Recalls

## Contents

---

<b>9.1. Recall on order and relations</b> . . . . .	<b>157</b>
9.1.1. Generalities . . . . .	157
9.1.2. Useful mathematical tools to deal with forsets and aftersets	158
<b>9.2. Recall on lattices and partitions</b> . . . . .	<b>160</b>
9.2.1. Lattices . . . . .	160
9.2.2. Partition . . . . .	161
<b>9.3. Recall on stochastic kernels</b> . . . . .	<b>161</b>
<b>9.4. Recall on star-difference</b> . . . . .	<b>162</b>
<b>9.5. Recall on risk measures</b> . . . . .	<b>163</b>
9.5.1. Definition of monetary risk measure . . . . .	163
9.5.2. Examples of risk measure . . . . .	164
9.5.3. Dual representation of coherent risk measure . . . . .	165
<b>9.6. Recall on Moreau's transform and Moreau's addition</b> .	<b>166</b>

---

## 9.1. Recall on order and relations

### 9.1.1. Generalities

A *binary relation* on the set  $\mathbb{A}$  is a subset  $\mathcal{A}$  of  $\mathbb{A}^2 = \mathbb{A} \times \mathbb{A}$ . As is traditional, we denote from now on

$$a \mathcal{A} a' \Leftrightarrow (a, a') \in \mathcal{A}, \quad \forall (a, a') \in \mathbb{A}^2. \quad (9.1)$$

The *equality* or *diagonal relation* is

$$\Delta_{\mathbb{A}} = \{(a, a) \mid a \in \mathbb{A}\}. \quad (9.2)$$

**Definition 9.1.1.** Let  $\mathcal{A}$  be a relation on the set  $\mathbb{A}$ . The *forset*  $\mathcal{A}a'$  of the element  $a' \in \mathbb{A}$  for the relation  $\mathcal{A}$  is defined by

$$\mathcal{A}a' = \{a \in \mathbb{A} \mid a \mathcal{A} a'\}. \quad (9.3)$$

The *afterset*  $a\mathcal{A}$  of the element  $a \in \mathbb{A}$  for the relation  $\mathcal{A}$  has a straightforward definition.

## 9. Recalls

Let  $\mathbb{A}$  be a set and  $\mathcal{A} \subset \mathbb{A} \times \mathbb{A}$  be a binary relation on  $\mathbb{A}$ .

We say that the relation  $\mathcal{A}$  is *reflexive* if for all element  $a \in \mathbb{A}$ ,  $(a, a) \in \mathcal{A}$ . We will say that it is *transitive* if for all  $(a_1, a_2, a_3) \in \mathbb{A}^3$ ,  $(a_1, a_2) \in \mathcal{A}$  and  $(a_2, a_3) \in \mathcal{A} \Rightarrow (a_1, a_3) \in \mathcal{A}$ . The relation  $\mathcal{A}$  is said to be *antisymmetric* if  $\forall (a_1, a_2) \in \mathbb{A}^2$ ,  $(a_1, a_2) \in \mathcal{A}$  and  $(a_2, a_1) \in \mathcal{A} \Rightarrow a_1 = a_2$ .

A *preorder* is a reflexive and transitive relation. A preorder is said to be total if  $\forall (a_1, a_2) \in \mathbb{A}^2$  we have either  $(a_1, a_2) \in \mathcal{A}$  or  $(a_2, a_1) \in \mathcal{A}$ . A *chain* is a totally ordered subset. If every chain has a supremum (See Appendix 9.2, the preorder is *complete*. An antisymmetric preorder is an *order*.

**Recall on special type of relation** From now on, we consider that  $\mathbb{A}$  is a vectorial space. A vectorial relation is a relation applied coordinates by coordinates. We also denote the translation of vector  $t$  by  $\tau_t$  meaning that  $\tau_t(a) = a + t$ .

In this context a relation  $\mathcal{A}$  on the set  $\mathbb{A}$  is *congruent* if  $a \mathcal{A} a' \Rightarrow (a + x) \mathcal{A} (a' + x)$ ,  $\forall x \in \mathbb{A}$ . We will say that it is translation invariant if  $\tau_t^* \mathcal{A} \subset \mathcal{A}$ ,  $\forall t > 0$  and that it is *monotonous* whenever  $a \geq a' \Rightarrow a \mathcal{A} a'$ . The relation  $\mathcal{A}$  is said to be *convex* if  $(\theta a + (1 - \theta)a') \mathcal{A} x$  whenever  $a \mathcal{A} x$  and  $a' \mathcal{A} x$  and  $\theta \in ]0; 1[$ , or equivalently  $\mathcal{A} x$  is convex. It said to be *positively homogeneous* if  $a \mathcal{A} a' \Rightarrow (\theta a) \mathcal{A} (\theta a')$  when  $\theta > 0$ .

It is well known that there is a one to one correspondence between vectorial orders  $\leq_K$  and convex cones  $K$  by  $a \leq_K a' \Leftrightarrow a - a' \in K$  (See Khan et al. (2015)).

We can write similar definition using the notion of forset and afterset.

### 9.1.2. Useful mathematical tools to deal with forsets and aftersets

**Definition 9.1.2.** We adopt a naming convention for forsets of a relation. The forsets of a equivalence relation are the *equivalence classes*, the ones of orders are the *predecessors*, the ones of preorders are the *lower sets*.

When we have particular structures on  $\mathbb{A}$  and  $\mathbb{B}$ , we can precise the nature of forsets and aftersets. The following proposition explicits the nature of product and intersection of forsets and aftersets.

**Proposition 9.1.3.** Let  $\mathcal{A}$  be a relation on the set  $\mathbb{A}$  and  $\mathcal{B}$  be a relation on the set  $\mathbb{B}$ .

1. If  $\mathcal{A}$  and  $\mathcal{B}$  are equivalence relations, then  $\mathcal{A} \times \mathcal{B}$  is an equivalence relation. In addition, the equivalent classes of the product relation are the products of the equivalent classes of relations, that is,

$$(\mathcal{A} \times \mathcal{B})(a', b') = \mathcal{A} a' \times \mathcal{B} b', \quad \forall a \in \mathbb{A}, \quad \forall b \in \mathbb{B}. \quad (9.4)$$

2. If  $\mathcal{A}$  and  $\mathcal{B}$  are preorders, then  $\mathcal{A} \times \mathcal{B}$  is a preorder. In addition, the lower sets of the product preorder are the products of lower sets of preorders, that is,

$$(\mathcal{A} \times \mathcal{B})(a', b') = \mathcal{A} a' \times \mathcal{B} b', \quad \forall a \in \mathbb{A}, \quad \forall b \in \mathbb{B}. \quad (9.5)$$

The same holds true for the product of upper sets of preorders.

3. If  $\mathcal{A}$  and  $\mathcal{B}$  are preorders, then  $\mathcal{A} \cap \mathcal{B}$  is a preorder. In addition, the lower sets of the intersection preorder are the intersections of lower sets of preorders, that is,

$$(a, b) (\mathcal{A} \cap \mathcal{B}) = a \mathcal{A} \cap b \mathcal{B}, \quad \forall a \in \mathbb{A}, \quad \forall b \in \mathbb{B}. \quad (9.6)$$

The same holds true for intersection of the aftersets of relations.

Let  $\mathcal{A}$  and  $\mathcal{A}'$  be two relations on the set  $\mathbb{A}$ . Recalling that  $\mathcal{A}$  and  $\mathcal{A}'$  can be seen as subsets of  $\mathbb{A}^2$ , the intersection of relations has the straightforward definition

$$a \mathcal{A} \cap \mathcal{A}' a' \Leftrightarrow a \mathcal{A} a' \text{ and } a \mathcal{A}' a'. \quad (9.7)$$

**Proposition 9.1.4.** *The following properties are inherited by the intersection of two relations: reflexivity, transitivity, symmetry, congruence, positive homogeneity, antisymmetry, monotonicity, convexity, invariance by translation and completeness.*

*Totality is not preserved by the intersection of two relations.*

The following recalls rely upon the references [Lewis and Papadimitriou \(1997\)](#), [Butzer et al. \(1967\)](#) and [Schmidt and Ströhlein \(2012\)](#). We first define three notions of product of relations.

**Definition 9.1.5** ([Imrich \(1998\)](#)). Let  $\mathcal{A}$  be a relation on the set  $\mathbb{A}$  and  $\mathcal{B}$  be a relation on the set  $\mathbb{B}$ . The *cardinal product*  $\mathcal{A} \times \mathcal{B}$  on the set  $\mathbb{A} \times \mathbb{B}$  is defined by the subset  $\mathcal{A} \times \mathcal{B}$  of the set  $\mathbb{A}^2 \times \mathbb{B}^2 \equiv (\mathbb{A} \times \mathbb{B})^2$  that is,

$$(a, b) (\mathcal{A} \times \mathcal{B}) (a', b') \Leftrightarrow (a, a') \in \mathcal{A} \text{ and } (b, b') \in \mathcal{B} \\ \forall (a, b, a', b') \in (\mathbb{A} \times \mathbb{B})^2. \quad (9.8)$$

The *strong product*  $\mathcal{A} \boxtimes \mathcal{B}$  is defined by

$$\mathcal{A} \boxtimes \mathcal{B} = ((\Delta_{\mathbb{A}} \cup \mathcal{A}) \times (\Delta_{\mathbb{B}} \cup \mathcal{B})) \setminus \Delta_{\mathbb{A} \times \mathbb{B}}. \quad (9.9)$$

The *Cartesian product*<sup>1</sup>  $\mathcal{A} \square \mathcal{B}$  is defined by

$$\mathcal{A} \square \mathcal{B} = (\Delta_{\mathbb{A}} \times \mathcal{B}) \cup (\mathcal{A} \times \Delta_{\mathbb{B}}). \quad (9.10)$$

By developping [\(9.9\)](#) we easily see that

$$\mathcal{A} \boxtimes \mathcal{B} = (\mathcal{A} \square \mathcal{B}) \cup (\mathcal{A} \times \mathcal{B}) \quad (9.11)$$

We focus on the cardinal product of relations.

**Proposition 9.1.6.** *The following properties are inherited by the cardinal product: reflexivity, transitivity, symmetry, congruence, positive homogeneity, antisymmetry, monotonicity, convexity and invariance by translation*

*Totality and completeness are not preserved by cardinal product.*

---

<sup>1</sup>The definition of Cartesian product for relations differs from the usual one for sets which corresponds here to the cardinal product.



## 9. Recalls

Let  $\mathbb{A}$  and  $\mathbb{B}$  be two sets. Given a relation  $\mathcal{C}$  on the set  $\mathbb{A} \times \mathbb{B}$  the *parametric relation*  $\mathcal{C}_a$  on the set  $\mathbb{B}$  is defined by

$$b \mathcal{C}_a b' \Leftrightarrow (a, b) \mathcal{C} (a, b'), \quad (9.12)$$

and we then have

$$b \bigcap_{a \in \mathbb{A}} \mathcal{C}_a b' \Leftrightarrow \forall a \in \mathbb{A}, b \mathcal{C}_a b'. \quad (9.13)$$

The next proposition deals with the specific case of orders.

**Proposition 9.1.7.** *If the relation  $\mathcal{C}$  is an order on the set  $\mathbb{A} \times \mathbb{B}$ , then  $\mathcal{C}_a$  defined in (9.12) is an order on the set  $\mathbb{B}$  and the intersection  $\bigcap_{a \in \mathbb{A}} \mathcal{C}_a$  is also an order on the set  $\mathbb{B}$ .*

Here is a general result dealing with intersections of parametric relations

**Proposition 9.1.8.** *If the relation  $\mathcal{C}$  is an order on the set  $\mathbb{A} \times \mathbb{B}$ , then*

1.  $(\forall (a, a') \in \mathbb{A}^2, \mathcal{C}_a = \mathcal{C}_{a'}) \Rightarrow (\Delta_{\mathbb{A}} \times \mathcal{C}_a \subset \mathcal{C}, \forall a \in \mathbb{A})$ ;
2. *if  $\mathcal{C}$  is reflexive, then  $\Delta_{\mathbb{B}} \subset \bigcap_{a \in \mathbb{A}} \mathcal{C}_a$ .*

## 9.2. Recall on lattices and partitions

### 9.2.1. Lattices

The following definitions rely upon [Carpentier, Cohen, Chancelier, and De Lara \(2015\)](#). The *top element* (of a preordered set) is an element which is greater than any other element of the set; the top is denoted by  $\top$ . The bottom element  $\perp$  (of a preordered set) has a similar definition. A *maximum element* (of a subset) is an element of the subset which is greater than any other element of the subset. If it exists, it is unique, and it coincides with the top element if the subset is equal to the whole set. A *minimum element* (of a subset) has a similar definition. A maximal element (of a subset) is an element of the subset which is not less than any other element of the subset. A *majorant* (of a subset), also called *upper bound*, is an element not necessarily belonging to the subset, which is greater than any other element of the subset. If a majorant belongs to the subset, it is the maximum element. A *minorant* (of a subset), also called *lower bound*, has a similar definition. A *least upper bound* (of a subset) is the *least majorant*, that is, the minimum element of the subset of majorants. A greatest lower bound has a similar definition. A *sup-semilattice* is a preordered set such that there exists an upper bound for each pair of elements. An inf-semilattice has a similar definition. A *lattice* is a preordered set which is both a sup- and an inf-semilattice. A *complete sup-semilattice* is a preordered set such that there exists an upper bound for each finite or infinite subset. A *complete inf-semilattice* has a similar definition. A *complete lattice* has an obvious definition. A fundamental result asserts that a complete sup-semilattice which has a bottom element is a complete lattice.

### 9.2.2. Partition

**Definition 9.2.1.** Let  $\Omega$  be a set. A collection  $\mathcal{P}$  of subsets of  $\Omega$  is a *partition* if it consists of mutually disjoint nonempty subsets whose union is  $\Omega$ .

$$\mathcal{P} = \{\Omega_i\}_{i \in I} \text{ with } \begin{cases} \Omega = \cup_{i \in I} \Omega_i, \\ \Omega_i \cap \Omega_j = \emptyset, \forall i \neq j, \Omega_i \neq \emptyset, \forall i \in I. \end{cases} \quad (9.14)$$

**Definition 9.2.2.** The *least upper bound*  $\pi \vee \pi'$  of the two partitions  $\pi$  and  $\pi'$  is the partition made of all nonempty intersections between elements of  $\pi$  and elements of  $\pi'$ :  $G \in \pi \vee \pi'$  if, and only if,  $\exists i \in I$  and  $j \in I'$  such that  $G = \Omega_i \cap \Omega'_j$  and  $G \neq \emptyset$ .

We can extend the definition to a family of partition  $\pi_i$ . In this case, we use the notation  $\bigvee_i \pi_i$ .

**Definition 9.2.3.** Consider a collection  $\mathcal{G}$  of subsets of  $\Omega$ . An *atom* of  $\mathcal{G}$ , or a  $\mathcal{G}$ -atom is a nonempty subset  $G \in \mathcal{G}$  such that  $K \in \mathcal{G}$  and  $K \subset G$  imply that  $K = \emptyset$  or  $K = G$ .

**Definition 9.2.4.** Consider two partition fields  $\mathcal{G}^b$  and  $\mathcal{G}^\sharp$  on  $\Omega$ . When  $\mathcal{G}^b \subset \mathcal{G}^\sharp$  (that is,  $G \in \mathcal{G}^b \Rightarrow G \in \mathcal{G}^\sharp$ ),  $\mathcal{G}^\sharp$  is said to be *finer* than  $\mathcal{G}^b$ , and this is denoted by  $\mathcal{G}^b \preceq \mathcal{G}^\sharp$ .

## 9.3. Recall on stochastic kernels

This recall can be found on the appendix of [Carpentier, Cohen, Chancelier, and De Lara \(2015\)](#)

**Definition 9.3.1.** Let  $(\mathbb{X}, \mathcal{X})$  and  $(\mathbb{Y}, \mathcal{Y})$  be two measurable spaces. A *stochastic kernel* from  $(\mathbb{X}, \mathcal{X})$  to  $(\mathbb{Y}, \mathcal{Y})$  is a mapping  $p : \mathbb{X} \times \mathcal{Y} \rightarrow [0; 1]$  such that

- for any  $F \in \mathcal{Y}$ ,  $p(\cdot, F)$  is  $\mathcal{X}$ -measurable;
- for any  $x \in \mathbb{X}$ ,  $p(x, \cdot)$  is a probability on  $\mathcal{Y}$ .

A stochastic kernel may equivalently be seen as a measurable mapping from  $(\mathbb{X}, \mathcal{X})$  to  $\mathcal{P}(\mathcal{Y})$ . Thus, as for notation and terminology, we shall speak of a *stochastic kernel*  $p(x, dy)$  from  $\mathbb{X}$  to  $\mathbb{Y}$  or of a *stochastic kernel*  $p(dy|x)$  on  $\mathbb{Y}$  given  $\mathbb{X}$ .

Here is a composition operation on stochastic kernels.

**Definition 9.3.2.** Let  $(\mathbb{X}, \mathcal{X})$ ,  $(\mathbb{Y}, \mathcal{Y})$  and  $(\mathbb{Z}, \mathcal{Z})$  be three measurable spaces. Consider two stochastic kernels,  $p(dy | x)$  on  $\mathbb{Y}$  given  $\mathbb{X}$  and  $q(dz | y)$  on  $\mathbb{Z}$  given  $\mathbb{Y}$ . Then, the following expression defines a stochastic kernel  $p \otimes q$  on  $\mathbb{Z}$  given  $\mathbb{X}$ :

$$(p \otimes q)(F | x) = \int_{\mathbb{Y}} p(dy | x) \int_F q(dz | y), \quad \forall F \in \mathcal{Z}. \quad (9.15)$$

**Proposition 9.3.3.** Let  $\mathbb{Y}$  and  $\mathbb{Z}$  be Borel spaces and  $q \in \mathcal{P}(\mathbb{Y} \times \mathbb{Z})$ . Then, there exists a stochastic kernel  $r(dz|y)$  over  $\mathbb{Z}$  knowing  $\mathbb{Y}$  such that  $q = r \otimes s$  :

$$q(dy dz) = r(dz|y)s(dy) \text{ where } s(dy) = \int_{\mathbb{Z}} d(dy dz). \quad (9.16)$$

## 9.4. Recall on star-difference

The definition which follows comes from the book [Hiriart-Urruty and Lemaréchal \(2013\)](#) and introduces a concept of difference between convex sets. We consider for this section the vectorial space  $\mathbb{R}^n$  endowed with a norm denoted  $\|\cdot\|$ .

**Definition 9.4.1.** Given two convex sets  $C_1$  and  $C_2$ , we call *star-difference* the relation defined by

$$C_1 \underline{*} C_2 = \bigcap_{c_2 \in C_2} \{C_1 - c_2\} = \{a \in \mathbb{R}^n : a + C_2 \subset C_1\}. \quad (9.17)$$

The star difference  $C_1 \underline{*} C_2$  is a convex set even if  $C_2$  is not convex. It is a “small” set and it is very often empty. Here are some examples to illustrate the notion.

**Example 9.4.2.** Let  $C_1 = \mathcal{B}(o_1, r_1)$  and  $C_2 = \mathcal{B}(o_2, r_2)$  be two balls with  $0 < r_2 \leq r_1$ . We have then  $C_1 \underline{*} C_2 = \mathcal{B}(o_1 - o_2, r_1 - r_2)$ . Let  $C_1$  and  $C_2$  be two convex cones in the positive orthant. If  $C_2 \subset C_1$  then  $C_1 \underline{*} C_2 = C_1$  else  $C_1 \underline{*} C_2 = \emptyset$ .

The first property we state analyses the result of “adding” and “deleting” a set  $S$  to a set  $R$ .

**Proposition 9.4.3.** *Given a set  $S$ , we have*

$$S + (R \underline{*} S) \subset R. \quad (9.18)$$

*Proof.* Let  $x$  be an element of  $R \underline{*} S$  then by definition of the star-difference we have that

$$x + S \subset R. \quad (9.19)$$

The result (9.19) is true for all  $x \in R \underline{*} S$  which ends the proof.

$$(R \underline{*} S) + S \subset R. \quad (9.20)$$

□

**Proposition 9.4.4.** *Let  $C$  be a convex set. Given two convex sets  $D_1$  and  $D_2$  such that  $D_1 \subset D_2$ , we have*

$$C \underline{*} D_2 \subset C \underline{*} D_1. \quad (9.21)$$

*Proof.* By definition (9.17) of the star-difference,

$$C \underline{*} D_2 = \{x \in \mathbb{R}^n : x + D_2 \subset C\}, \quad (9.22)$$

Let  $x \in C \underline{*} D_2$  and  $d_1$  be an element of the set  $D_1$ . By assumption,  $d_1 \in D_2$  therefore  $x + d_1 \in C$ . This implies  $x + D_1 \subset C$  which ends the proof. □

The next proposition explores the effects of interverting star difference and intersection.

**Proposition 9.4.5.** *Given three convex sets  $C_1$ ,  $C_2$  and  $C_3$  we have*

$$(C_1 \underline{*} C_2) \cap (C_1 \underline{*} C_3) \subset C_1 \underline{*} (C_2 \cap C_3). \quad (9.23)$$

*Proof.* By definition we have

$$C_1 \star (C_2 \cap C_3) = \bigcap_{x \in C_2 \cap C_3} \{C_1 - x\}, \quad (9.24)$$

and

$$(C_1 \star C_2) \cap (C_1 \star C_3) = \underbrace{\bigcap_{y \in C_2} \{C_1 - y\}}_{I_2} \cap \underbrace{\bigcap_{y \in C_3} \{C_1 - y\}}_{I_3}. \quad (9.25)$$

We denote by  $R_2$  the rest  $C_2 \setminus (C_2 \cap C_3)$  and by  $R_3$  the rest  $C_3 \setminus (C_2 \cap C_3)$ . We have thus  $I_i = \bigcap_{y \in (C_2 \cap C_3)} \{C_1 - y\} \cap \bigcap_{y \in R_i} \{C_1 - y\}$ . We can rewrite equation (9.24)

$$(C_1 \star C_2) \cap (C_1 \star C_3) = \bigcap_{y \in (C_2 \cap C_3)} \{C_1 - y\} \cap \bigcap_{y \in R_2} \{C_1 - y\} \cap \bigcap_{y \in R_3} \{C_1 - y\}, \quad (9.26)$$

which by definition of the star difference reads

$$(C_1 \star C_2) \cap (C_1 \star C_3) = C_1 \star (C_2 \cap C_3) \cap \bigcap_{y \in R_2} \{C_1 - y\} \cap \bigcap_{y \in R_3} \{C_1 - y\}, \quad (9.27)$$

$$\subset C_1 \star (C_2 \cap C_3). \quad (9.28)$$

This ends the proof.  $\square$

## 9.5. Recall on risk measures

The results presented rely upon [Acerbi \(2002\)](#), [Ruszczyński \(2010\)](#) and [Föllmer and Schied \(2016\)](#). Let  $(\Omega, \mathcal{T})$  be a measurable space and  $\mathbb{P}$  be a probability distribution over  $\Omega$ . The set  $\mathcal{M}_1(\Omega)$  is the set of probability distributions over  $\Omega$  and let  $\mathbb{X}$  be a subspace of  $L^\infty(\Omega, \mathcal{T}, \mathbb{P})$ .

### 9.5.1. Definition of monetary risk measure

**Definition 9.5.1.** A risk measure  $\rho : \mathbb{X} \rightarrow \mathbb{R}$  is said to be

- *monotonous* if  $\mathbf{X} \leq \mathbf{Y} \Rightarrow \rho(\mathbf{X}) \leq \rho(\mathbf{Y})$ ,
- *translation invariant* if  $\forall m \in \mathbb{R}, \rho(\mathbf{X} + m) = \rho(\mathbf{X}) + m$ ,
- *convex* if  $\rho(\lambda \mathbf{X} + (1 - \lambda) \mathbf{Y}) \leq \lambda \rho(\mathbf{X}) + (1 - \lambda) \rho(\mathbf{Y})$  for  $\lambda \in [0; 1]$ ,
- *positively homogeneous* if  $\forall \lambda \geq 0, \rho(\lambda \mathbf{X}) = \lambda \rho(\mathbf{X})$

**Proposition 9.5.2.** A monetary risk measure  $\rho$  is Lipschitz continuous with respect to the norm  $\|\cdot\|_\infty$

A risk measure which is monetary, convex and positively homogeneous is said to be *coherent*.

### 9.5.2. Examples of risk measure

Before defining some well known risk measures, we need to introduce notation for cumulative distribution function

- $F_{\mathbf{X}}(x) = \mathbb{P}(\mathbf{X} \leq x)$ ,
- $F_{\mathbf{X}}^{-1}(p) = \inf \{x | F_{\mathbf{X}}(x) \geq p\}$ .

Let  $\alpha$  be a real in  $[0; 1]$ . We defined

#### Value at Risk:

$$\begin{aligned} VaR_{\alpha}(\mathbf{X}) &= \inf \{m \in \mathbb{R} | \mathbb{P}(\mathbf{X} \leq m) \geq \alpha\} , & (9.29a) \\ &= F_{\mathbf{X}}^{-1}(\alpha) . & (9.29b) \end{aligned}$$

#### Tail Conditional Expectation:

$$TCE_{\alpha}(\mathbf{X}) = \mathbb{E}_{\mathbb{P}}[\mathbf{X} | \mathbf{X} \geq VaR_{\alpha}] . \quad (9.30)$$

#### Entropic risk measure:

$$RE_{\alpha}(\mathbf{X}) = \frac{1}{1 - \alpha} \log (\mathbb{E}_{\mathbb{P}}[e^{(1-\alpha)\mathbf{X}}]) . \quad (9.31)$$

#### Worst Conditional Expectation:

$$WCE_{\alpha}(\mathbf{X}) = \sup \{ \mathbb{E}_{\mathbb{P}}[\mathbf{X} | A]; A \in \mathcal{F}, \mathbb{P}(A) \geq 1 - \alpha \} . \quad (9.32)$$

#### Conditional Value at Risk:

$$AV@R_{\alpha}[\mathbf{X}] = \inf s \in \mathbb{R} \left\{ \frac{\mathbb{E}[(\mathbf{X} - s)^+]}{1 - \alpha} + s \right\} . \quad (9.33)$$

#### Expected Shortfall:

$$ES_{\alpha}(\mathbf{X}) = \frac{1}{1 - \alpha} \mathbb{E}[(\mathbf{X} - VaR_{\alpha})^+] + VaR_{\alpha} . \quad (9.34)$$

#### Average Value at Risk :

For all  $\alpha$  in  $[0; 1)$

$$AVaR_{\alpha}(\mathbf{X}) = \frac{1}{1 - \alpha} \int_{\alpha}^1 VaR_p(\mathbf{X}) dp . \quad (9.35)$$

and by extension, for  $\alpha = 1$ , we use the Worst Case:

$$AVaR_1(\mathbf{X}) = \text{ess.sup}(\mathbf{X}) . \quad (9.36)$$

**Theorem 9.5.3.** *With previous definition we have*

$$CVaR_{\alpha} = ES_{\alpha} = AVaR_{\alpha} . \quad (9.37)$$

**Proposition 9.5.4.** *The risk measures  $VaR_{\alpha}$ ,  $TCE_{\alpha}$  and  $RE_{\alpha}$  are not coherent in general. the risk measure  $WCE_{\alpha}$  and  $AVaR_{\alpha}$  are coherent.*

### 9.5.3. Dual representation of coherent risk measure

#### General case

We recall that the topological dual of  $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$  is the set of finite  $\sigma$ -additive measures absolutely continuous with respect to  $\mathbb{P}$  and is denoted by  $\mathcal{M}_{1,f}$ .

**Theorem 9.5.5** (Representation theorem for convex risk measures). *Every convex monetary risk measure  $\rho$  on  $\mathbb{X}$  has the form*

$$\rho(\mathbf{X}) = \max_{\mathbb{Q} \in \mathcal{M}_{1,f}} (\mathbb{E}_{\mathbb{Q}}[X] - c_{\min}(\mathbb{Q})) , \quad (9.38)$$

where the penalty function  $c_{\min}$  is given by

$$c_{\min}(\mathbb{Q}) = \sup_{\mathbf{X} \in \mathcal{A}_\rho} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] . \quad (9.39)$$

Besides, every penalty function  $c$  is such that  $\rho(\mathbf{X}) = \max_{\mathbb{Q} \in \mathcal{M}_{1,f}} (\mathbb{E}_{\mathbb{Q}}[X] - c(\mathbb{Q}))$  satisfies  $c(\mathbb{Q}) \geq c_{\min}(\mathbb{Q})$

**Theorem 9.5.6** (Representation theorem for coherent risk measures). *The penalty function  $c_{\min}$  of a coherent risk measure  $\rho$  only takes the values 0 or  $+\infty$ . In particular,*

$$\rho(\mathbf{X}) = \max_{\mathbb{Q} \in \mathcal{Q}_{\max}} \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] . \quad (9.40)$$

with the convex set

$$\mathcal{Q}_{\max} := \{\mathbb{Q} \in \mathcal{M}_{1,f} \mid c_{\min} = 0\} . \quad (9.41)$$

#### Application to examples

##### Proposition 9.5.7.

$$RE_\alpha(\mathbf{X}) = \sup_{\mathbb{Q} \in \mathcal{M}_1} \left\{ \mathbb{E}_{\mathbb{Q}}[\mathbf{X}] - \frac{1}{1-\alpha} H(\mathbb{Q}|\mathbb{P}) \right\} . \quad (9.42)$$

where  $H(\mathbb{Q}|\mathbb{P}) := \mathbb{E}_{\mathbb{Q}}(\log \frac{d\mathbb{Q}}{d\mathbb{P}})$

##### Proposition 9.5.8.

The set  $\mathcal{Q}_{\max}$  associated with the risk measure  $AVaR_\alpha$  is defined by

$$\mathcal{Q}_{\max} := \left\{ \mathbb{Q} \in \mathcal{M}_1(\mathbb{P}); \frac{d\mathbb{Q}}{d\mathbb{P}} \leq \frac{1}{1-\alpha} \right\} . \quad (9.43)$$

The set  $\mathcal{Q}_{\max}$  associated with the risk measure  $WCE_\alpha$  is defined by

$$\mathcal{Q}_{\max} := \{ \mathbb{P}[\cdot | A]; A \in \mathcal{F}, \mathbb{P}[A] \geq 1 - \alpha \} . \quad (9.44)$$

We notice that when  $\alpha$  goes to 1, the set  $\mathcal{Q}_{\max}$  of the  $AVaR_\alpha$  goes to  $\mathcal{P}(\Omega)$  which defines the worst case risk measure.

## 9.6. Recall on Moreau's transform and Moreau's addition

We first present ingredients that are used to state our main contribution. The formal tools of coupling and Fenchel-Moreau conjugate were introduced in [Moreau](#). We recall that

$$\bar{\mathbb{R}} = [-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\} . \quad (9.45)$$

When we manipulate functions with values in  $\bar{\mathbb{R}} = [-\infty, +\infty]$ , we adopt the following Moreau *lower addition* or *upper addition*, depending on whether we deal with sup or inf operations. In the sequel,  $u, v$  and  $w$  are any elements of  $\bar{\mathbb{R}}$ .

### Moreau lower addition

The Moreau *lower addition* extends the usual addition with

$$(+\infty) \dagger (-\infty) = (-\infty) \dagger (+\infty) = -\infty . \quad (9.46a)$$

With the *lower addition*,  $(\bar{\mathbb{R}}, \dagger)$  is a convex cone, with  $\dagger$  commutative and associative. The lower addition displays the following properties:

$$u \leq u' , \quad v \leq v' \Rightarrow u \dagger v \leq u' \dagger v' , \quad (9.46b)$$

$$(-u) \dagger (-v) \leq -(u \dagger v) , \quad (9.46c)$$

$$(-u) \dagger u \leq 0 , \quad (9.46d)$$

$$\sup_{a \in \mathbb{A}} f(a) \dagger \sup_{b \in \mathbb{B}} g(b) = \sup_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dagger g(b)) , \quad (9.46e)$$

$$\inf_{a \in \mathbb{A}} f(a) \dagger \inf_{b \in \mathbb{B}} g(b) \leq \inf_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dagger g(b)) , \quad (9.46f)$$

$$t < +\infty \Rightarrow \inf_{a \in \mathbb{A}} f(a) \dagger t = \inf_{a \in \mathbb{A}} (f(a) \dagger t) . \quad (9.46g)$$

### Moreau upper addition

The Moreau *upper addition* extends the usual addition with

$$(+\infty) \dot{\dagger} (-\infty) = (-\infty) \dot{\dagger} (+\infty) = +\infty . \quad (9.47a)$$

With the *upper addition*,  $(\bar{\mathbb{R}}, \dot{\dagger})$  is a convex cone, with  $\dot{\dagger}$  commutative and associative. The upper addition displays the following properties:

$$u \leq u' , \quad v \leq v' \Rightarrow u \dot{\dagger} v \leq u' \dot{\dagger} v' , \quad (9.47b)$$

$$(-u) \dot{\dagger} (-v) \geq -(u \dot{\dagger} v) , \quad (9.47c)$$

$$(-u) \dot{\dagger} u \geq 0 , \quad (9.47d)$$

$$\inf_{a \in \mathbb{A}} f(a) \dot{\dagger} \inf_{b \in \mathbb{B}} g(b) = \inf_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{\dagger} g(b)) , \quad (9.47e)$$

$$\sup_{a \in \mathbb{A}} f(a) \dot{\dagger} \sup_{b \in \mathbb{B}} g(b) \geq \sup_{a \in \mathbb{A}, b \in \mathbb{B}} (f(a) \dot{\dagger} g(b)) , \quad (9.47f)$$

$$-\infty < t \Rightarrow \sup_{a \in \mathbb{A}} f(a) \dot{\dagger} t = \sup_{a \in \mathbb{A}} (f(a) \dot{\dagger} t) . \quad (9.47g)$$

**Joint properties of the Moreau lower and upper addition**

We obviously have that

$$u \dot{+} v \leq u \dot{+} v . \quad (9.48a)$$

The Moreau lower and upper additions are related by

$$-(u \dot{+} v) = (-u) \dot{+} (-v) , \quad -(u \dot{+} v) = (-u) \dot{+} (-v) . \quad (9.48b)$$

They satisfy the inequality

$$(u \dot{+} v) \dot{+} w \leq u \dot{+} (v \dot{+} w) . \quad (9.48c)$$

with

$$(u \dot{+} v) \dot{+} w < u \dot{+} (v \dot{+} w) \iff \begin{cases} u = +\infty \text{ and } w = -\infty , \\ \text{or} \\ u = -\infty \text{ and } w = +\infty \text{ and } -\infty < v < +\infty . \end{cases} \quad (9.48d)$$

Finally, we have that

$$u \dot{+} (-v) \leq 0 \iff u \leq v \iff 0 \leq v \dot{+} (-u) . \quad (9.48e)$$

**Background on Fenchel-Moreau conjugacy with respect to a coupling** Let be given two sets  $\mathbb{C}$  and  $\mathbb{C}^\sharp$ . Consider a *coupling* function  $\Phi : \mathbb{C} \times \mathbb{C}^\sharp \rightarrow [-\infty, +\infty]$ . We also use the notation  $\mathbb{C} \overset{\Phi}{\leftrightarrow} \mathbb{C}^\sharp$  for a coupling, so that

$$\mathbb{C} \overset{\Phi}{\leftrightarrow} \mathbb{C}^\sharp \iff \Phi : \mathbb{C} \times \mathbb{C}^\sharp \rightarrow [-\infty, +\infty] . \quad (9.49)$$

**Definition 9.6.1.** The *Fenchel-Moreau conjugate* of a function  $f : \mathbb{C} \rightarrow [-\infty, +\infty]$ , with respect to the coupling  $\Phi$  in (9.49), is the function  $f^\Phi : \mathbb{C}^\sharp \rightarrow [-\infty, +\infty]$  defined by

$$f^\Phi(c^\sharp) = \sup_{c \in \mathbb{C}} \left( \Phi(c, c^\sharp) \dot{+} (-f(c)) \right) , \quad \forall c^\sharp \in \mathbb{C}^\sharp . \quad (9.50)$$

The *Fenchel-Moreau biconjugate* is the function  $f^{\Phi\Phi} : \mathbb{C} \rightarrow [-\infty, +\infty]$  defined by

$$f^{\Phi\Phi}(c) = \sup_{c^\sharp \in \mathbb{C}^\sharp} \left( \Phi(c, c^\sharp) \dot{+} (-f^\Phi(c^\sharp)) \right) , \quad \forall c \in \mathbb{C} . \quad (9.51)$$





# 10. Translation invariant mappings on ordered group lattices

## Contents

---

<b>10.1. Translation invariant mappings on lattices</b> . . . . .	<b>169</b>
10.1.1. Basic notions . . . . .	169
10.1.2. Translation Invariant mapping and acceptance set . . . . .	170

---

## 10.1. Translation invariant mappings on lattices

Inspiring ourselves from [Föllmer and Schied \(2016\)](#), we are going to define a general framework for translation invariant mappings and acceptance sets. All the mappings will be defined on a set  $\mathbb{X}$ . We will first define a cone and use it to define what it means to be positive with a preorder coming from this cone. Then we will use this order to define acceptance set via an infimum operation.

### 10.1.1. Basic notions

Let  $\mathbb{I}$  be a subset of the set  $\mathbb{X}$  that is equipped with a partial order denoted by  $\leq$ . We recall that a *complete lattice*  $\mathbb{I}$  is a set such that there exists a greatest lower bound and a least upper bound for each (finite or infinite) subset. We denote by  $\perp$  the bottom element of the set  $\mathbb{I}$ , by  $\top$  the top element of the set  $\mathbb{I}$  by  $\wedge$  the infimum operation and by  $\vee$  the supremum operation.

**Definition 10.1.1.** Let  $\mathbb{F}$  be a subset of  $\mathbb{I}$ . We denote by  $\mathbb{F}^+$  the set

$$\mathbb{F}^+ = \{f \in \mathbb{F} \mid f \geq 0\} . \quad (10.1)$$

The reverse set  $\mathbb{F}^-$  has an evident definition. The set  $\mathbb{F}_*^+$  is the set

$$\{f \in \mathbb{F} \mid f > 0\} , \quad (10.2)$$

and the set  $\mathbb{F}_*^-$  has an evident definition. More generally, we define the lower set  $\downarrow \{i\}$  and the upper set  $\uparrow \{i\}$  of  $i$  by

$$\downarrow \{i\} = \{i' \in \mathbb{I} \mid i' \leq i\} , \quad (10.3a)$$

$$\uparrow \{i\} = \{i' \in \mathbb{I} \mid i' \geq i\} . \quad (10.3b)$$

Let  $\mathbb{F}$  be a subset of  $\mathbb{I}$

$$\downarrow \mathbb{F} = \{i' \in \mathbb{I} \mid i' \leq i \quad \forall i \in \mathbb{F}\} , \quad (10.4a)$$

$$\uparrow \mathbb{F} = \{i' \in \mathbb{I} \mid i' \geq i \quad \forall i \in \mathbb{F}\} . \quad (10.4b)$$

### 10.1.2. Translation Invariant mapping and acceptance set

We suppose that  $(\mathbb{I}, \oplus)$  is a commutative group whose neutral element is 0 i.e. the operation  $\oplus$  is intern commutative and associative, and each element  $i \in \mathbb{I}$  have an inverse denoted by  $\ominus i$  such that  $\forall i \in \mathbb{I}, i \oplus (\ominus i) = (\ominus i) \oplus i = 0$ .

Let  $\leq$  be a partial order on  $\mathbb{I}$ . We say that the order  $\leq$  is compatible with  $\oplus$  if

$$i_1 \leq i_2 \Rightarrow i_1 \oplus i_3 \leq i_2 \oplus i_3, \quad \forall (i_1, i_2, i_3) \in \mathbb{I}^3. \quad (10.5)$$

**Definition 10.1.2.** Let  $(\mathbb{I}, \oplus)$  be a non trivial commutative group and  $\gamma : \mathbb{I} \rightarrow \mathbb{X}$  be a given mapping. A  $\mathbb{I}$ -translation invariant mapping on  $\mathbb{X}$  is a mapping  $\rho : \mathbb{X} \rightarrow \mathbb{I}$  which satisfies

$$\rho(x - \gamma(i)) = \rho(x) \oplus (\ominus i), \quad \forall i \in \mathbb{I}. \quad (10.6)$$

When  $\mathbb{I} \subset \mathbb{X}$  and  $\gamma = \text{Id}$ , Equation (10.6) reads  $\rho(x - i) = \rho(x) - i$ .

**Definition 10.1.3.** Let  $\rho$  be a  $\mathbb{I}$ -translation invariant mapping on  $\mathbb{X}$ . and let  $\leq$  be a partial order on  $\mathbb{I}$  compatible with  $\oplus$ . The *acceptance set*  $\mathcal{A}_\rho$  of the mapping  $\rho$  is defined by

$$\mathcal{A}_\rho = \{x \in \mathbb{X} \mid \rho(x) \leq 0\}. \quad (10.7)$$

**Proposition 10.1.4.** Let  $\mathcal{A}$  be a subset of  $\mathbb{X}$ ,  $(\mathbb{I}, \oplus, \leq)$  a commutative group compatible with the partial order  $\leq$  which is also a complete lattice and  $\gamma : \mathbb{I} \rightarrow \mathbb{X}$  a given mapping. Then, the mapping  $\rho_{\mathcal{A}}$  defined by

$$\rho_{\mathcal{A}}(x) = \wedge \{i \in \mathbb{I} \mid x - \gamma(i) \in \mathcal{A}\}, \quad \forall x \in \mathbb{X}, \quad (10.8)$$

is an  $\mathbb{I}$ -translation invariant risk mapping.

*Proof.* Using the fact that  $\mathbb{I}$  is a complete lattice, the mapping  $\rho_{\mathcal{A}}$  is well defined and take values in  $\mathbb{I}$ . We fix an element  $j \in \mathbb{I}$ . We have the following equalities

$$\begin{aligned} \rho_{\mathcal{A}}(x - \gamma(j)) &= \wedge \{i \in \mathbb{I} \mid i \in \mathbb{I} \text{ and } x - \gamma(j) - \gamma(i) \in \mathcal{A}\}, \quad (\text{by definition (10.8),}) \\ &= \wedge \{i \oplus j \ominus j \mid i \in \mathbb{I} \text{ and } x - \gamma(j) - \gamma(i) \in \mathcal{A}\}, \end{aligned}$$

as  $j \ominus j = 0$  and associativity of  $\oplus$ ,

$$\begin{aligned} &= \wedge \{i \oplus j \ominus j \mid i \in \mathbb{I} \text{ and } x - \gamma(i \oplus j) \in \mathcal{A}\}, \quad (\text{as } \gamma \text{ is linear,}) \\ &= \wedge \{i \oplus j \mid i \in \mathbb{I} \text{ and } x - \gamma(i \oplus j) \in \mathcal{A}\} \ominus j, \end{aligned}$$

as  $j$  is a silent variable for the infimum operation and  $\leq$  is compatible with  $\oplus$ ,

$$\begin{aligned} &= \wedge \{i' \mid i' \ominus j \in \mathbb{I} \text{ and } x - \gamma(i') \in \mathcal{A}\} \ominus j, \quad (i' = i \oplus j) \\ &= \wedge \{i' \mid i' \in \mathbb{I} \text{ and } x - \gamma(i') \in \mathcal{A}\} \ominus j, \quad (\mathbb{I} = j + \mathbb{I}) \\ &= \rho_{\mathcal{A}}(x) \ominus j, \quad (\text{by definition (10.8)}) \end{aligned}$$

This ends the proof.  $\square$

**Proposition 10.1.5.** *Let  $\rho$  be a  $\mathbb{I}$ -translation invariant mapping on  $\mathbb{X}$ . Then we have*

$$\rho(x) = \rho_{\mathcal{A}_\rho}(x), \quad \forall x \in \mathbb{X}. \quad (10.9)$$

*Proof.* Let  $x$  be an element of  $\mathbb{X}$ . We have the following equalities

$$\rho_{\mathcal{A}_\rho}(x) = \wedge\{i \mid i \in \mathbb{I} \text{ and } x - \gamma(i) \in \mathcal{A}_\rho\}, \quad (\text{by Definition (10.8),})$$

$$= \wedge\{i \mid i \in \mathbb{I} \text{ and } \rho(x - \gamma(i)) \leq 0\}, \quad (\text{by Definition 10.1.3})$$

$$= \wedge\{i \mid i \in \mathbb{I} \text{ and } \rho(x) \ominus i \leq 0\}, \quad (\text{by Definition 10.1.2})$$

$$= \wedge\{i \mid i \in \mathbb{I} \text{ and } \rho(x) \ominus i \oplus i \leq 0 \oplus i\}, \quad (\text{by compatibility of } \oplus \text{ and } \leq)$$

$$= \wedge\{i \mid i \in \mathbb{I} \text{ and } \rho(x) \leq i\}, \quad (10.10)$$

$$= \rho(x), \quad (10.11)$$

as  $\rho(x)$  is a minorant and belongs to the minimizing set.  $\square$

In the next proposition we detail assumptions that can be given on a set  $\mathcal{A}$  in order to define a *proper* risk mapping  $\rho_{\mathcal{A}}$  i.e. a risk mapping which is not identically equals to the top element and never take the bottom element as value.

**Definition 10.1.6.** A mapping  $\rho : \mathbb{X} \rightarrow \mathbb{I}$  is said to be *proper* if

$$\exists x \in \mathbb{X}, \quad \rho(x) < \top, \quad (10.12)$$

$$\forall x \in \mathbb{X}, \quad \rho(x) > \perp. \quad (10.13)$$

**Proposition 10.1.7.** *Let  $\mathcal{A}$  be a subset of  $\mathbb{X}$ . If*

1.  $\mathcal{A} \neq \emptyset$ ,

2.  $\{i \mid i \in \mathbb{I} \text{ and } x - \gamma(i) \in \mathcal{A}\} \geq \alpha(x)$  for all  $x$  where  $\alpha : \mathbb{X} \rightarrow \mathbb{I} \setminus \{\top, \perp\}$ ,

*then the translation invariant risk mapping  $\rho_{\mathcal{A}}$  is proper.*

*By item 2 and definition (10.8),  $\rho_{\mathcal{A}}$  never takes the value  $\perp$ . This ends the proof.*

*Proof.* We fix  $x \in \mathcal{A}_\rho$  as  $\mathcal{A}_\rho$  is non empty. Then  $\rho_{\mathcal{A}}(x) \leq 0 < \top$  and  $\rho_{\mathcal{A}}$  takes at least one finite value.  $\square$

**Proposition 10.1.8.** *Let  $\rho$  be a  $\mathbb{I}$ -translation invariant mapping on  $\mathbb{X}$  that is proper. Then*

1.  $\mathcal{A}_\rho \neq \emptyset$ ,

2.  $\{i \mid i \in \mathbb{I} \text{ and } x - \gamma(i) \in \mathcal{A}_\rho\} \geq \rho(x)$  for all  $x$  where  $\alpha : \mathbb{X} \rightarrow \mathbb{I} \setminus \{\top, \perp\}$ .

*Proof.* As the mapping  $\rho$  is proper, we fix  $x \in \mathbb{X}$  such that  $\rho(x) < \top$ . Then,  $\rho(x - \gamma(\rho(x))) = \rho(x) \ominus \rho(x) = 0 \leq 0$  and the element  $x - \gamma(\rho(x))$  belongs to  $\mathcal{A}_\rho$ .

By Proposition 10.1.5, we know that  $\rho = \rho_{\mathcal{A}_\rho} = \wedge\{i \mid i \in \mathbb{I} \text{ and } x - \gamma(i) \in \mathcal{A}_\rho\}$ .

This ends the proof.  $\square$



**Part VI.**  
**Bibliography**



# Bibliography

- C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.
- S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 131–142, 1966.
- K. J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pages 265–290, 1954.
- P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and Bellman’s principle. *Annals of Operations Research*, 152(1):5–22, 2007a.
- P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and bellman’s principle. *Annals of Operations Research*, 152(1):5–22, 2007b.
- M. Basseville. Divergence measures for statistical data processing—an annotated bibliography. *Signal Processing*, 93(4):621–633, 2013.
- H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
- A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.
- A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3):411–424, 2000.
- A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University Press, 2009.
- A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.



## Bibliography

- D. P. Bertsekas. *Dynamic Programming and Stochastic Control*. Academic Press, 1976.
- D. P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice Hall, Englewood Cliffs, 1987.
- T. R. Bielecki, I. Cialenco, and M. Pitera. A survey of time consistency of dynamic risk measures and dynamic performance measures in discrete time: Lm-measure perspective. *Probability, Uncertainty and Quantitative Risk*, 2(1):3, 2017.
- L. M. Briceno-Arias, G. Chierchia, E. Chouzenoux, and J.-C. Pesquet. A random block-coordinate douglas-rachford splitting method with low computational complexity for binary logistic regression. *arXiv preprint arXiv:1712.09131*, 2017.
- A. Brook, D. Kendrick, and A. Meeraus. GAMS, a user's guide. *ACM Signum Newsletter*, 23(3-4):10–11, 1988.
- P. P. Butzer et al. Fundamentals of semi-group theory. In *Semi-Groups of Operators and Approximation*, pages 1–82. Springer, 1967.
- P. Carpentier, G. Cohen, J.-P. Chancelier, and M. De Lara. *Stochastic Multi-Stage Optimization*. Springer, 2015.
- A. Chambolle and C. Dossal. *Journal of Optimization Theory and Applications*, 166(3):968–982, 2015.
- P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11(3):57–106, 2006.
- P. L. Combettes. Strong convergence of block-iterative outer approximation methods for convex optimization. *SIAM Journal on Control and Optimization*, 38(2):538–565, 2000.
- P. L. Combettes. A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Transactions on Signal Processing*, 51(7):1771–1782, 2003.
- P. L. Combettes and C. L. Müller. Perspective functions: Proximal calculus and applications in high-dimensional statistics. *Journal of Mathematical Analysis and Applications*, 457(2):1283–1306, 2018.
- P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2010.
- P. L. Combettes, D. Dung, and B. C. Vũ. Dualization of signal recovery problems. *Set-Valued and Variational Analysis*, 18(3):373–404, Dec. 2010.

- I. Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, 8:85–108, 1964.
- M. De Lara and V. Leclère. Building up time-consistency for risk measures and dynamic optimization. *European Journal of Operations Research*, 249(1):177 – 187, 2016.
- G. de Maere d’Aertrycke, A. Ehrenmann, and Y. Smeers. Investment with incomplete markets for risk: The need for long-term contracts. *Energy Policy*, 105: 571–583, 2017.
- D. Dentcheva and A. Ruszczyński. Risk forms: Representation, disintegration, and application to partially observable two-stage systems. *arXiv preprint arXiv:1807.02300*, 2018.
- K. Detlefsen and G. Scandolo. Conditional and dynamic convex risk measures. *Finance and Stochastics*, 9(4):539–561, October 2005.
- S. Dreyfus. Richard Bellman on the birth of dynamic programming. *Operations Research*, 50(1):48–51, 2002.
- J. Duchi, P. Glynn, and H. Namkoong. Statistics of robust optimization: A generalized empirical likelihood approach. *arXiv preprint arXiv:1610.03425*, 2016.
- A. Ehrenmann and Y. Smeers. Generation capacity expansion in a risky environment: a stochastic equilibrium analysis. *Operations Research*, 59(6):1332–1346, 2011.
- L. G. Epstein and M. Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31, 2003a.
- L. G. Epstein and M. Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31, 2003b.
- P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116*, 2015.
- P. M. Esfahani, S. Shafieezadeh-Abadeh, G. A. Hanasusanto, and D. Kuhn. Data-driven inverse optimization with imperfect information. *Mathematical Programming*, pages 1–44, 2017.
- F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *4OR: A Quarterly Journal of Operations Research*, 5(3):173–210, 2007.
- M. C. Ferris and T. S. Munson. Complementarity problems in GAMS and the PATH solver. *Journal of Economic Dynamics and Control*, 24(2):165–188, 2000.
- M. C. Ferris, S. P. Dirkse, J.-H. Jagla, and A. Meeraus. An extended mathematical programming framework. *Computers & Chemical Engineering*, 33(12):1973–1982, 2009.

## Bibliography

- H. Föllmer and A. Schied. *Stochastic finance: an introduction in discrete time. Fourth revised and extended edition.* Walter de Gruyter, 2016.
- H. Gérard, M. De Lara, and J.-P. Chancelier. Equivalence between time consistency and nested formula. *arXiv preprint arXiv:1711.08633*, 2017.
- H. Gérard, V. Leclère, and A. Philpott. On risk averse competitive equilibrium. *Operations Research Letters*, 46(1):19–26, 2018.
- J.-y. Gotoh, M. J. Kim, and A. E. Lim. Robust empirical optimization is almost the same as mean–variance optimization. *Operations Research Letters*, 2018.
- P. J. Hammond. Changing tastes and coherent dynamic choice. *The Review of Economic Studies*, 43(1):159–173, February 1976.
- Y. Haugazeau. Sur les inéquations variationnelles et la minimisation de fonctionnelles convexes. *These, Universite de Paris*, 1968.
- J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I: fundamentals*, volume 305. Springer science & business media, 2013.
- Z. Hu and L. J. Hong. Kullback-leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, 2013.
- W. Imrich. Factoring cardinal product graphs in polynomial time. *Discrete Mathematics*, 192(1):119–144, 1998.
- A. A. Khan, C. Tammer, and C. Zalinescu. *Set-valued optimization.* Springer, 2015.
- T. C. Koopmans. Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, pages 287–309, 1960.
- D. Kreps and E. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, 46(1):185–200, 1978a.
- D. M. Kreps and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, pages 185–200, 1978b.
- D. M. Kreps and E. L. Porteus. Temporal von Neumann-Morgenstern and induced preferences. *Journal of Economic Theory*, 20(1):81–109, February 1979.
- A. Kurakin, I. Goodfellow, and S. Bengio. Adversarial examples in the physical world. *arXiv preprint arXiv:1607.02533*, 2016.
- B. S. Laetitia Andrieu, Michel De Lara. Taking risk into account in electricity portfolio management. In M. P. Steffen Rebennack, Panos M. Pardalos and N. Iliadis, editors, *Handbook of Power Systems 1.* Springer Verlag, 2009.
- H. R. Lewis and C. H. Papadimitriou. *Elements of the Theory of Computation.* Prentice Hall PTR, 1997.

- R. Mattheij and J. Molenaar. *Ordinary differential equations in theory and practice*. SIAM, 2002.
- S. Moghaddam and M. Mahlooji. Robust simulation optimization using  $\varphi$ -divergence. *International Journal of Industrial Engineering Computations*, 7(4): 517–534, 2016.
- J. Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *Journal of Mathematics*, 63:361–382.
- T. Morimoto. Markov processes and the h-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.
- H. Namkoong and J. C. Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems*, pages 2208–2216, 2016.
- N. Papernot, P. McDaniel, and I. Goodfellow. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. *arXiv preprint arXiv:1605.07277*, 2016.
- B. Peleg and M. E. Yaari. On the existence of a consistent course of action when tastes are changing. *Review of Economic Studies*, 40(3):391–401, July 1973.
- G. C. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer, 2014.
- A. Philpott, M. Ferris, and R. Wets. Equilibrium, uncertainty and risk in hydro-thermal electricity systems. *Mathematical Programming*, 157(2):483–513, 2016. ISSN 1436-4646.
- G. Pritchard, G. Zakeri, and A. Philpott. A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Operations Research*, 58:1210–1219, 2010.
- D. Ralph and Y. Smeers. Risk trading and endogenous probabilities in investment equilibria. *SIAM Journal on Optimization*, 25(4):2589–2611, 2015.
- F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112(2):185–200, 2004.
- R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–41, 2000a.
- R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000b.
- R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1):33–53, 2013.
- R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

## Bibliography

- R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Risk tuning with generalized linear regression. *Mathematics of Operations Research*, 33(3):712–729, 2008.
- A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, 125(2):235–261, 2010.
- A. Ruszczyński. Risk-averse dynamic programming for markov decision processes. *Mathematical programming*, 125(2):235–261, 2010.
- A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of operations research*, 31(3):544–561, 2006a.
- A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452, 2006b.
- P. A. Samuelson. The stability of equilibrium: comparative statics and dynamics. *Econometrica: Journal of the Econometric Society*, pages 97–120, 1941.
- G. Schmidt and T. Ströhlein. *Relations and graphs: discrete mathematics for computer scientists*. Springer Science & Business Media, 2012.
- S. Shafieezadeh-Abadeh, P. M. Esfahani, and D. Kuhn. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pages 1576–1584, 2015.
- A. Shapiro. On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters*, 37(3):143–147, 2009.
- A. Shapiro. Rectangular sets of probability measures. *Operations Research*, 64(2):528–541, 2016.
- A. Shapiro, A. Ruszczyński, and D. Dentcheva. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2014.
- R. H. Strotz. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180, 1955-1956.
- H. Uzawa. Walras’ tatonnement in the theory of exchange. *The Review of Economic Studies*, 27(3):182–194, 1960.
- S. Wong and J. Fuller. Pricing energy and reserves using stochastic optimization in an alternative electricity market. *IEEE Transactions on Power Systems*, 22(2):631–638, 2007.
- G. Zakeri, G. Pritchard, M. Bjorndal, and E. Bjorndal. Pricing wind: A revenue adequate cost recovering uniform price for electricity markets with intermittent generation. *NHH Dept. of Business and Management Science*, 2016.