Projet de Fin d'Études (PFE)
ENSTA PARIS — MPRO
Parcours : Sciences des Données et de l'Optimisation (SOD)
Année scolaire : 2022-2023

Perturbation-Duality Scheme in Combinatorial Optimization and Algorithms in Generalized Convexity

NON CONFIDENTIEL

Auteur : Seta Rakotomandimby **Promotion** : 2023

Enseignante référente
ENSTA PARIS et MPRO
Sourour Elloumi

Encadrant
CERMICS
Michel de Lara

Stage effectué du 11/04/2023 au 29/09/2023
Organisme d'accueil : CERMICS, École Nationale des Ponts et Chaussées Paris Tech
Cité Descartes, 6-8 Avenue Blaise Pascal, 77455 Champs-Sur-Marne, France

ENSTA

IP PARIS

Attestation de confidentialité / non-confidentialité

A remplir par l'entreprise d'accueil et à remettre à la bibliothèque d'ENSTA Paris

Rapport de Projet de Recherche (PFE)

Par la présente, je soussigné (e) ~~Madame~~ / Monsieur [1] NOM, Prénom

DE LARA, Michel

Employé(e) en tant que (qualité)

chercheur

dans la société (NOM et ADRESSE de l'établissement) :

ECOLE DES PONTS

Atteste sur l'honneur que LES DONNÉES contenues dans le rapport de

~~Madame~~ / Monsieur[1] NOM Prénom RAKOTOMANDIMBY Seta **SONT:**

☐ **CONFIDENTIELLES**

Par conséquent, **l'entreprise** (ou organisme) d'accueil **S'ENGAGE A CONSERVER** le rapport de stage **pendant la Durée d'Utilité Administrative (DUA) soit 5 ans.**

A l'issue de la DUA, l'entreprise s'engage (cocher le choix correspondant) :

☐ A conserver le document

☐ A détruire le document

☒ **NON CONFIDENTIELLES** (**cocher** le choix correspondant concernant la publication des métadonnées sur internet)

☒ L'entreprise autorise une **mise en ligne du rapport de stage.** Dans ce cas, seules les métadonnées (titre, résumé mots-clés, entreprise, noms de l'élève et des encadrants...) sont accessibles à tous. Le rapport lui-même n'est accessible qu'aux personnes de la communauté ENSTA Paris). L'étudiant doit déposer son rapport sur le site des **archives institutionnelles de l'école** BibNum - **https://bibnum.ensta.fr.**

☐ L'entreprise autorise un accès restreint au document (uniquement sur place et au format électronique). L'étudiant doit déposer son rapport sur le site des **archives institutionnelles de l'école** BibNum - **https://bibnum.ensta.fr** mais il ne sera pas mis en ligne.

Un rapport consultable uniquement sur place est stocké sur un PC non connecté au réseau et à internet. Les copies électroniques et imprimées ne sont pas autorisées.

À Champs s/Marne Le 18-09-2023

Signature et cachet

Cette attestation doit être transmise par mail à documentation@ensta-paris.fr, si le document est confidentiel. Si le rapport est non confidentiel, elle devra être déposée par l'élève en même temps que le rapport sur https://bibnum.ensta.fr.

[1] Rayer la mention inutile.

Abstract

This report is structured into two parts, each encompassing an aspect of Operations Research. In the first part, we delve into the duality of integer linear programs using the Rockafellar perturbation-duality scheme. The second part focuses on the numerical evaluation of a cutting plane algorithm applied to a class of generalized convex problems. Through these two distinct parts, we explore both theoretical insights and practical applications.

Contents

Introduction

Ce rapport comportent deux parties distinctes. Dans la première partie, on étudie la dualité des programmes linéaires en nombres entiers à l'aide du schéma de perturbation-dualité de Rockafellar. Dans la deuxième partie, nous testons numériquement un algorithme de plans coupants sur des problèmes parcimonieux en convexité généralisée.

Partie 1 : Schéma de perturbation-dualité en optimisation combinatoire

Les résultats de dualité en programmation linéaire continue sont bien connus et largement utilisés dans des méthodes de résolution — comme par exemple l'algorithme du simplexe dual dans un «branch-and-cut» ou en programmation bi-niveau. Il n'en va pas de même pour les résultats de dualité en programmation linéaire en nombres entiers (PLNE) qui restent plus méconnus et peu, voire pas, utilisés. Dans [19], Jeroslow a introduit, pour un programme PLNE, un problème dual sous-additif qui possède la propriété de dualité forte avec le PLNE d'origine et les conditions de complémentarité sur les solutions que l'on appelle usuellement «primales-duales».

Avant d'aller plus loin, il faut préciser ce qu'on entend par «dualité». En cours de programmation linéaire au MPRO, la «dualité» d'un programme linéaire est présentée sous la forme d'un autre programme linéaire, obtenu en appliquant certaines règles de calcul au programme linéaire initial. La «dualité» dans le schéma de perturbation-dualité de Rockafellar, que (nous allons présenter dans le Chapitre 1) correspond à un couplage $c : \mathcal{U} \times \mathcal{V} \to \mathbb{R} \cup \{-\infty, +\infty\}$ entre un espace primal $\mathcal{U}$ et un espace dual $\mathcal{V}$. Cette deuxième notion de «dualité», que nous allons adopter pour le reste du rapport, permet de récupérer la première notion de «dualité» d'un programme linéaire continue: on applique le schéma de perturbation-dualité au programme linéaire d'origine, avec une perturbation du membre de droite des constraintes et le produit scalaire $\langle \cdot, \cdot \rangle : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ comme couplage.

L'enjeu et la contribution de cette première partie du rapport est d'appliquer la méthodologie du schéma de perturbation-dualité de Rockafellar à la PLNE, d'étudier ce que cette méthodologie *systématique* apporte à la compréhension de la dualité en PLNE, et de retrouver notamment le programme dual que Jeroslow a défini.

Partie 2 : Algorithmes en convexité généralisée

En convexité généralisée, le produit de dualité est remplacé par un couplage c, la conjugaison de Fenchel est remplacée par la conjugaison associée au couplage. Les fonctions convexes fermées sont remplacées par des fonctions c-convexes, (définies comme des fonctions égales à leurs biconjuguées) [25]. Ce domaine a principalement fait l'objet de travaux théoriques [28, 32, 35] et, plus rarement, d'approches algorithmiques [28, Chapitre 9], [32, Chapitre 9]. Des travaux récents [8] sur la pseudonorme ℓ_0 ont mis en lumière un couplage et une conjugaison, dits E-Capra, qui pourraient ouvrir la voie à l'application d'algorithmes de convexité généralisée à l'optimisation parcimonieuse.

Dans ce contexte, nous avons défini, en convexité généralisée, la distance de Bregman pour un couplage quelconque et l'opérateur proximal pour les couplages unilatéralement linéaires (OSL) qui permettraient de généraliser les méthodes proximales de la convexité ordinaire. D'autre part, nous avons implémenté et adapté la méthode des plans coupants abstrait à des problèmes E-Capra convexes et nous avons testé son efficacité sur trois problèmes E-Capra spécifiques.

Plan et contributions personnelles

Dans le Chapitre 1 nous présentons le schéma de perturbation-dualité de Rockafellar et comment la «relaxation Lagrangienne» (que nous appelons relaxation «Geoffrion Lagrangienne») peut être interprétée comme un schéma de perturbation-dualité.

Dans le Chapitre 2, nous présentons les résultats sur la dualité des PLNE de Jeroslow et nous les retrouvons en appliquant appliquant cinq schémas de perturbations-dualité à la PLNE. Notre contribution se trouve :

- dans les Tableaux 2.1, 2.2, 2.3 et 2.4 qui résument notre application des cinq schémas à la PLNE ;
- dans la Proposition 2.11, où nous exhibons un lien entre les conditions de complémentarité des schémas avec le sous-différentiel sous-additive de la fonction valeur ;
- dans la Proposition 2.19, où nous donnons l'exemple d'un nouveau programme dual quasi-affine.

Dans le Chapitre 3, nous appliquons la méthodologie du schéma de perturbation-dualité à des problèmes linéaires en variables binaires. Nous traitons le cas particulier du problème du sac à dos. Notre contribution se trouve

- dans la Proposition 3.6 où nous établissons un résultat de dualité forte pour un PLNE où seule une partie des contraintes est perturbée (similairement à la relaxation Lagrangienne «à la Geoffrion» mais avec un nouvel espace dual).

Dans le Chapitre 4, nous présentons des algorithmes d'optimisation globale appliquées à la minimisation de fonction convexes abstraites. On y retrouve la méthode des plans coupants, le «branch-and-bound» et la recherche tabou. Nous présentons aussi une généralisation de l'opérateur proximal pour les couplages dits OSL (One Sided Linear). Notre contribution se trouve

- dans les six tableaux de §4.2.1 où nous rassemblons une étude systématique de la conjugaison OSL.

Dans le Chapitre 5, nous présentons la conjugaison E-Capra (un cas particulier d'OSL), trois problèmes E-Capra, l'adaptation des plans coupants à ces problèmes et les résultats numériques de nos expériences numériques. Notre contribution se trouve

- en la production d'un code Julia de résolution de ces trois problèmes décrit en §5.2;
- en la génération par un code Julia d'instances des problèmes décrit en §5.3.1;
- en la présentation des résultats numériques en §5.3.2 et §5.3.3 qui tendent à montrer que la méthode de plans coupants convergent pour des problèmes d'optimisation parcimonieux et sont prometteurs, en particulier pour le problème du spark d'une matrice (voir la Figure 5.3.4).

Part I

Perturbation-duality scheme in combinatorial optimization

Chapter 1

The Rockafellar perturbation-duality scheme

In §1.1, we present the framework of the perturbation-duality scheme with coupling. In §1.2, we show how the perturbation-duality scheme covers the well-known 'Geoffrion Lagrangian relaxation'.

1.1 Perturbation-duality scheme with generalized coupling

Introduced in 1974 in its modern form by Rockafellar [30], the goal of the perturbation-duality scheme is to define a framework that allows to systematically produce a dual optimization problem from a given optimization problem. This framework relies on the choice of perturbation of the original problem and on the choice of a coupling between the perturbation space and a dual space.

While Rockafellar was considering bilinear couplings (basically the scalar product of $\mathbb{R}^n$ and the Fenchel conjugacy), Balder considered general couplings [1] which linked the notions of weak and strong duality with the one of abstract convexity obtained by a coupling (also see [34, 32]).

In §1.1.1, we outline each step of the perturbation-duality scheme with general coupling. In §1.1.2, we sum-up in a table the usual functions of the perturbation-duality scheme.

1.1.1 Outline of the general perturbation-duality scheme

Here we present an outline of the scheme and sum up the usual objects encountered in the scheme. We remind that $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and that the Moreau additions $\dot{+}, \underset{\cdot}{+}$ extend the usual addition over $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ by

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty \ , \tag{1.1a}$$

$$(+\infty) \underset{\cdot}{+} (-\infty) = (-\infty) \underset{\cdot}{+} (+\infty) = -\infty \ . \tag{1.1b}$$

Original optimization problem

The perturbation-duality scheme is applied to an minimization problem

$$\inf_{w \in \mathcal{W}} h(w) \ , \tag{1.2}$$

defined by some set $\mathcal{W}$ and an objective function $h : \mathcal{W} \to \overline{\mathbb{R}}$.

Perturbation of the original problem by Rockafellian

We first choose a perturbation of the original minimization problem (1.2). To do so, we introduce a *Rockafellian* $\mathfrak{R} : \mathcal{W} \times \mathcal{U} \to \overline{\mathbb{R}}$ that will represent the perturbation. The perturbation is parametrized by a *(primal-)perturbation set* $\mathcal{U}$.

Definition 1.1 ([31])**.** *We say that the bivariate function* $\mathfrak{R} : \mathcal{W} \times \mathcal{U} \to \overline{\mathbb{R}}$ *is a* Rockafellian *for the original minimization problem* (1.2) *if there is* $\bar{u} \in \mathcal{U}$*, called* anchor*, such that*

$$h(w) = \mathfrak{R}(w, \bar{u}) \ , \ \ \forall w \in \mathcal{W} \ . \tag{1.3}$$

Thus, we obtain a family of perturbed minimization problems $\{\inf_{w \in \mathcal{W}} \mathfrak{R}(w, u)\}_{u \in \mathcal{U}}$ such that $\inf_{w \in \mathcal{W}} \mathfrak{R}(w, \bar{u})$ is the original minimization problem (1.2).

The perturbation function

As the original minimization problem (1.2) has been parametrized by the elements of the set $\mathcal{U}$, we can consider the value of the minimization problem for each perturbation $u \in \mathcal{U}$. This leads to the definition of the *perturbation function*.

Definition 1.2. *For a given Rockafellian* $\mathfrak{R} : \mathcal{W} \times \mathcal{U} \to \overline{\mathbb{R}}$ *of the original minimization problem* (1.2)*, the function* $\varphi : \mathcal{U} \to \overline{\mathbb{R}}$ *defined by*

$$\varphi(u) = \inf_{w \in \mathcal{W}} \mathfrak{R}(w, u) \ , \ \ \forall u \in \mathcal{U} \ , \tag{1.4}$$

is called the perturbation function*.*

Coupling the perturbation set with a dual set

Now that the perturbation of the original problem has been set, the last choice that needs to be made to define a dual problem of (1.2) is the choice of the coupling $c : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$ between the perturbation set $\mathcal{U}$ and a dual set $\mathcal{V}$.

It is worth noting that we do not assume any structure on the perturbation and dual sets, nor on the coupling function c.

Lagrangian function and dual function

Two other functions of the perturbation-duality scheme are the *Lagrangian function* and the *dual function.* We refer the reader to [11] for a study of the duality between Rockafellians and Lagrangians.

Definition 1.3.

- *For a given Rockafellian* $\mathfrak{R} : \mathcal{U} \to \overline{\mathbb{R}}$ *of the minimization problem* (1.2) *and a given coupling* $c : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$*, we call* the Lagrangian function *the function* $\mathcal{L} : \mathcal{W} \times \mathcal{V} \to \overline{\mathbb{R}}$ *defined by*

$$\mathcal{L}(w, v) = \inf_{u \in \mathcal{U}} \left\{ \mathfrak{R}(w, u) \dot{+} \left(-c(u, v) \right) \right\} , \quad \forall w \in \mathcal{W} , \quad \forall v \in \mathcal{V} . \tag{1.5a}$$

- *For a given Rockafellian* $\mathfrak{R} : \mathcal{U} \to \overline{\mathbb{R}}$ *of the minimization problem* (1.2) *and a given coupling* $c : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$*, we call* the dual function *the function* $\Psi : \mathcal{V} \to \overline{\mathbb{R}}$ *defined by*

$$\Psi(v) = \inf_{w \in \mathcal{W}} \mathcal{L}(w, v) , \quad \forall v \in \mathcal{V} . \tag{1.5b}$$

The dual objective function and the dual problem

We can now introduce the *dual problem* given by the *dual objective function* defined by the Rockafellian and the coupling.

Definition 1.4.

- *For a given Rockafellian* $\mathfrak{R} : \mathcal{W} \times \mathcal{U} \to \overline{\mathbb{R}}$ *of the original minimization problem* (1.2) *with an anchor* $\bar{u} \in \mathcal{U}$*, and a given coupling* $c : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$*, we call* dual objective function *the function* $\Phi_{\bar{u}} : \mathcal{V} \to \overline{\mathbb{R}}$ *defined by*

$$\Phi_{\bar{u}}(v) = c(\bar{u}, v) \underset{\cdot}{+} (-\varphi^{c}(v)) , \quad \forall v \in \mathcal{V} . \tag{1.6a}$$

- *The* generalized maximization dual problem *is*

$$\sup_{v \in \mathcal{V}} \Phi_{\bar{u}}(v) = \sup_{v \in \mathcal{V}} c(\bar{u}, v) \underset{\cdot}{+} (-\varphi^{c}(v)) . \tag{1.6b}$$

Generalized weak and strong duality

Now that we have introduced the generalized maximization dual problem (1.6b), we can define the notions of weak and strong duality using the definition of c-convexity.

By properties of biconjugates (A.10) , we have that

$$\sup_{v \in \mathcal{V}} \Phi_{\bar{u}}(v) = \varphi^{cc'}(\bar{u}) \leq \varphi(\bar{u}) = \inf_{w \in \mathcal{W}} h(w) , \tag{1.7}$$

thus giving *weak duality.* Furthermore, if the perturbation function φ is c-convex at $\bar{u}$, which means $\varphi^{cc'}(\bar{u}) = \varphi(\bar{u})$ according to Definition A.6, we have *strong duality.*

1.1.2 Summary of the perturbation-duality scheme functions

Here we sum up the functions that arise in the perturbation-duality scheme and the relations between them. The Table 1.1 is taken from [11, Table 3], where the proofs of the properties are also given. We have added the dual objective function.

bivariate functions	univariate functions	definition	property
Rockafellian $\mathfrak{R} : \mathcal{W} \times \mathcal{U} \to \overline{\mathbb{R}}$			
Lagrangian $\mathcal{L} : \mathcal{W} \times \mathcal{V} \to \overline{\mathbb{R}}$		$\mathcal{L}(w, v) = \inf_{u\in\mathcal{U}} \big\{\mathfrak{R}(w, u) \dotplus \big(-c(u, v)\big)\big\}$	$-\mathcal{L}(w, \cdot) = \big(\mathfrak{R}(w, \cdot)\big)^c$
	perturbation $\varphi : \mathcal{U} \to \overline{\mathbb{R}}$	$\varphi(u) = \inf_{w\in\mathcal{W}} \mathfrak{R}(w, u)$	
	dual $\Psi : \mathcal{V} \to \overline{\mathbb{R}}$	$\Psi(v) = \inf_{w\in\mathcal{W}} \mathcal{L}(w, v)$	$-\Psi = \varphi^c$
	dual objective $\Phi_{\bar{u}} : \mathcal{V} \to \overline{\mathbb{R}}$	$\Phi_{\bar{u}}(v) = c(\bar{u}, v) \overset{\cdot}{+} (-\varphi^c(v))$	$\Phi_{\bar{u}}(v) = c(\bar{u}, v) \overset{\cdot}{+} \Psi(v)$ $\varphi^{cc'}(\bar{u}) = \sup_{v\in\mathcal{V}} \Phi_{\bar{u}}(v)$

Table 1.1: Functions in the perturbation-duality scheme

1.2 Hidden perturbation-duality scheme in Geoffrion Lagrangian relaxation

In §1.2.1, we provide a simple background on the so-called Geoffrion Lagrangian relaxation. Then, in §1.2.2 we show that the Geoffrion Lagrangian relaxation is a special case of the perturbation-duality scheme.

1.2.1 Background on Geoffrion Lagrangian relaxation

In 1974, Geoffrion published his seminal article [14] where he coined the term 'Lagrangian relaxation' for the following method. We consider a pure integer linear program with constraints split in two blocks

$$\begin{aligned} \alpha_{\mathrm{PILP}} = \ & \inf_x && \langle k,\ x\rangle \ , \\ & \text{s.t.} && Ax = \bar{b}\ , \\ & && \tilde{A}x = \tilde{b}\ , \\ & && x \in \mathbb{Z}^n_+\ , \end{aligned} \tag{1.8}$$

where $k \in \mathbb{Q}^n$, $A \in \mathbb{Q}^{m\times n}$, $\tilde{A} \in \mathbb{Q}^{\tilde{m}\times n}$, $\bar{b} \in \mathbb{Q}^m$ and $b' \in \mathbb{Q}^{\tilde{m}}$. We relax the first block of constraint using a Lagrangian multiplier $\lambda \in \mathbb{R}^m$, thus defining a concave function $g : \mathbb{R}^m \to$

$\overline{\mathbb{R}}$, which Geoffrion called 'Lagrangian function' [1], given by

$$g(\lambda) = \begin{array}{ll} \inf\limits_{x} & \langle k,\, x\rangle + \left\langle \lambda,\, \bar{b} - Ax\right\rangle \ , \\ \text{s.t.} & \tilde{A}x = \tilde{b} \ , \\ & x \in \mathbb{Z}_+^n \ . \end{array} \tag{1.9}$$

For every $\lambda \in \mathbb{R}^m$, the quantity $g(\lambda)$ gives a lower estimate of the value α_{PILP} of the PILP (1.8). Furthermore, if we look for the best one, that means $\sup_{\lambda\in\mathbb{R}^m} g(\lambda)$, we get a potentially better lower estimate than the continuous relaxation, as stated by the following theorem.

Theorem 1.5. *[14, Theorem 1] [10, Theorem 8.2] Suppose that the PILP* (1.8) *is feasible. If we denote α_{LP} the value of the continuous relaxation of the PILP* (1.8)*, which is the PILP* (1.8) *where $x \in \mathbb{Z}_+^n$ have been replaced by $x \in \mathbb{R}_+^n$, then*

$$\alpha_{\text{LP}} \leq \sup_{\lambda\in\mathbb{R}^m} g(\lambda) = \inf\left\{\langle k,\, x\rangle \,\middle|\, Ax = \bar{b}\ ,\ \ x \in \text{co}\{x \in \mathbb{Z}_+^n | \tilde{A}x = \tilde{b}\}\right\} \leq \alpha_{\text{PILP}} \ , \tag{1.10}$$

where co *denotes the convex closure of a set.*

Furthermore, as the function g is concave, we can apply subgradient descent methods to compute $\sup_{\lambda\in\mathbb{R}^m} g(\lambda)$. For a review of the history of Geoffrion Lagrangian relaxation and of the recent advancements to compute its optimal value we refer the reader to [6].

1.2.2 Comparison of Geoffrion Lagrangian relaxation and generalized perturbation-duality scheme

Now, we show that the Geoffrion Lagrangian relaxation is a special case of the perturbation-duality scheme, and that the Geoffrion Lagrangian function is in fact the dual objective function from the perturbation-duality scheme presented in Table 1.1.

Proposition 1.6. *The Geoffrion Lagrangian function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$* (1.9) *coincides with the dual objective function given by the Rockafellian $\mathfrak{R} : \mathbb{Z}_+^n \times \mathbb{Q}^m \to \overline{\mathbb{R}}$ given by*

$$\mathfrak{R}(x, b) = \langle k,\, x\rangle \dot{+} \delta_{Ax=\bar{b}} \dot{+} \delta_{\tilde{A}x=\tilde{b}} \ ,\ \ \forall(x, b) \in \mathbb{Z}_+^n \times \mathbb{Q}^m \ , \tag{1.11}$$

and the coupling $c : \mathbb{Q}^m \times \mathbb{R}^m \to \mathbb{R}$ given by

$$c(b, \lambda) = \langle b,\, \lambda\rangle \ \ ,\ \ \forall(b, \lambda) \in \mathbb{Q}^m \times \mathbb{R}^m \ . \tag{1.12}$$

Proof. We prove that $g(\lambda) = \Phi_{\bar{b}}(\lambda)$, $\forall\lambda \in \mathbb{R}^m$, where the dual objective function $\Phi_{\bar{b}}$ is given by the Rockafellian (1.11) and the coupling (1.12) .

[1] This terminology is confusing as the Lagrangian function already exists in the perturbation-duality scheme. Thus when we refer to Geoffrion's 'Lagrangian function', we say *Geoffrion Lagrangian function*.

We have, for $\lambda \in \mathbb{R}^m$,

$$
\begin{aligned}
\Phi_{\bar{b}}(\lambda) &= c(\bar{b}, \lambda) \dot{+} \big(-\varphi^c(\lambda)\big) \ , && \text{(by 1.6a)}\\
&= \big\langle \bar{b},\ \lambda \big\rangle \dot{+} \Big(- \sup_{b \in \mathbb{Q}^m} \Big\{ \langle b,\ \lambda \rangle \dot{+} \big(- \inf_{x \in \mathbb{Z}^n_+} \{\langle k,\ x\rangle \dot{+} \delta_{Ax=b} \dot{+} \delta_{\tilde{A}x=\tilde{b}}\}\big)\Big\}\Big) \ , \\
& && \text{(by Definition 1.2 and (A.8))}\\
&= \big\langle \bar{b},\ \lambda \big\rangle \dot{+} \Big(- \sup_{b \in \mathbb{Q}^m} \Big\{ \langle b,\ \lambda \rangle \dot{+} \sup_{\substack{\tilde{A}x=\tilde{b}\\ x \in \mathbb{Z}^n_+}} \big\{ - \langle k,\ x\rangle \dot{+} -\delta_{Ax=b}\big\}\Big\}\Big) \ , \\
& && \text{(as } -\inf(\cdot) = \sup(-\cdot) \text{ and } -(\alpha \dot{+} \beta) = (-\alpha) \dot{+} (-\beta))\\
&= \big\langle \bar{b},\ \lambda \big\rangle \dot{+} \Big(- \sup_{\substack{\tilde{A}x=\tilde{b}\\ x \in \mathbb{Z}^n_+}} \Big\{ \sup_{b \in \mathbb{Q}^m} \big\{ \langle b,\ \lambda \rangle - \langle k,\ x\rangle \dot{+} -\delta_{Ax=b}\big\}\Big\}\Big) \ , \\
& && \text{(by switching the two sup)}\\
&= \big\langle \bar{b},\ \lambda \big\rangle \dot{+} \Big(- \sup_{\substack{\tilde{A}x=\tilde{b}\\ x \in \mathbb{Z}^n_+}} \Big\{ \langle Ax,\ \lambda \rangle - \langle k,\ x\rangle \Big\}\Big) \ , \\
&= \inf_{\substack{\tilde{A}x=\tilde{b}\\ x \in \mathbb{Z}^n_+}} \Big\{ \langle k,\ x\rangle + \big\langle \bar{b} - Ax,\ \lambda \big\rangle \Big\} \ , && \text{(as } \dot{+} = + \text{ as all terms take finite values)}\\
&= g(\lambda) \ . && \text{(by the definition of the function } g \text{ (1.9))}
\end{aligned}
$$

□

Thus, the generalized perturbation-duality scheme covers the Geoffrion Lagrangian relaxation, while offering a larger variety of dual elements. Indeed, in the Geoffrion Lagrangian relaxation, the dual elements are the Lagrangian multipliers in $\mathbb{R}^m$, while the perturbation-duality scheme allows for more general dual elements by considering dual spaces 'larger' than $\mathbb{R}^m$ (for instance subadditive functions as we will see in Chapter 2).

It is worth to note that Geoffrion Lagrangian relaxation and the generalized perturbation-duality scheme serve the same purpose: to produce a dual problem that will give a lower estimate of the value of an minimization PILP. The value of this lower estimate should be as tight as possible to the PILP value.

Geoffrion Lagrangian relaxation is used to get a gap tighter than the *continuous relaxation* (which is a special case of Geoffrion Lagrangian relaxation) as we can see by rewriting loosely the inequations (1.10),

$$\textit{continuous relaxation value} \ \leq \ \textit{sup of Geoffrion Lagrangian function} \ \leq \ \textit{PILP value} \ .$$

The *generalized perturbation-duality schemes* goes one step further, *if* the *Rockafellian* $\mathfrak{R}$ and the *coupling* c are chosen such that the scheme they define covers the Geoffrion Lagrangian relaxation *and* includes other dual elements than the linear functions $\{\langle \cdot,\ \lambda\rangle\}_{\lambda \in \mathbb{R}^n}$, then the supremum of the *dual objective function* potentially yields a tighter gap,

$$\textit{sup of Geoffrion Lagrangian function} \ \leq \ \textit{sup of dual objective function} \ \leq \ \textit{PILP value} \ .$$

Chapter 2

Perturbation-duality scheme for pure integer linear programming (PILP)

When solving a mixed integer linear program (MILP), *branch-and-cut* algorithm is the staple method implemented by every MILP solver. Basically, for a minimizing program, branch-and-cut is an enumeration method that takes the form of a research tree keeping track of a lower bound and an upper bound of the optimal value of the MILP. The upper bound is updated when an integral feasible solution is found, while the lower bound is updated at nodes when a surrogate problem minimizing the original problem is solved. As branches of the tree can be pruned if the lower bound computed at a node is greater than the current upper bound, the tightness of the surrogate problems is crucial to speed up the enumeration. Thus, studying dual problems of MILPs comes naturally to design surrogate problems for branch-and-cut, as dual problems are, by essence, lower approximation of the original problem. In §2.1, we introduce the *subadditive dual problem* of a PILP, which the analog of the usual dual of linear programming (LP) but for PILP, and present the result of Blair and Jeroslow [5] restricting the elements of the subadditive dual problem to Chvátal functions. In §2.2, we rewrite these classical results into the framework of the generalized perturbation-duality scheme we presented in Chapter 1. In §2.3, we focus on the linear couplings introduced in §2.2. In §2.4, we discuss the results and branch out with other perturbation-duality scheme ideas.

2.1 The subadditive dual problem of PILP

For an extensive review of integer programming duality, we refer the reader to [17]. In §2.1.1, we present the subadditive dual problem of a PILP where the dual elements are subadditive functions. In §2.1.1, we present the result of Blair and Jeroslow who state that we can restrict the subadditive function space to the Chvátal function space in the subadditive dual problem.

2.1.1 Definition of the subadditive dual problem of PILP

We consider a perturbed PILP in its standard form (that means with equality constraints). For that, we define a perturbation function (or value function) $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$. Let $A \in \mathbb{Q}^{m\times n}$ be a constraint matrix, $k \in \mathbb{Q}^n$ be a cost vector, and consider

$$\forall b \in \mathbb{Q}^m \ , \ \ G_{\mathbb{Z}}(b) = \begin{array}{ll} \inf\limits_{x} & \langle k,\ x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathbb{Z}^n_+ . \end{array} \tag{2.1}$$

We set an anchor $\bar{b} \in \mathbb{Q}^m$ and we call $G_{\mathbb{Z}}(\bar{b})$ the *value of the original PILP*, and the associated PILP the *original PILP*.

We choose rational coefficients for the constraints and for the cost as it is done in the literature [5, 33, 10]. Doing so guarantees convergence of cutting methods such as the Gomory's cutting plane method [33, Theorem 23.2].

In 1979, Jeroslow introduced a subadditive dual problem for MILPs and proved strong duality [19]. We present the PILP version of this subadditive dual problem [5, Theorem 2.15].

Definition 2.1. *For an integer $m \geq 1$, we call*

$$\mathcal{S}^m = \left\{ F \in \overline{\mathbb{R}}^{\mathbb{Q}^m} \middle| F(u_1 + u_2) \leq F(u_1) \dot{+} F(u_2) \ , \ \ \forall u_1, u_2 \in \mathbb{Q}^m \right\} , \tag{2.2}$$

the (rational) subadditive functional space*. Its elements are called* subadditive functions*.*

Theorem 2.2. *[19, Theorem 1], [17, Theorem 3] The following subadditive program*

$$\begin{array}{ll} \sup\limits_{F} & F(\bar{b}) \ , \\ s.t. & F(A_j) \leq k_j \ , \ \ j = 1, \ldots, n \ , \\ & F(0) \leq 0 \ , \\ & A(\mathbb{Z}^n_+) \subset \mathrm{dom} F \ , \\ & F \in \mathcal{S}^m \ , \end{array} \tag{2.3a}$$

where $\mathrm{dom}F$ *is the* effective domain *of the function F defined by $\{b \in \mathbb{Q}^m | \mathcal{S}(b) < +\infty\}$, satisfies*

1. weak duality*: for all feasible $x \in \mathbb{Z}^n$ in the original PILP* (2.1) *and all feasible $F \in \mathcal{S}^m$ in the subadditive dual problem* (2.3a)*, we have the inequality*
$$F(\bar{b}) \leq \langle k,\ x\rangle \ ; \tag{2.3b}$$

2. strong duality*: when the original PILP* (2.1) *is feasible, for all optimal $\hat{x} \in \mathbb{Z}^n$ in the original PILP* (2.1) *and for all optimal $\widehat{F} \in \mathcal{S}^m$ of the subadditive dual problem* (2.3a)*, we have the equality*
$$\widehat{F}(\bar{b}) = \langle k,\ \hat{x}\rangle \ . \tag{2.3c}$$
Furthermore, the value function $G_{\mathbb{Z}}$ in (2.1) *is a feasible solution of the subadditive dual problem* (2.3a).

3. *If the original PILP* (2.1) *(resp., the subadditive dual problem* (2.3a)*) is unbounded, then the subadditive dual problem (resp., the original PILP) is infeasible.*

4. *If the original PILP* (2.1) *(resp., the subadditive dual problem* (2.3a)*) is infeasible, then the subadditive dual problem (resp., the original PILP) is infeasible or unbounded.*

Remark 2.3.

1. *The constraints of the subadditive dual problem* (2.3a) *are similar to the constraints of the usual dual problem of a continuous LP, but the dual variable has been replaced by a dual subadditive function. In a continuous LP, elements of the dual can be identified with linear functions.*

2. *In [19], the subadditive functions were not defined as we did with $\mathcal{S}^m$ in Definition 2.1. In [19], subadditive functions are only defined on $A(\mathbb{Z}^n_+)$ and only take finite values, so there is no need to use Moreau addition for the definition of subadditive functions.*

   However in this report, we decided to allow subadditive functions to be defined on $\mathbb{Q}^m$, so that their definition does not rely on the definitions of a particular PILP. As discussed in Remark 2.10, we want to include the value function $G_{\mathbb{Z}}$ to the dual space, so we have to allow the functions of the dual space to take infinite values as possibly do the value function $G_{\mathbb{Z}}$. Thus, the upper $\dot{+}$ Moreau addition is needed in Definition 2.1 of subadditive functions to sum $-\infty$ and $+\infty$.

Moreover, Jeroslow also obtained complementary slackness conditions in [18, Equation 2.4.C].

Theorem 2.4. *[18, Equation 2.4.C],[17, Theorem 4] Let us denote $a_1, \dots, a_n$the columns of $A = [a_1 \dots a_n] \in \mathbb{Q}^{m\times n}$. Let $x \in \mathbb{Z}^n_+$ and $F \in \mathcal{S}^m$ be feasible points of, respectively, the original PILP* (2.1) *and the subadditive program* (2.3a)*.*

Then the integer vector x and the proper subadditive function F are optimal iff

$$x_j\big(k_j - F(a_j)\big) = 0 \ , \ \ \forall j = 1, \dots, n \ , \tag{2.4a}$$

$$\sum_{j=1}^{n} F(a_j)x_j = F(\bar{b}) \ . \tag{2.4b}$$

2.1.2 Blair's and Jeroslow's result

In their 1982 paper [5], Blair and Jeroslow strenghtened Theorem 2.2 by restricting the subadditive functional space $\mathcal{S}^m$ in (2.3a) to a smaller space of subadditive functions called *Chvátal functions.*

Definition 2.5.

- *The* Gomory function space $\mathcal{G}^m$ *is the intersection of all sets* $E \subset \overline{\mathbb{R}}^{\mathbb{Q}^m}$ *satisfying*

$$\big(u \in \mathbb{Q}^m \mapsto \langle v,\, u\rangle\big) \in E \ , \ \ \forall v \in \mathbb{Q}^m \ , \tag{2.5a}$$
$$\alpha F_1 + \beta F_2 \in E \ , \ \ \forall F_1, F_2 \in E \ , \tag{2.5b}$$
$$\lceil F \rceil \in E \ , \ \ \forall F \in E \ , \tag{2.5c}$$
$$\max\{F_1, F_2\} \in E \ , \ \ \forall F_1, F_2 \in E \ , \tag{2.5d}$$

  where, for any set $\mathcal{W}$ *and any real valued function* $H : \mathcal{W} \to \mathbb{R}$*, the* ceiling function $\lceil \cdot \rceil$ *is defined by*
$$\lceil H \rceil(w) = \inf\big\{k \in \mathbb{Z} \big| H(w) \leq k\big\} \ , \ \ \forall w \in \mathcal{W} \ ; \tag{2.5e}$$

- *the* Chvátal function space $\mathcal{C}^m \subset \mathcal{G}^m$ *is the intersection of all sets* E *satisfying* (2.5a), (2.5b), (2.5c).

Theorem 2.6. *[5, Theorem 5.1, Theorem 5.2] If the original PILP* (2.1) *is feasible, there is a Gomory function which is the optimal solution of the subadditive dual problem*(2.3a) *and which satisfies strong duality.*

Furthermore, there is a Gomory function $F \in \mathcal{G}^m$ *such that the original PILP* (2.1) *is feasible iff* $F(\bar{b}) \leq 0$*.*

Theorem 2.7. *[5, Theorem 5.1, Theorem 5.2, Proposition 2.18] If the original PILP* (2.1) *is feasible, there is a Chvátal function which is the optimal solution of the subadditive dual problem*(2.3a) *and which satisfies strong duality.*

Furthermore, there is a Chvátal function $F \in \mathcal{C}^m$ *such that the original PILP* (2.1) *is feasible iff* $F(\bar{b}) \leq 0$*.*

2.2 Rewriting PILP duality results in the perturbation-duality framework

In §2.2.1, we introduce the perturbation space, Rockafellian and perturbation function which are common to all the five schemes we will consider. In §2.2.2, we present five couplings which are divided into three evaluation couplings and two linear couplings. In §2.2.3, we write the Lagrangians and the dual functions for each scheme (see Table 1.1 for their definitions). In §2.2.4, we present the dual problems resulting from each scheme. In §2.2.5, we present the subdifferentials of the perturbation function for each scheme and their link with complementary slackness. In §2.2.6, we cover the case of canonical PILP (that means PILP with inequality constraints).

2.2.1 Rockafellian and perturbation function

To begin the presentation of the schemes, we introduce a Rockafellian, which will be the same for each of the schemes.

Primal perturbation space and Rockafellian

Throughout this chapter, the primal perturbation space is $\mathcal{U} = \mathbb{Q}^m$ and the Rockafellian $\mathfrak{R} : \mathbb{Z}_+^n \times \mathbb{Q}^m \to \overline{\mathbb{R}}$ is given by

$$\mathfrak{R}(x, b) = \langle k,\ x \rangle + \delta_{\{0\}}(Ax - b) \ , \ \ \forall x \in \mathbb{Z}_+^n \ , \ \ \forall b \in \mathbb{Q}^m \ . \tag{2.6}$$

Let $\bar{b} \in \mathbb{Q}^m$ be its anchor.

Perturbation function (value function)

Following Definition 1.2, the perturbation function $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ of the original PILP (2.1) (or value function in this case as named in [5]) is given by

$$G_{\mathbb{Z}}(b) = \inf_{x \in \mathbb{Z}_+^n} \mathfrak{R}(x, b) = \inf_{x \in \mathbb{Z}_+^n} \left\{ \langle k,\ x \rangle + \delta_{\{0\}}(Ax - b) \right\} \ , \ \ \forall b \in \mathbb{Q}^m \ . \tag{2.7}$$

2.2.2 Three evaluation couplings and two linear couplings

The couplings are divided in two groups: the *(nonlinear) evaluation couplings* and the *linear couplings.*

Evaluation couplings for PILP

We consider three evaluations couplings with the following dual spaces:

- *subadditive functions space* $\mathcal{S}^m$, given by Definition 2.1;
- *Gomory functions space* $\mathcal{G}^m$, given by Definition 2.5 ;
- *Chvátal functions space* $\mathcal{C}^m$, given by Definition 2.5.

The evaluation couplings corresponding to each of these dual spaces are respectively denoted

$$c_{\mathcal{S}} : \mathbb{Q}^m \times \mathcal{S}^m \to \overline{\mathbb{R}} \ , \ \ c_{\mathcal{G}} : \mathbb{Q}^m \times \mathcal{G}^m \to \mathbb{R} \ , \ \ c_{\mathcal{C}} : \mathbb{Q}^m \times \mathcal{C}^m \to \overline{\mathbb{R}} \ , \tag{2.8}$$

and are defined by

$$c_{\mathcal{F}}(b, F) = F(b) \ , \ \ \forall \mathcal{F} \in \{\mathcal{S}^m, \mathcal{G}^m, \mathcal{C}^m\} \ , \ \ \forall b \in \mathbb{Q}^m \ , \ \ \forall F \in \mathcal{F} \ . \tag{2.9}$$

Linear couplings for PILP

We also consider *linear couplings*:

- *real linear coupling* $\star_{\mathbb{R}} : \mathbb{Q}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by

$$\star_{\mathbb{R}}(b, p) = \langle b,\ p \rangle \ , \ \ \forall b \in \mathbb{Q}^m \ , \ \ \forall p \in \mathbb{R}^m \ , \tag{2.10}$$

  here the dual space is $\mathbb{R}^m$;

- *rational linear coupling* $\star_{\mathbb{Q}} : \mathbb{Q}^m \times \mathbb{Q}^m \to \overline{\mathbb{R}}$ defined by

$$\star_{\mathbb{Q}}(b,p) = \langle b,\ p \rangle \ ,\ \ \forall b \in \mathbb{Q}^m \ ,\ \ \forall p \in \mathbb{Q}^m \ , \tag{2.11}$$

  here the dual space is $\mathbb{Q}^m$.

Remark 2.8. *These linear couplings are restrictions of the usual Fenchel coupling* $\star :\ \mathbb{R}^m \times \mathbb{R}^m$*, given by* $\star(b,p) = \langle b,\ p \rangle$, $\forall b \in \mathbb{R}^m, p \in \mathbb{R}^m$*.*

2.2.3 Lagrangians and dual functions

Before defining the dual problems coming for each of these five schemes, let us present the corresponding Lagrangian function and dual functions. The definitions of Lagrangian function and dual function are found in Table 1.1,

Lagrangians

Coupling	Dual space	Lagrangian
Subadditive coupling	$\mathcal{S}^m$	$\mathcal{L} : \mathbb{Z}^n_+ \times \mathcal{S}^m \to \overline{\mathbb{R}}$ $\mathcal{L}(x,F) = \langle k,\ x \rangle - F(Ax)$
Gomory coupling	$\mathcal{G}^m$	$\mathcal{L} : \mathbb{Z}^n_+ \times \mathcal{G}^m \to \mathbb{R}$ $\mathcal{L}(x,F) = \langle k,\ x \rangle - F(Ax)$
Chvátal coupling	$\mathcal{C}^m$	$\mathcal{L} : \mathbb{Z}^n_+ \times \mathcal{C}^m \to \mathbb{R}$ $\mathcal{L}(x,F) = \langle k,\ x \rangle - F(Ax)$
Rational linear coupling	$\mathbb{Q}^m$	$\mathcal{L} : \mathbb{Z}^n_+ \times \mathbb{Q}^m \to \mathbb{R}$ $\mathcal{L}(x,p) = \left\langle k - A^T p,\ x \right\rangle$
Real linear coupling	$\mathbb{R}^m$	$\mathcal{L} : \mathbb{Z}^n_+ \times \mathbb{R}^m \to \mathbb{R}$ $\mathcal{L}(x,p) = \left\langle k - A^T p,\ x \right\rangle$

Table 2.1: Lagrangians for each of the five perturbation-duality schemes

Proof. Let us prove the formulas of the Lagrangians, by identifying $\mathbb{Q}^m$ and $\mathbb{R}^m$ with linear functions spaces.

Let $x \in \mathbb{Z}^n_+, F \in \mathcal{F}$, where $\mathcal{F} \in \{\mathcal{S}^m, \mathcal{G}^m, \mathcal{C}^m, \mathbb{Q}^m, \mathbb{R}^m\}$. Then the Lagrangian is given by

$$\begin{aligned}
\mathcal{L}(x,F) &= \inf_{b \in \mathbb{Q}^m} \left\{ \mathfrak{R}(x,b) \dotplus \left(-c(b,F)\right) \right\} \ , && \text{(according to Table 1.1)} \\
&= \inf_{b \in \mathbb{Q}^m} \left\{ \langle k,\ x \rangle \dotplus \delta_{\{0\}}(Ax - b) \dotplus \left(-F(b)\right) \right\} \ , && \text{(by (2.6) and (2.8))} \\
&= \langle k,\ x \rangle \dotplus \inf_{b \in \mathbb{Q}^m} \left\{ \delta_{\{0\}}(Ax - b) \dotplus \left(-F(b)\right) \right\} \ , \\
&= \langle k,\ x \rangle \dotplus \left(-F(Ax)\right) \ , && (\text{as } Ax \neq b \iff \delta_{\{0\}}(Ax-b) = +\infty) \\
&= \langle k,\ x \rangle - F(Ax) \ .
\end{aligned}$$

In the special case where there is $p \in \mathbb{Q}^m$ or $p \in \mathbb{R}^m$ such that, we get that $F = \langle p, \cdot \rangle$

$$\mathcal{L}(x,p) = \langle k,\ x \rangle - \langle p,\ Ax \rangle = \left\langle k - A^T p,\ x \right\rangle \ .$$

This concludes the proof. □

Dual functions

Coupling	Dual space	Dual function
Subadditive coupling	$\mathcal{S}^m$	$\Psi : \mathcal{S}^m \to \overline{\mathbb{R}}$ $\Psi(F) = \inf_{x \in \mathbb{Z}_+^n} \left\{ \langle k,\ x \rangle - F(Ax) \right\}$
Gomory coupling	$\mathcal{G}^m$	$\Psi : \mathcal{G}^m \to \overline{\mathbb{R}}$ $\Psi(F) = \inf_{x \in \mathbb{Z}_+^n} \left\{ \langle k,\ x \rangle - F(Ax) \right\}$
Chvátal coupling	$\mathcal{C}^m$	$\Psi : \mathcal{C}^m \to \overline{\mathbb{R}}$ $\Psi(F) = \inf_{x \in \mathbb{Z}_+^n} \left\{ \langle k,\ x \rangle - F(Ax) \right\}$
Rational linear coupling	$\mathbb{Q}^m$	$\Psi : \mathbb{Q}^m \to \overline{\mathbb{R}}$ $\Psi(p) = -\delta_{\mathbb{R}_+^n}(k - A^T p)$
Real linear coupling	$\mathbb{R}^m$	$\Psi : \mathbb{R}^m \to \overline{\mathbb{R}}$ $\Psi(p) = -\delta_{\mathbb{R}_+^n}(k - A^T p)$

Table 2.2: Dual functions for each of the five perturbation-duality scheme

Proof.

- The formulas of the dual functions for the evaluation couplings come directly from the definition of a dual function in Table 1.1.

- For $p \in \mathbb{R}^m$, we have

$$\begin{aligned}
\Psi(p) &= \inf_{x \in \mathbb{Z}_+^n} \left\langle k - A^T p,\ x \right\rangle \ , && \text{(by definition of a dual function, see Table 1.1)} \\
&= \inf_{x \in \overline{\mathrm{co}} \mathbb{Z}_+^n} \left\langle k - A^T p,\ x \right\rangle \ , \\
&= \inf_{x \in \mathbb{R}_+^n} \left\langle k - A^T p,\ x \right\rangle \ , \\
&= -\delta_{\mathbb{R}_+^n}(k - A^T p) \ ,
\end{aligned}$$

which concludes the proof.

□

2.2.4 Dual problems

We can now define the dual problems coming from the schemes, using the definition of the *dual objective function* $\Phi_{\bar{b}}(F) = F(\bar{b}) \underset{\cdot}{+} \Psi(F)$ from the Table 1.1. We remind that $\bar{b} \in \mathbb{Q}^m$ is the anchor of the Rockafellian (2.6).

Coupling	Dual space	Dual problem
Subadditive coupling	$\mathcal{S}^m$	Formulation 1: ${G_{\mathbb{Z}}}^{c_{\mathcal{S}} c_{\mathcal{S}}'}(\bar{b}) = \sup_{F \in \mathcal{S}^m} \Big\{ F(\bar{b}) \underset{\cdot}{+} \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) \dot{+} \big(-F(b)\big) \big\} \Big\}$ Formulation 2: ${G_{\mathbb{Z}}}^{c_{\mathcal{S}} c_{\mathcal{S}}'}(\bar{b}) = \sup_{F \in \mathcal{S}^m} \Big\{ F(\bar{b}) \underset{\cdot}{+} \inf_{x \in \mathbb{Z}^n_+} \big\{ \langle k,\ x \rangle - F(Ax) \big\} \Big\}$
Gomory coupling	$\mathcal{G}^m$	Formulation 1: ${G_{\mathbb{Z}}}^{c_{\mathcal{G}} c_{\mathcal{G}}'}(\bar{b}) = \sup_{F \in \mathcal{G}^m} \Big\{ F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - F(b) \big\} \Big\}$ Formulation 2: ${G_{\mathbb{Z}}}^{c_{\mathcal{G}} c_{\mathcal{G}}'}(\bar{b}) = \sup_{F \in \mathcal{G}^m} \Big\{ F(\bar{b}) + \inf_{x \in \mathbb{Z}^n_+} \big\{ \langle k,\ x \rangle - F(Ax) \big\} \Big\}$
Chvátal coupling	$\mathcal{C}^m$	Formulation 1: ${G_{\mathbb{Z}}}^{c_{\mathcal{C}} c_{\mathcal{C}}'}(\bar{b}) = \sup_{F \in \mathcal{C}^m} \Big\{ F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - F(b) \big\} \Big\}$ Formulation 2: ${G_{\mathbb{Z}}}^{c_{\mathcal{C}} c_{\mathcal{C}}'}(\bar{b}) = \sup_{F \in \mathcal{C}^m} \Big\{ F(\bar{b}) + \inf_{x \in \mathbb{Z}^n_+} \big\{ \langle k,\ x \rangle - F(Ax) \big\} \Big\}$
Rational linear coupling	$\mathbb{Q}^m$	Formulation 1: ${G_{\mathbb{Z}}}^{\star_{\mathbb{Q}} \star_{\mathbb{Q}}'}(\bar{b}) = \sup_{p \in \mathbb{Q}^m} \Big\{ F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - \langle b,\ p \rangle \big\} \Big\}$ Formulation 2: ${G_{\mathbb{Z}}}^{\star_{\mathbb{Q}} \star_{\mathbb{Q}}'}(\bar{b}) = \sup\limits_{\substack{A^T p \leq k \\ p \in \mathbb{Q}^m}} \big\langle \bar{b},\ p \big\rangle$
Real linear coupling	$\mathbb{R}^m$	Formulation 1: ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}} \star_{\mathbb{R}}'}(\bar{b}) = \sup_{p \in \mathbb{R}^m} \Big\{ F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - \langle b,\ p \rangle \big\} \Big\}$ Formulation 2: ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}} \star_{\mathbb{R}}'}(\bar{b}) = \sup\limits_{\substack{A^T p \leq k \\ p \in \mathbb{R}^m}} \big\langle \bar{b},\ p \big\rangle$

Table 2.3: Dual problems for each of the five perturbation-duality scheme

Proposition 2.9. *The perturbation function $G_{\mathbb{Z}}$ satisfies the following strong duality results:*

1. ${G_{\mathbb{Z}}}^{c_{\mathcal{S}}c_{\mathcal{S}}'} = G_{\mathbb{Z}}$*, which means that the function $G_{\mathbb{Z}}$ is $c_{\mathcal{S}}$-*convex*; furthermore, there exists a subadditive function $F \in \mathcal{S}^m$ such that $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b})$, for all vector $\bar{b} \in \mathbb{Q}^m$;*

2. *if $G_{\mathbb{Z}}(0) > -\infty$, then ${G_{\mathbb{Z}}}^{c_{\mathcal{G}}c_{\mathcal{G}}'}(\bar{b}) = G_{\mathbb{Z}}(\bar{b})$, for all vector $\bar{b} \in \mathrm{dom}\ G_{\mathbb{Z}}$, which means that the function $G_{\mathbb{Z}}$ is $c_{\mathcal{G}}$-*convex on $\mathrm{dom} G_{\mathbb{Z}}$*; furthermore, there exists a Gomory function $H \in \mathcal{G}^m$ such that $G_{\mathbb{Z}}(\bar{b}) = H(\bar{b})$, for all vector $\bar{b} \in \mathrm{dom}\ G_{\mathbb{Z}}$;*

3. *if $G_{\mathbb{Z}}(0) > -\infty$, then ${G_{\mathbb{Z}}}^{c_{\mathcal{C}}c_{\mathcal{C}}'}(\bar{b}) = G_{\mathbb{Z}}(\bar{b})$, for all vector $\bar{b} \in \mathrm{dom}\ G_{\mathbb{Z}}$, which means that the function $G_{\mathbb{Z}}$ is $c_{\mathcal{C}}$-*convex on $\mathrm{dom} G_{\mathbb{Z}}$*; furthermore, for all vector $\bar{b} \in \mathrm{dom}\ G_{\mathbb{Z}}$, there exists a Chvátal function $H_{\bar{b}} \in \mathcal{C}^m$ such that $G_{\mathbb{Z}}(\bar{b}) = H_{\bar{b}}(\bar{b})$.*

Proof. 1. We prove ${G_{\mathbb{Z}}}^{c_{\mathcal{S}}c_{\mathcal{S}}'} = G_{\mathbb{Z}}$. As the inequality ${G_{\mathbb{Z}}}^{c_{\mathcal{S}}c_{\mathcal{S}}'} \leq G_{\mathbb{Z}}$ is true according to Proposition A.5, we have to prove ${G_{\mathbb{Z}}}^{c_{\mathcal{S}}c_{\mathcal{S}}'} \geq G_{\mathbb{Z}}$.

Let $\bar{b} \in \mathbb{Q}^m$. We have that

$$\begin{aligned} {G_{\mathbb{Z}}}^{c_{\mathcal{S}}c_{\mathcal{S}}'}(\bar{b}) = & \sup_{F \in \mathcal{S}^m} \Big\{ F(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) \dot{+} \big(-F(b)\big) \big\} \Big\} , \\ & \text{(according to Formulation 1 in Line 1 in Table 2.3)} \\ \geq & G_{\mathbb{Z}}(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) \dot{+} \big(-G_{\mathbb{Z}}(b)\big) \big\} , \\ & \text{(as it is easy to check that the value function } G_{\mathbb{Z}} \in \mathcal{S} \text{ is subadditive)} \\ \geq & G_{\mathbb{Z}}(\bar{b}) . \qquad \text{(as } \inf\big\{ G_{\mathbb{Z}}(b) \dot{+} \big(-G_{\mathbb{Z}}(b)\big) \big\} \geq 0 \text{, according to (A.5d))} \end{aligned}$$

Thus, we have proven 1.

2. Similarly to 1, we just have to prove that ${G_{\mathbb{Z}}}^{c_{\mathcal{G}}c_{\mathcal{G}}'}(\bar{b}) \geq G_{\mathbb{Z}}(\bar{b})$, for all vector $\bar{b} \in \mathrm{dom} G_{\mathbb{Z}} \subset \mathbb{Q}^m$.

   Let $\bar{b} \in \mathrm{dom} G_{\mathbb{Z}}$. Let $H \in \mathcal{G}^m \subset \mathbb{R}^{\mathbb{Q}^m}$ be the Gomory function that coincides with the value function $G_{\mathbb{Z}}$ on its domain $\mathrm{dom} G_{\mathbb{Z}}$ [5, Theorem 5.2]. We have that,

$$\begin{aligned} {G_{\mathbb{Z}}}^{c_{\mathcal{G}}c_{\mathcal{G}}'}(\bar{b}) = & \sup_{F \in \mathcal{G}^m} \Big\{ F(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - F(b) \big\} \Big\} , \\ & \text{(according to Formulation 1 in Line 2 in Table 2.3)} \\ \geq & H(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - H(b) \big\} , \qquad \text{(as } H \in \mathcal{G}^m) \\ \geq & H(\bar{b}) , \qquad \text{(as } G_{\mathbb{Z}}(b) = H(b) \text{, for all } b \in \mathrm{dom} G_{\mathbb{Z}}) \\ \geq & G_{\mathbb{Z}}(\bar{b}) . \qquad \text{(as } \bar{b} \in \mathrm{dom} G_{\mathbb{Z}} \implies G_{\mathbb{Z}}(\bar{b}) = H(\bar{b})) \end{aligned}$$

   Thus, we have proven 2.

3. Following the same arguments, we have to prove that ${G_{\mathbb{Z}}}^{ccc c'}(\bar{b}) \geq G_{\mathbb{Z}}(\bar{b})$, for all vector $\bar{b} \in \mathrm{dom} G_{\mathbb{Z}} \subset \mathbb{Q}^m$.

   Let $H \in \mathcal{G}^m \subset \mathbb{R}^{\mathbb{Q}^m}$ be the Gomory function that coincides with the value function $G_{\mathbb{Z}}$ on its domain $\mathrm{dom} G_{\mathbb{Z}}$.

   Let $\bar{b} \in \mathrm{dom} G_{\mathbb{Z}}$. According to [5, Proposition 2.18], there exists a Chvátal function $H_{\bar{b}} \in \mathcal{C}^m$ such that $H(\bar{b}) = H_{\bar{b}}(\bar{b})$. We have that

$$
\begin{aligned}
{G_{\mathbb{Z}}}^{ccc c'}(\bar{b}) = \sup_{F \in \mathcal{C}^m} & \left\{ F(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \left\{ G_{\mathbb{Z}}(b) - F(b) \right\} \right\} , \\
& \qquad \text{(according to Formulation 1 in Line 23in Table 2.3)} \\
\geq & H_{\bar{b}}(\bar{b}) \dot{+} \inf_{b \in \mathbb{Q}^m} \left\{ G_{\mathbb{Z}}(b) - H_{\bar{b}}(b) \right\} , \qquad \text{(as } H_{\bar{b}} \in \mathcal{C}^m \text{)} \\
\geq & H_{\bar{b}}(\bar{b}) , \qquad \text{(as } H_{\bar{b}}(\bar{b}) = H(\bar{b}) = G_{\mathbb{Z}}(\bar{b}) \text{)} \\
\geq & G_{\mathbb{Z}}(\bar{b}) . \qquad \text{(as } H_{\bar{b}}(\bar{b}) = H(\bar{b}) = G_{\mathbb{Z}}(\bar{b}) \text{)}
\end{aligned}
$$

   Thus we have proven 3.

□

Remark 2.10.

1. *In the proof of Proposition 2.9, the strong duality essentially comes from the following property: the perturbation function (here $G_{\mathbb{Z}}$) coincides with a function in the dual space (here $\mathcal{S}, \mathcal{G}$ or $\mathcal{C}$) on its domain. Thus, in order to get strong duality with an evaluation coupling, we can distinguish two extreme cases for the choice of the dual space $\mathcal{F}$: we could take $\mathcal{F} = \{G_{\mathbb{Z}}\}$ containing only the perturbation function, or we could take $\mathcal{F} = \overline{\mathbb{R}}^{\mathbb{R}^m}$ containing every function from $\mathbb{R}^m$ to $\overline{\mathbb{R}}$. In the first case, the dual space is too 'small' as we do not usually know the perturbation function. In the second case, the dual space is too 'big' as every perturbation function in every point would be c-convex.*

2. *Looking at the formulation 2 in the first line of Table 2.3, we see that we implicitly retrieved the constraint $A(\mathbb{Z}^n_+) \subset \mathrm{dom} F$ from Jeroslow's subadditive dual problem* (2.3a)*. Indeed, if $Ax \notin \mathrm{dom} F$, then $\inf_{x \in \mathbb{Z}^n_+} \left\{ \langle k,\, x \rangle - F(Ax) \right\} = -\infty$. However, we still do not have the constraints $F(A_j) \leq k_j$, $F(0) \leq 0$ from Jeroslow's subadditive dual problem* (2.3a)*. For them, we need to look at the $c_{\mathcal{S}}$-subdifferential of the perturbation function $G_{\mathbb{Z}}$ in §2.2.5.*

2.2.5 Generalized subdifferentials of the perturbation function

To complete the study of PILP through the perturbation-duality scheme, we address the link between the complementary slackness conditions and the subdifferential of the perturbation function. More generally speaking, the subdifferential of the perturbation function is linked to the *abstract Karush-Kuhn-Tucker conditions*(see [30, Section 7, Section 10]), when the optimization space is paired to another space by a coupling (here the optimization variable is $x \in \mathbb{Z}^n$ and $\mathbb{Z}^n$ is paired with $\mathbb{Q}^n$ by the scalar product). $\langle k, \, x \rangle$).

Coupling	Subdifferential Characterization (1)	Subdifferential Characterization (2)
$c_{\mathcal{S}}$	$F \in \partial^{c_{\mathcal{S}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{S}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) \underset{\cdot}{+} \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) \dotplus \big(-F(b)\big) \big\}$	$F \in \partial^{c_{\mathcal{S}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{S}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) \underset{\cdot}{+} \inf\limits_{x \in \mathbb{Z}^n_+} \big\{ \langle k, \, x \rangle - F(Ax) \big\}$
$c_{\mathcal{G}}$	$F \in \partial^{c_{\mathcal{G}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{G}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - F(b) \big\}$	$F \in \partial^{c_{\mathcal{G}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{G}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) + \inf\limits_{x \in \mathbb{Z}^n_+} \big\{ \langle k, \, x \rangle - F(Ax) \big\}$
$c_{\mathcal{C}}$	$F \in \partial^{c_{\mathcal{C}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{C}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) + \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - F(b) \big\}$	$F \in \partial^{c_{\mathcal{C}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathcal{C}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) + \inf\limits_{x \in \mathbb{Z}^n_+} \big\{ \langle k, \, x \rangle - F(Ax) \big\}$
$\star_{\mathbb{Q}}$	$p \in \partial^{\star_{\mathbb{Q}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathbb{Q}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = \big\langle \bar{b}, \, p \big\rangle \underset{\cdot}{+} \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - \langle p, \, b \rangle \big\}$	$p \in \partial^{\star_{\mathbb{Q}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathbb{Q}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = \big\langle \bar{b}, \, p \big\rangle + \delta_{\mathbb{R}^n_+}(k - A^T p)$
$\star_{\mathbb{R}}$	$p \in \partial^{\star_{\mathbb{R}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathbb{R}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = \big\langle \bar{b}, \, p \big\rangle \underset{\cdot}{+} \inf\limits_{b \in \mathbb{Q}^m} \big\{ G_{\mathbb{Z}}(b) - \langle p, \, b \rangle \big\}$	$p \in \partial^{\star_{\mathbb{R}}} G_{\mathbb{Z}}(\bar{b}) \subset \mathbb{R}^m \iff$ $G_{\mathbb{Z}}(\bar{b}) = \big\langle \bar{b}, \, p \big\rangle + \delta_{\mathbb{R}^n_+}(k - A^T p)$

Table 2.4: Subdifferential characterization for each of the five perturbation-duality scheme

Proposition 2.11. *Let $G_{\mathbb{Z}}$ be the perturbation function of the PILP* (2.1) *defined by $k \in \mathbb{Q}^n$, $A = \big(A_j\big)_{j=1,\ldots,n} \in \mathbb{Q}^{m \times n}$, and $\bar{b} \in \mathbb{Q}^n$. Let $\hat{x} \in \{x \in \mathbb{Z}^n_+ | Ax = \bar{b}\}$ and $\widehat{F} \in \mathcal{G}^m$. Suppose that $G_{\mathbb{Z}}(\bar{b}) = \langle k, \, \hat{x} \rangle$. Then the following assertions are equivalent:*

$$\widehat{F} \in \partial^{c_{\mathcal{G}}} G_{\mathbb{Z}}(\bar{b}) \; , \tag{2.12a}$$

$$-k \in \partial\big(-\widehat{F} \circ A \dotplus \delta_{\mathbb{Z}^n_+}\big)(\hat{x}) \; . \tag{2.12b}$$

Furthermore, if $\widehat{F}(A_j) \leq k_j$, for all $\forall j = 1, \ldots, n$, then the following assertion is equivalent to (2.12b)*:*

$$\widehat{F}(0) = 0 \; , \; \; \widehat{F}(\bar{b}) = G_{\mathbb{Z}}(\bar{b}) \text{ and } \big(k_j - \widehat{F}(A_j)\big)\hat{x}_j = 0 \; , \; \; \forall j = 1, \ldots, n \; . \tag{2.12c}$$

Proof.

- Let $\hat{x} \in \{x \in \mathbb{Z}^n_+ | Ax = \bar{b}\}$ and $F \in \mathcal{G}^m$. Suppose that $G_{\mathbb{Z}}(\bar{b}) = \langle k, \, \hat{x} \rangle$ Then, we have that

$$
\begin{aligned}
&\widehat{F} \in \partial^{cg} G_{\mathbb{Z}}(\bar{b}) \ , \\
\iff & G_{\mathbb{Z}}(\bar{b}) \leq \widehat{F}(\bar{b}) + \langle k,\ x\rangle - \widehat{F}(Ax) \ , \ \ \forall x \in \mathbb{Z}_+^n \ , \qquad \text{(according to Table 2.4)} \\
\iff & \langle k,\ \hat{x}\rangle \leq \widehat{F}(A\hat{x}) + \langle k,\ x\rangle - \widehat{F}(Ax) \ , \ \ \forall x \in \mathbb{Z}_+^n \ , \qquad (\text{as } G_{\mathbb{Z}}(\bar{b}) = \langle k,\ \hat{x}\rangle) \\
\iff & \langle -k,\ x-\hat{x}\rangle - \widehat{F}(A\hat{x}) \leq -\widehat{F}(Ax) \ , \ \ \forall x \in \mathbb{Z}_+^n \ , \\
\iff & \langle -k,\ x-\hat{x}\rangle - \widehat{F}(Ax) \dotplus \delta_{\mathbb{Z}_+^n}(x) \leq -\widehat{F}(A\hat{x}) \dotplus \delta_{\mathbb{Z}_+^n}(x) \ , \ \ \forall x \in \mathbb{R}^n \ , \\
\iff & -k \in \partial\big(-\widehat{F} \circ A \dotplus \delta_{\mathbb{Z}_+^n}\big)(\hat{x}) \ .
\end{aligned}
$$

- We prove $\big(\widehat{F}(A_j) \leq k_j, \forall j \in= 1, \ldots, n\big) \implies \big((2.12\text{a}) \iff (2.12\text{c})\big)$
  - It is easy to check that $\big(\widehat{F}(A_j) \leq k_j, \forall j \in= 1, \ldots, n\big) \implies \big((2.12\text{c}) \implies (2.12\text{a})\big)$
  - Now we prove the reciprocal implication. Let us notice that

$$0 \geq G_{\mathbb{Z}}(\bar{b}) - \widehat{F}(\bar{b}) = \langle k,\ \hat{x}\rangle - \widehat{F}(A\hat{x}) \ .$$

Indeed, we have that

$$
\begin{aligned}
& \langle k,\ \hat{x}\rangle - \widehat{F}(A\hat{x}) \leq \langle k,\ x\rangle - \widehat{F}(Ax) \ , \ \ \forall x \in \mathbb{Z}_+^n \ , \\
& \qquad\qquad (\text{as } \widehat{F} \in \partial^{cg} G_{\mathbb{Z}}(\bar{b}) \text{ and } G_{\mathbb{Z}}(\bar{b}) = \langle k,\ \hat{x}\rangle) \\
\implies & \langle k,\ \hat{x}\rangle - \widehat{F}(A\hat{x}) \leq -\widehat{F}(0) \ , \\
\implies & \langle k,\ \hat{x}\rangle - \widehat{F}(A\hat{x}) \leq 0 \ .
\end{aligned}
$$

(as $\widehat{F}(0) \geq \frac{F(N0)}{N} \xrightarrow[N\to\infty]{} 0$, where the limit comes from Fekete's lemma [21],[12, Proposition IX])

  - Suppose $\widehat{F}(A_j) \leq k_j, \forall j \in= 1, \ldots, n$. We have that

$$\sum_{i=1}^{n} \big(k_i - \widehat{F}(A_i)\big)x_i \geq 0 \ . \qquad (\text{by the assumption} \widehat{F}(A_j) \leq k_j, \forall j \in= 1, \ldots, n)$$

Furthermore,

$$\widehat{F}(A\hat{x}) \geq \sum_{i=1}^{n} \big(k_i - \widehat{F}(A_i)\big)x_i \ . \qquad (\text{by subadditivity of } F)$$

Thus

$$0 \geq G_{\mathbb{Z}}(\bar{b}) - \widehat{F}(\bar{b}) = \langle k,\ \hat{x}\rangle - \widehat{F}(A\hat{x}) \geq 0 \ ,$$

which concludes the proof.

Suppose now that $\widehat{F}(A_j) \leq k_j$. $\square$

Remark 2.12. *Thus, we have completed the rewriting of Jeroslow's subadditive dual problem* (2.3a) *and we reinterpret it as follows: in Jeroslow's subadditive dual* (2.3a) *the subadditive space is restricted to the subadditive minimizers F of the PILP value function $G_{\mathbb{Z}}$ such that F is a $c_{\mathcal{S}}$ upper-subgradient of $G_{\mathbb{Z}}$ and coincides with $G_{\mathbb{Z}}$ at the anchor $\bar{b}$.*

2.2.6 Case of canonical PILP

Until now, we have considered PILP in its standard form, that means with equality constraints. We could have considered PILP in canonical form, that means with inequatlity constraints:

$$\begin{array}{ll} \inf\limits_{x} & \langle k,\ x\rangle\ , \\ \text{s.t.} & Ax \geq \bar{b}, \\ & x \geq 0, \\ & x \in \mathbb{Z}^n. \end{array} \tag{2.13}$$

Thankfully, the results we have established remain true for the natural corresponding Rockafellian

$$\mathfrak{R}_{\geq}(x, b) = \langle k,\ x\rangle + \delta_{\mathbb{R}^n_+}(b - Ax) + \delta_{\mathbb{Z}^n_+}(x)\ ,\ \ \forall b \in \mathbb{Q}^m\ ,\ \ \forall x \in \mathbb{R}^n\ , \tag{2.14}$$

by slightly changing the dual spaces according to [5, Theorem 5.17].

To do so, let us introduce for a real valued functional set $\mathcal{F} \subset \overline{\mathbb{R}}^{\mathbb{R}^m}$, the set of the nondecreasing functions $\mathcal{F}_{\nearrow} \subset \mathcal{F}$, according to the usual order on $\mathbb{R}^m$. Furthermore, we identify $\mathbb{Q}^m, \mathbb{R}^m$ respectively with the rational linear functions and the real linear functions.

Perturbation-duality	Standard ($=$)	Canonical ($\geq$)
Subadditive	$\mathcal{S}^m$	$\mathcal{S}^m_{\nearrow}$
Gomory	$\mathcal{G}^m$	$\mathcal{G}^m_{\nearrow}$
Chvátal	$\mathcal{C}^m$	$\mathcal{C}^m_{\nearrow}$
Rational linear	$\mathbb{Q}^m$	$\mathbb{Q}^m_{\nearrow} = \mathbb{Q}^m_+$
Real linear	$\mathbb{R}^m$	$\mathbb{R}^m_{\nearrow} = \mathbb{R}^m_+$

Table 2.5: Comparison between standard PILP dual spaces and canonical PILP dual spaces

2.3 Linear coupling convexity of the integer value function

While Proposition 2.9 states that we have $c_{\mathcal{S}}$-, $c_{\mathcal{G}}$- and $c_{\mathcal{C}}$-strong duality between the corresponding dual problem and the original PILP under weak assumptions, nothing is said about $\star_{\mathbb{Q}}$- and $\star_{\mathbb{R}}$-strong duality. In this section, we discuss why $\star_{\mathbb{Q}}$- and $\star_{\mathbb{R}}$- convexity for the value function $G_{\mathbb{Z}}$ are generally not achieved.

To do so, let us first define the *(continuously) relaxed value function* $G_{\mathbb{R}} : \mathbb{R}^m \to \overline{\mathbb{R}}$, given by

$$G_{\mathbb{R}}(b) = \inf_{x \in \mathbb{R}^n} \left\{ \langle k,\ x \rangle + \delta_{\{0\}}(Ax - b) \right\} ,\ \ \forall b \in \mathbb{R}^m \ . \tag{2.15}$$

As indicated by its name, it is the value function of the usual continuous relaxation of the PILP (2.1).

In §2.3.1, we show that the value of dual problems coming from the $\star_{\mathbb{Q}}$- and $\star_{\mathbb{R}}$ couplings are the same and coincide with the value of the continuous relaxation. In §2.3.2, we show that there is no better convex lower approximation of the PILP value function $G_{\mathbb{Z}}$ than the value function of the continuous relaxation.

2.3.1 The same dual problem for the linear couplings

Looking at the last two lines of Table 2.3 could give the impression that $\star_{\mathbb{Q}}$ and $\star_{\mathbb{R}}$ couplings give two different dual problems when used in the perturbation-duality scheme. It is actually not the case.

Proposition 2.13. *We remind that $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ is the value function of the PILP defined in* (2.1)*. The $\star_{\mathbb{Q}}$-biconjugate ${G_{\mathbb{Z}}}^{\star_{\mathbb{Q}}\star_{\mathbb{Q}}{}'} : \mathbb{Q}^m \to \overline{\mathbb{R}}$, the $\star_{\mathbb{R}}$-biconjugate ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}{}'} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ satisfy*

$${G_{\mathbb{Z}}}^{\star_{\mathbb{Q}}\star_{\mathbb{Q}}{}'}(b) = {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}{}'}(b) = G_{\mathbb{R}}(b) \ ,\ \ \forall b \in \mathrm{dom}\ G_{\mathbb{Z}} \ . \tag{2.16}$$

Proof. As

$$\overline{\mathrm{co}}\big(\mathbb{Q}^m \cap \{p \in \mathbb{R}^m | A^T p \leq k\}\big) = \{p \in \mathbb{R}^m | A^T p \leq k\} \ ,$$

properties of support functions give:

$${G_{\mathbb{Z}}}^{\star_{\mathbb{Q}}\star_{\mathbb{Q}}{}'}(b) = \sup_{\substack{A^T p \leq k \\ p \in \mathbb{Q}^m}} \langle p,\ b \rangle = \sup_{\substack{A^T p \leq k \\ p \in \mathbb{R}^m}} \langle p,\ b \rangle = {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}{}'}(b) \ .$$

The last inequality follows from the usual strong duality in continuous linear programming. □

Thus, there no hope for strong duality with the linear couplings as it is well-known that the value $G_{\mathbb{R}}(\bar{b})$ of the relaxed problem is usually smaller than the value $G_{\mathbb{Z}}(\bar{b})$ of the PILP.

Example 2.14. *[5, Equation (3.20)]*

For example, let us consider the following PILP:

$$\begin{array}{rll} G_{\mathbb{Z}}(b) = & \inf\limits_{x} & x_1 - x_2 \\ & s.t. & x_1 - x_2 - x_3 = b \quad (\forall b \in \mathbb{Q}) \\ & & x \in \mathbb{Z}^3_+ . \end{array} \tag{2.17}$$

We can easily see that the value function is $G_{\mathbb{Z}}(b) = \lceil b \rceil$, $\forall b \in \mathbb{Q}$, and that the relaxed value function is $G_{\mathbb{R}}(b) = b$, $\forall b \in \mathbb{Q}$.

Here, $G_{\mathbb{R}}(b) = G_{\mathbb{Z}}(b) \iff b \in \mathbb{Z}$, thus $G_{\mathbb{R}}(b) < G_{\mathbb{Z}}(b)$, $\forall b \in \mathbb{Q} \setminus \mathbb{Z}$. We see in this example that strong duality for linear couplings is not achieved.

2.3.2 Best convex approximation of the PILP value function $G_{\mathbb{Z}}$

It is well-known that the relaxed value function $G_{\mathbb{R}}$ is convex and is a minimizer of the PILP value function $G_{\mathbb{Z}}$ in the sense $G_{\mathbb{R}|\mathbb{Q}^m} \leq G_{\mathbb{Z}}$ [30, Section 5]. Then we may ask: is there a better convex lower approximation of $G_{\mathbb{Z}}$? Jeroslow answered no [19, Theorem 2]. Here, we present another proof of it using conjugacy.

First, we define the *best convex lower approximation* of a function defined on some set $\mathcal{W} \subset \mathbb{R}^m$.

Definition 2.15. *Let $\mathcal{W} \subset \mathbb{R}^m$ be a subset and $g : \mathcal{W} \to \overline{\mathbb{R}}$ be a function. We call the function* $\mathrm{co}_{\mathcal{W}} g$ *the* best convex lower approximation of g on $\mathcal{W}$ *defined by*

$$\mathrm{co}_{\mathcal{W}} g = \sup_{\substack{h|_{\mathcal{W}} \leq g \\ h \text{ proper convex lsc}}} h \ . \tag{2.18}$$

Proposition 2.16. *Let* $\mathrm{co}_{\mathbb{Q}^m} G_{\mathbb{Z}} : \mathbb{R}^m \to \overline{\mathbb{R}}$ *be* the best convex lower approximation *of the value function $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ as defined in* (2.18)*, and $G_{\mathbb{R}} : \mathbb{R}^m \to \overline{\mathbb{R}}$ be the relaxed value function defined in* (2.15)*. Then*

$$\big(\mathrm{co}_{\mathbb{Q}^m} G_{\mathbb{Z}}\big)|_{\mathbb{Q}^m} = G_{\mathbb{R}}|_{\mathbb{Q}^m} \ . \tag{2.19}$$

Proof. We have that

1. for any function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ coinciding with the value function $G_{\mathbb{Z}}$ in $\mathbb{Q}^m$, $\big(f \dotplus \delta_{\mathbb{Q}^m}\big)^\star = {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}}$, because, for any $p \in \mathbb{R}^m$, we have that

$$\begin{aligned}
\big(f \dotplus \delta_{\mathbb{Q}^m}\big)^\star(p) &= \sup_{b \in \mathbb{R}^m} \langle b,\, p\rangle \stackrel{\cdot}{+} \big(-f(b) \stackrel{\cdot}{+} (-\delta_{\mathbb{Q}^m}(b))\big) \ , \\
&= \sup_{b \in \mathbb{Q}^m} \langle b,\, p\rangle \stackrel{\cdot}{+} \big(-G_{\mathbb{Z}}(b)\big) \ , \\
&= {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}}(p) \ ,
\end{aligned}$$

2. ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}'}|_{\mathbb{Q}^m} = {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star}|_{\mathbb{Q}^m}$, because for all $b \in \mathbb{Q}^m$,

$$\begin{aligned}
{G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}'}(b) &= \sup_{p \in \mathbb{R}^m} \{\langle b,\, p\rangle \stackrel{\cdot}{+} (-{G_{\mathbb{Z}}}^{\star_{\mathbb{R}}}(p))\} \ , \\
&= {G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star}(b) \ . \qquad \text{(by definition of the Fenchel conjugate)}
\end{aligned}$$

3. ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}'}|_{\mathbb{Q}^m} = G_{\mathbb{R}}|_{\mathbb{Q}^m}$ according to Proposition 2.13.

Combining all of these results, we get that $\left(f \dot{+} \delta_{\mathbb{Q}^m}\right)^{\star\star}|_{\mathbb{Q}^m} = G_{\mathbb{R}}|_{\mathbb{Q}^m}$, for any function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ coinciding with the value function $G_{\mathbb{Z}}$ in $\mathbb{Q}^m$. Finally, Lemma 2.17 allow us to conclude that $\left(\mathrm{co}_{\mathbb{Q}^m} G_{\mathbb{Z}}\right)|_{\mathbb{Q}^m} = G_{\mathbb{R}}|_{\mathbb{Q}^m}$. $\square$

Lemma 2.17. *Let $\mathcal{W} \subset \mathbb{R}^m$ and $g : \mathcal{W} \to \overline{\mathbb{R}}$. Let $\mathrm{co}_{\mathcal{W}} g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be the* best convex lower approximation *of g defined in* (2.18)*. Then, we have that*

$$\mathrm{co}_{\mathcal{W}} g = \left(f \dot{+} \delta_{\mathcal{W}}\right)^{\star\star} , \tag{2.20}$$

for any function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ coinciding with the function g in $\mathcal{W}$.

Proof. Let $\mathcal{A} \subset \mathbb{R}^{\mathbb{R}^m}$ be the set of affine functions, and $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ a function coinciding with the function g in $\mathcal{W}$.

$$\begin{aligned}
\mathrm{co}_{\mathcal{W}} g &= \sup_{\substack{h|_{\mathcal{W}} \leq g \\ h \text{ proper convex lsc}}} h , \\
&= \sup_{\substack{h|_{\mathcal{W}} \leq g \\ h \text{ proper convex lsc}}} \sup_{\substack{p \in \mathcal{A} \\ p \leq h}} p , \\
&\quad \text{(as } h \text{ is proper convex lsc, it is the sup of its affine minorants [2, Corollary 13.42])} \\
&= \sup_{p \in \mathcal{A}} \sup_{\substack{h|_{\mathcal{W}} \leq g \\ h \text{ proper convex lsc} \\ p \leq h}} p , \\
&= \sup_{p \in \mathcal{A}} \begin{cases} -\infty , & \text{if } p|_{\mathcal{W}} \not\leq g \\ p , & \text{otherwise} \end{cases} , \\
&\quad \text{(as } h|_{\mathcal{W}} \leq g \text{ and } p \leq h \implies p|_{\mathcal{W}} \leq g \text{ and as } p \text{ is affine, thus proper convex lsc)} \\
&= \sup_{p \in \mathcal{A}} \left\{ p \dotplus \left(-\delta_{p|_{\mathcal{W}} \leq g}\right) \right\} , \\
&= \sup_{\substack{p \in \mathcal{A} \\ p|_{\mathcal{W}} \leq g}} p , \\
&= \sup_{\substack{p \in \mathcal{A} \\ p \leq f \dot{+} \delta_{\mathcal{W}}}} p , \qquad \text{(as } p|_{\mathcal{W}} \leq g \iff p \leq f \dot{+} \delta_{\mathcal{W}} \text{ because } f \text{ coincides with } g \text{ in } \mathcal{W}) \\
&= \left(f \dot{+} \delta_{\mathcal{W}}\right)^{\star\star} .
\end{aligned}$$

(because the biconjugate is the best affine minorant approximation at each point [2, Corollary 13.42])

$\square$

All in all, we cannot do better than the function $G_{\mathbb{R}}$ in the proper convex lsc world to approximate the value function $G_{\mathbb{Z}}$ from below. It can be seen in the fact that $G_{\mathbb{Z}}$ coincides with a Gomory function on its effective domain, and as there are integer round-up operations in its expression, it makes $G_{\mathbb{Z}}$ non-convex.

2.4 Conclusion and discussion

PILP duality and the perturbation-duality scheme

On one hand, Jeroslow provided a subadditive dual problems in [19] for PILP, where the dual variables belong to the space of subadditive functions. He and Blair proved in [5] that the space of subadditive functions could be restricted to the space of Gomory functions and even to the space of Chvátal functions. On the other hand, the perturbation-duality scheme of Rockafellar presented in [30] gives a systematic method to construct a dual problems from an original minimization problem. The scheme highlights the two choices behind the construction of a dual problem: the choice of a perturbation of the original minimization problem and the choice of the coupling between the space of perturbation and a dual space.

Our contribution was to rewrite the results of Jeroslow using the perturbation-duality of Rockafellar with the intent to see what the systematic point of of view of the perturbation-duality scheme could bring to the understanding of the duality of a PILP. We realized that the Jeroslow's subadditive dual problem corresponds to the problem of finding a $c_{\mathcal{S}}$-subdifferential $F \in \mathcal{S}$ of the value function $G_{\mathbb{Z}}$ at the anchor $\bar{b} \in \mathbb{Q}^m$ which coincides with the value function $G_{\mathbb{Z}}$ at the anchor $\bar{b}$.

We sum up the strong duality results that we have rewritten in the perturbation duality scheme in the following proposition and table.

Proposition 2.18. *If the original PILP* (2.1) *is bounded from below (which is equivalent to $G_{\mathbb{Z}}(0) > -\infty$), then the biconjugates of the value function $G_{\mathbb{Z}}$ for each scheme satisfy the following equalities and inequalities*

$$G_{\mathbb{R}}(\bar{b}) = G_{\mathbb{Z}}{}^{\star_{\mathbb{R}}\star_{\mathbb{R}}{}'}(\bar{b}) = \underbrace{G_{\mathbb{Z}}{}^{\star_{\mathbb{Q}}\star_{\mathbb{Q}}{}'}(\bar{b}) \leq G_{\mathbb{Z}}{}^{c_{\mathcal{C}}c_{\mathcal{C}}{}'}(\bar{b})}_{\textit{generally strict inequality}} = G_{\mathbb{Z}}{}^{c_{\mathcal{G}}c_{\mathcal{G}}{}'}(\bar{b}) = G_{\mathbb{Z}}{}^{c_{\mathcal{S}}c_{\mathcal{S}}{}'}(\bar{b}) = G_{\mathbb{Z}}(\bar{b}) \; . \tag{2.21}$$

Coupling	Weak duality	Strong duality	Reference
Subadditive	Yes	Yes	[19, Theorem 1]
Gomory	Yes	Yes	[5, Theorem 5.2]
Chvátal	Yes	Yes	[5, Proposition 2.18,Theorem 5.2]
Rational linear	Yes	No	[19, Theorem 2]
Real linear	Yes	No	[19, Theorem 2]

Table 2.6: PILP strong or weak duality for the five schemes

We complete the parallel between convex optimization and PILP that Blair and Jeroslow made in [19].

Convex optimization	PILP
Polyhedral function	Gomory function
Linear function	Chvátal function
Convex cone	Monoid [19]
Polyhedron	Slice [19]

Table 2.7: Comparison between convex optimization and PILP

Perspective of new dual problems for PILP

Looking at the inequality in (2.21), we realize that the duality gap for PILP is between the Fenchel biconjugate ${G_{\mathbb{Z}}}^{\star_{\mathbb{R}}\star_{\mathbb{R}}'}$ (which is equal to the continuous relaxation) and the Chvátal biconjugate ${G_{\mathbb{Z}}}^{cccc'}$ (which is equal to the value of the PILP). In an ideal world, each time we would want to solve a PILP, we would work on its Jeroslow dual problem (2.3a) with Chvátal functions and obtain strong duality. Unfortunately, the Jeroslow dual problem with Chvátal functions might not be easier to solve than the original PILP as it implies to find a Chvátal function that achieves the strong duality and as a Chvátal function $F \in \mathcal{C}^m$ have no restriction on the number time the ceiling function $\lceil \cdot \rceil$ is used in its expression (potentially a exponential number of time in the number of constraints m).

However, if we renounce to strong duality, and only look for dual problems that achieves weak duality with a tighter gap than the continuous relaxation, we could try to solve the Jeroslow dual problem with Chvátal functions that have a fixed number of the ceiling function in their expression. Going even further, by considering an even smaller set of Chvátal functions, we would get the following dual problem.

Proposition 2.19. *Let $A \in \mathbb{Q}^{m\times n}$, $k \in \mathbb{Q}^n$, $\bar{b} \in \mathbb{Q}^m$ define the usual PILP value function $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ as in* (2.1)*. Let $\alpha \in \mathbb{Q}_+$ and . Then,*

$$\tau = \sup_{\substack{\lambda\in\mathbb{Q}^m \\ \langle \lambda,\, A_j\rangle + \alpha\lceil\langle \lambda,\, A_j\rangle\rceil \le k_j\ , \\ \forall j \in \{1,\dots,n\}}} \langle \lambda,\, \bar{b}\rangle + \alpha\lceil\langle \lambda,\, \bar{b}\rangle\rceil \tag{2.22}$$

is a quasi-affine program*, meaning that the objective function is quasi-affine [24] and that the constraints are quasi-affine and $\tau \geq G_{\mathbb{R}}(\bar{b})$ if the affine program* (2.22) *is feasible.*

In Proposition 2.19, for a fixed nonnegative rational $\alpha \in \mathbb{Q}_+$, the considered Chvátal functions $F \in \mathcal{C}^m$ belongs to the family $\{\langle \lambda,\, \cdot\rangle + \alpha\lceil\langle \lambda,\, \cdot\rangle\rceil\}_{\lambda\in\mathbb{Q}^m}$ whose function only has one ceiling function in there expressions. It remains to be seen if tighter gaps are obtained by such quasi-affine program is tighter than continuous relaxation and if the computing time of quasi-affine programs is worth the gain in hypothetical gap tightness.

Chapter 3

Partial perturbation-duality scheme for PILP

In this chapter, we conclude the analogy between between the 'Geoffrion Lagrangian relaxation' method and the perturbation-duality scheme that we started in §1.2, by applying the perturbation-duality scheme to PILP with a partial perturbation of the right-hand side of the constraints. Especially, we apply the scheme to the knapsack problem.

In §3.1, we apply a real linear perturbation-duality scheme to 0-1 linear programs (particularly to the knapsack problem). In §3.2, we apply a Chvátal perturbation-duality scheme to an ILP only partially perturbed.

3.1 Partial linear perturbation-duality scheme for 0-1 linear programming

In §3.1.1, we present the knapsack problem, its continuous relaxation and the already known optimal solution of its continuous relaxation. In §3.1.2, we apply a real linear perturbation-duality scheme to 0-1LP, in order to retrieve the already known optimal solution of its continuous relaxation in §3.1.3 as a special case.

3.1.1 Knapsack problem continuous relaxation

We first write the knapsack problem in its minimization form. Then we present the well-known solution of the continuous knapsack problem, obtained by sorting the items by efficiency.

Knapsack problem in minimization form

The knapsack problem is a particular case of PILP. The problem is characterized by unsplittable n items indexed by $j \in \{1, \ldots, n\}$, which all have a value $k_j \in \mathbb{Q}_+ \setminus \{0\}$ and a weight $W_j \in \mathbb{Q}_+ \setminus \{0\}$, and by a capacity $L \in \mathbb{Q}_+$ such that

$$L \in [\min_{j\in\{1,\dots,n\}} W_j, \sum_{j=1}^{n} W_j] \cap \mathbb{Q}_+ \ . \tag{3.1}$$

The goal is to put as much value in the knapsack as possible without exceeding the capacity of the knapsack.

Usually, the knapsack problem is considered in its maximization form. To follow the perturbation-duality scheme of Rockafellar [30], we consider the knapsack problem in its minimization form.

$$\forall l \in \mathbb{Q} \ , \ \ G_{\mathbb{Z},K}(l) = \begin{array}{ll} \inf\limits_{x} & \langle -k,\, x\rangle \\ \text{s.t.} & -Wx \geq l \\ & x \in \{0,1\}^n, \end{array} \tag{3.2}$$

where $W = \big(W_j\big) \in \mathbb{Q}^{1\times n}$ and $k = (k_j) \in \mathbb{Q}^n$. As L is the capacity of the original knapsack, we set the anchor to be $\bar{l} = -L \leq 0$. Each component x_j of the optimization variable x corresponds to the decision of taking or not the item j in the knapsack.

The function $G_{\mathbb{Z},K} : \mathbb{Q} \to \overline{\mathbb{R}}$ is the value function of the perturbed knapsack problem where $-l \in \mathbb{Q}$ is the perturbed capacity of the knapsack. We immediately see that

$$G_{\mathbb{Z},K}(l) = +\infty \ , \ \ \forall l > -\min_j W_j \ ,$$

$$G_{\mathbb{Z},K}(l) = \sum_{j=1}^{n} k_j \ , \ \ \forall l < -\sum_{j=1}^{n} W_j \ .$$

The knapsack problem and the algorithms for solving it have been thoroughly studied in [23] and [20]. The value function obtained by perturbing the right-hand side has been studied in [16].

Solution of continuously relaxed knapsack problem

As usual, the continuous relaxed value function $G_{\mathbb{R},K} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ is given by

$$\forall l \in \mathbb{Q} \ , \ \ G_{\mathbb{R},K}(l) = \begin{array}{ll} \inf\limits_{x} & \langle -k,\, x\rangle \\ \text{s.t.} & -Wx \geq l \\ & x \in [0,1]^n. \end{array} \tag{3.3}$$

We first need to define what means '*items sorted by efficiency*'.

Definition 3.1. *We say that the items of the knapsack problem* (3.2)*, defined as couples of value-weight* $(k_j, W_j) \in \big(\mathbb{Q} \setminus \{0\}\big)^2, \forall j \in \{1,\dots,n\}$*, are* sorted by efficiency *if*

$$\frac{k_1}{W_1} \geq \frac{k_2}{W_2} \geq \cdots \geq \frac{k_n}{W_n} \ . \tag{3.4}$$

Theorem 3.2 ([23, Theorem 2.1]). *Suppose that the items are sorted by efficiency as defined in* (3.4). *Consider* $s = \min\left\{i \in \{2,\dots,n\} \middle| \sum_{j=1}^{i} W_j > L\right\}$. *Then, the vector* $\hat{x} \in [0,1]^n$ *given by*

$$\hat{x}_j = 1 \ , \ \ \forall j = 1,\dots,s-1 \ , \tag{3.5a}$$
$$\hat{x}_j = 0 \ , \ \ \forall j = s+1,\dots,n \ , \tag{3.5b}$$
$$\hat{x}_s = \frac{L - \sum_{j=1}^{s-1} W_j}{W_s} \ , \tag{3.5c}$$

is an optimal solution of the continuously relaxed knapsack problem (3.3).

This result, which is proven in [23] by an *ad hoc* argument, is retrieved in §3.1.3 using a linear perturbation-duality scheme.

3.1.2 Linear perturbation-duality scheme applied to general 0-1LP case

To retrieve the result of Theorem 3.2, we look at the more general case of 0-1 linear programming (where as usual the cost vector and the constraints are rational):

$$G_{\mathbb{Z}}(b) = \inf_{\substack{x \\ Ax = b \\ x \in \{0,1\}^n}} \langle k, \ x \rangle \ . \tag{3.6}$$

As discussed in Proposition 1.6, applying linear perturbation-duality scheme will yield the same result as applying the 'Geoffrion Lagrangian relaxation' method to all constraints except the 0-1 constraints on x. Nonetheless, we pursue our objective to rewrite these known results in the perturbation-duality scheme framework of Chapter 1.

Proposition 3.3. *Let* $\mathfrak{R} : \{0,1\}^n \times \mathbb{Q}^m \to \overline{\mathbb{R}}$ *be the Rockafellian defined by*

$$\mathfrak{R}(x,b) = \langle b, \ x \rangle + \delta_{\{0\}}(b - Ax) \ , \ \ \forall x \in \{0,1\}^n \ , \ \ \forall b \in \mathbb{Q}^m \ , \tag{3.7}$$

with an anchor $\bar{b} \in \mathbb{Q}^m$*, and let* $\star_{\mathbb{R}}$ *be the linear coupling defined in* (2.10). *We denote the perturbation function (value function)* $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$ *as in* (2.1) *and* $G_{\mathbb{R}} : \mathbb{R}^m \to \overline{\mathbb{R}}$ *its continuous relaxation as defined in* (2.15). *Then,*

1. *the dual function (see Table 1.1)* $\Psi : \mathbb{R}^m \to \overline{\mathbb{R}}$ *satisfies*

$$\Psi(p) = -\sum_{j=1}^{n} \left(\left(A^T p\right)_j - k_j \right)_+ \ , \ \ \forall p \in \mathbb{R}^m \ , \tag{3.8a}$$

where $(\cdot)_+$ *is the positive part;*

2. *the dual objective function at the anchor (see Table 1.1)* $\Phi_{\bar{b}} : \mathbb{R}^m \to \overline{\mathbb{R}}$ *is* concave *and satisfies*

$$\Phi_{\bar{b}}(p) = \langle \bar{b},\, p \rangle - \sum_{j=1}^{n} \Big(\big(A^T p\big)_j - k_j \Big)_+ \;,\;\; \forall p \in \mathbb{R}^m \;; \tag{3.8b}$$

3. *the dual problem satisfies*

$$\sup_{p \in \mathbb{R}^m} \Big\{ \langle \bar{b},\, p \rangle - \sum_{j=1}^{n} \big(A_j^T p - k_j \big)_+ \Big\} = G_{\mathbb{R}}(\bar{b}) \;; \tag{3.8c}$$

4. *the solutions* $p \in \mathbb{R}^m$ *of the dual problem satisfy*

$$p \in \arg\max_{\tilde{p} \in \mathbb{R}^m} \Phi_{\bar{b}}(\tilde{p}) \iff \exists \mu \in [0,1]^n \;,\;\; b = \sum_{j : A_j^T p > k_j} A_j + \sum_{j : A_j^T p = k_j} \mu_j A_j \;, \tag{3.8d}$$

   where A_j *are the columns of* A*.*

Proof. We prove 4.

$\Phi_{\bar{b}}$ is a sum of polyhedral functions f_j which have the same effective domain. So according to [15] (Proposition 3.71) $\partial \Phi_{\bar{b}} = \sum_j \partial f_j$. With similar arguments and according to [15] (Proposition 3.72) we can write $\partial \left(A_j^T \cdot - k_j \right)_+ (p) = A_j \partial \left(\cdot \right)_+ \left(A_j^T p - k_j \right)$. So

$$\begin{aligned} \partial \Phi_{\bar{b}}(p) &= \{\bar{b}\} - \sum_{i=1}^{n} A_j \partial \left(\cdot \right)_+ \left(A_j^T p - k_j \right) \,, \\ &= \{\bar{b}\} - \sum_{j : A_j^T p > k_j} \{A_j\} - \sum_{j : A_j^T p = k_j} A_j [0,1] \;. \qquad \left(\text{as } \partial \left(\cdot \right)_+ (w) = \begin{cases} \{0\} \text{ if } w < 0 \\ \{1\} \text{ if } w > 0 \\ [0,1] \text{ if } w = 0 \end{cases} \right) \end{aligned}$$

Moreover,

$$p \in \arg\max_{\tilde{p} \in \mathbb{R}^m} \Phi_{\bar{b}}(\tilde{p}) \iff 0 \in \partial \Phi_{\bar{b}}(p) \,,$$

which concludes the proof. □

The positive part appearing in the results was already mentioned by Geoffrion in its first example of application of 'Lagrangian relaxation' [14].

3.1.3 Partial linear perturbation-duality scheme applied to the knapsack problem

We now apply Proposition 3.3 to retrieve the same result as in Theorem 3.2.

Corollary 3.4. *Suppose that the items have been* sorted by efficiency*, meaning*

$$\frac{k_1}{W_1} \geq \frac{k_2}{W_2} \geq \cdots \geq \frac{k_n}{W_n} \ .$$

Let $s \in \{2, \ldots, n\}$ be such that the capacity $L \in \mathbb{Q}_+$ of the knapsack satisfies

$$\sum_{j=1}^{s-1} W_j \leq L < \sum_{j=1}^{s} W_j \ .$$

Let us consider the dual objective function Φ_{-L} *obtained by perturbation-duality using the knapsack value function $G_{\mathbb{Z},K}$ as* the perturbation function*, the coupling $\star_{\mathbb{R}}$, defined in* (2.10)*, and the anchor $\bar{l} = -L$. Then, we have that*

1. *the dual problem, defined by Φ_{-L}, satisfies*

$$\sup_{p\in\mathbb{R}^m} \Phi_{-L}(p) = -\Big(\sum_{j=1}^{s-1} k_j + k_s \frac{L - \sum_{j=1}^{s-1} W_j}{W_s}\Big) = -\sum_{j=1}^{n} k_j \hat{x}_j \ , \tag{3.9a}$$

   where the vector $\hat{x} \in [0, 1]$ is defined as in Theorem 3.2;

2. *the solutions of the dual problem* $\arg\max_{\tilde{p}\in\mathbb{R}^m} \Phi_{-L}(p)$ *are characterized by*

$$p \in \arg\max_{\tilde{p}\in\mathbb{R}^m} \Phi_{-L}(p) \iff \begin{cases} p \in [\frac{k_s}{W_s}, \frac{k_{s-1}}{W_{s-1}}] \ , & \textit{if } \frac{k_{s-1}}{W_{s-1}} \neq \frac{k_s}{W_s} \textit{ and } L = \sum_{j=1}^{s-1} W_j \ , \\ p = \frac{k_s}{W_s} \ , & \textit{if } \frac{k_{s-1}}{W_{s-1}} \neq \frac{k_s}{W_s} \textit{ and } L \neq \sum_{j=1}^{s-1} W_j \ , \\ p = \frac{k_s}{W_s} \ , & \textit{if } \frac{k_{s-1}}{W_{s-1}} = \frac{k_s}{W_s} \ ; \end{cases} \tag{3.9b}$$

3. *if $L = \sum_{j=1}^{s-1} W_j$, then $G_{\mathbb{Z},K}(\bar{l}) = G_{\mathbb{R},K}(\bar{l})$, which means that we have strong duality for the partial linear perturbation-duality scheme.*

3.2 Partial perturbation-duality scheme

In §1.2, we have drawn a parallel between the 'Geoffrion Lagrangian relaxation' method and the perturbation-duality scheme. However, in Chapter 2 we have only considered perturbations of every components b_i of the right-hand side in PILP (2.1), while in the 'Geoffrion Lagrangian relaxation' method the constraints were split in two parts, and only the first part was perturbed:

$$\begin{aligned} \forall b \in \mathbb{Q}^m \ , \ G_{\mathbb{Z}}(b) = \ & \inf_{x} \quad \langle k, \, x \rangle \ , \\ & \text{s.t.} \quad Ax = b \\ & \qquad \tilde{A}x = \tilde{b} \\ & \qquad x \in \mathbb{Z}^n_+ \ , \end{aligned} \tag{3.10}$$

where $G_{\mathbb{Z}} : \mathbb{Q}^m \to \overline{\mathbb{R}}$, $k \in \mathbb{Q}^n$, $A \in \mathbb{Q}^{m \times n}$, $\tilde{A} \in \mathbb{Q}^{\tilde{m} \times n}$, $\tilde{b} \in \mathbb{Q}^{\tilde{m}}$, and where we set an anchor $\bar{b} \in \mathbb{Q}^m$.

Thus, to complete the analogy with the 'Lagrangian relaxation' method, we apply the perturbation-duality scheme to the *partially perturbed PILP* (3.10) with *affine Chvátal evaluation coupling*, where *affine Chvátal functions are defined as follows.*

Definition 3.5. *The* affine Chvátal function space $\mathcal{C}_{\mathcal{A}}{}^m$ *is the intersection of all sets* $E \subset \overline{\mathbb{R}}^{\mathbb{Q}^m}$ *satisfying* (2.5b), (2.5c), *and*

$$\big(u \in \mathbb{Q}^m \mapsto \langle v,\, u\rangle + \beta\big) \in E \;,\;\; \forall v \in \mathbb{Q}^m \;,\;\; \forall \beta \in \mathbb{Q} \;. \tag{3.11}$$

The perturbation-duality scheme we use is given by the *affine Chvátal evaluation coupling* $c_{\mathcal{C}_{\mathcal{A}}}$ and the following Rockafellian $\mathfrak{R} : \{\tilde{x} \in \mathbb{Z}^n_+ | \tilde{A}\tilde{x} = \tilde{b}\} \times \mathbb{Q}^m \to \overline{\mathbb{R}}$ defined by:

$$\mathfrak{R}(x, b) = \langle k,\, x\rangle + \delta_{\{0\}}(b - Ax) \;,\;\; \forall x \in \{\tilde{x} \in \mathbb{Z}^n_+ | \tilde{A}\tilde{x} = \tilde{b}\} \;,\;\; \forall b \in \mathbb{Q}^m \;. \tag{3.12}$$

Proposition 3.6. *The partially perturbed PILP* (3.10) *satisfies the following strong duality property*

$$G_{\mathbb{Z}}(\bar{b}) = G_{\mathbb{Z}}{}^{c_{\mathcal{C}_{\mathcal{A}}} c_{\mathcal{C}_{\mathcal{A}}}{}'}(\bar{b}) = \sup_{F \in \mathcal{C}_{\mathcal{A}}{}^m} \Big\{ F(\bar{b}) \dotplus \inf_{\substack{\tilde{A}x = \tilde{b} \\ x \in \mathbb{Z}^n_+}} \big\{ \langle k,\, x\rangle - F(Ax) \big\} \Big\} \;, \tag{3.13a}$$

which also means that the value function $G_{\mathbb{Z}}$ of the partially perturbed PILP (3.10) *is $c_{\mathcal{C}_{\mathcal{A}}}$-convex.*

Furthermore, if $G_{\mathbb{Z}}(\bar{b}) \in \mathbb{R}$, there exists an affine Chvátal function $F \in \mathcal{C}_{\mathcal{A}}{}^m$, as in Definition 3.5, satisfying

$$G_{\mathbb{Z}}(\bar{b}) = F(\bar{b}) \;, \tag{3.13b}$$

$$\inf_{\substack{\tilde{A}x = \tilde{b} \\ x \in \mathbb{Z}^n_+}} \big\{ \langle k,\, x\rangle - F(Ax) \big\} = 0 \;. \tag{3.13c}$$

More specifically, the function F is given by a Chvátal function $H \in \mathcal{C}^{m+\tilde{m}}$ such that

$$H\big(b, \tilde{b}\big) = F(b) \;,\;\; \forall b \in \mathbb{Q}^m \;, \tag{3.13d}$$

$$H\big(A_j, \tilde{A}_j\big) \le k_j \;, \tag{3.13e}$$

$$H\big(0, 0\big) \le 0 \;, \tag{3.13f}$$

$$\tag{3.13g}$$

where $A_j, \tilde{A}_j$ are respectively the columns of A and $\tilde{A}$.

Proof. First, let us consider the perturbation-duality scheme of the problem (3.10) such that all the right-hand side are perturbed by $(b, \tilde{b}) \in \mathbb{Q}^{m+\tilde{m}}$, and such that $\mathbb{Q}^{m+\tilde{m}}$ is coupled with the Chvátal space $\mathcal{C}^{m+\tilde{m}}$. Then, according to [5, Proposition 2.18,Theorem 5.2], there is a Chvátal function $H \in \mathcal{C}^{m+\tilde{m}}$ that satisfies (3.13e), (3.13f) and

$$G_{\mathbb{Z}}(\bar{b}) = H(\bar{b}, \tilde{b}) \ .$$

Now, we prove that ${G_{\mathbb{Z}}}^{c\mathcal{C}_{\mathcal{A}}c\mathcal{C}_{\mathcal{A}}'} \geq G_{\mathbb{Z}}(\bar{b})$. We have that

$$\begin{aligned}
{G_{\mathbb{Z}}}^{c\mathcal{C}_{\mathcal{A}}c\mathcal{C}_{\mathcal{A}}'}(\bar{b}) = &\sup_{F \in {\mathcal{C}_{\mathcal{A}}}^m} \left\{ F(\bar{b}) \dot{+} \inf_{\substack{\tilde{A}x = \tilde{b} \\ x \in \mathbb{Z}^n_+}} \left\{ \langle k,\ x \rangle - F(Ax) \right\} \right\} , \\
\geq & H(\bar{b}, \tilde{b}) \dot{+} \inf_{\substack{\tilde{A}x = \tilde{b} \\ x \in \mathbb{Z}^n_+}} \left\{ \langle k,\ x \rangle - H(Ax, \tilde{b}) \right\} , && (\text{as } H(\cdot, \tilde{b}) \in \mathcal{C}_{\mathcal{A}}) \\
= & H(\bar{b}, \tilde{b}) \dot{+} \inf_{x \in \mathbb{Z}^n_+} \left\{ \langle k,\ x \rangle - H(Ax, \tilde{A}x) \right\} , \\
= & \sup_{G \in \mathcal{C}^{m+\tilde{m}}} \left\{ G(\bar{b}, \tilde{b}) \dot{+} \inf_{x \in \mathbb{Z}^n_+} \left\{ \langle k,\ x \rangle - G(Ax, \tilde{A}x) \right\} \right\} , \\
= & H(\bar{b}, \tilde{b}) \ , \\
= & G_{\mathbb{Z}}(\bar{b}) \ ,
\end{aligned}$$

which concludes the proof.

$\square$

Part II

Algorithms in abstract convexity

Chapter 4

Global and proximal methods

The first things to come to mind when talking about first order optimization algorithms are the *subgradient descent methods* for proper convex lower semi-continuous objective functions or the *gradient descent methods* for differentiable objective functions, and their *projected* counterparts. We refer the reader to [3] for a comprehensive study of first order methods. When minimizing $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, a proper convex lsc function, on $C \subset \mathbb{R}^n$, a closed convex set, a projected subgradient descent takes the form of a discrete trajectory $\{u_k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ defined by

$$u_{k+1} = \pi_C(u_k - t_k v_k) \ , \ \ v_k \in \partial f(u_k) \ , \ \ \forall k \in \mathbb{N} \ , \ \ u_0 \in \mathbb{R}^n \ , \tag{4.1}$$

where π_C is the projection over the closed convex set C and $\{t_k\}_{k \in \mathbb{N}}$ are the *step sizes.*

In the definition of the trajectory (4.1), we are able to write $u_k - t_k v_k$ because the vectors u_k and the subgradient v_k live in the same vector space, $\mathbb{R}^n$. However, in abstract convexity, subgradients do not necessarily belong to a dual space $\mathcal{V}$ that can be identified to is paired primal space $\mathcal{U}$ (as it is the case in Chapter 2 where a primal space $\mathbb{Q}^m$ is paired with the Chvátal functional space $\mathcal{C}^m$ for instance). Then, the trajectory equation (4.1) does not make sense anymore. In order to define first-order optimization, other methods need to be considered.

In §4.1, we present some global optimization in generalized convexity from [32]. In §4.2, we present a definition of proximal methods for *one sided linear* convexity.

4.1 Global optimization methods

Rubinov presented several global optimization methods in [32] for generalized convexity that remind methods in MILP and metaheuristics such as cutting methods, branch-and-bound methods and tabu search. In §4.1.1, we provide background on generalized convexity and generalized subgradients. In §4.1.2, we present the *abstract cutting plane method*, the *abstract branch-and-bound method* and the *abstract tabu search.*

4.1.1 Background on abstract convexity and generalized subgradients

First we provide a definition for abstract convexity, then we present generalized subgradients.

Abstract convexity

There exists many definitions of *abstract convexity* concerning sets and functions [34]. Here, we are interested by a definition for abstract convex functions that stems from the following characterization of usual proper convex lower semicontinuous functions: a function is proper convex lsc if and only if it is the supremum of its affine minorants [2, Proposition 8.16,Corollary 13.42]

Definition 4.1 ([32, Definition 1.1])**.** *Let $\mathcal{W}$ be a set. Let $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ be a nonempty set of functions, which we call* set of elementary functions*. A function $f : \mathcal{W} \to \overline{\mathbb{R}}$ is called* abstract convex *with respect to H (or H-convex) if there exists a set $D \subset H$ such that f is the upper envelope of the set D:*

$$f(w) = \sup\{h(w) | h \in D\} \ , \ \ \forall w \in \mathcal{W} \ . \tag{4.2}$$

For usual convexity, the set H is the set of affine functions.

Remark 4.2. *Pallaschke and Rolewicz provided a result in [28, §1.2,Equation (1.1.7), Theorem 1.2.6] which states that if the set $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ is stable by the addition of constant, i.e. $H + k \subset H$ for all real number $k \in \mathbb{R}$, then H-convexity is equivalent to c-convexity, with respect to a coupling c.*

Generalized subgradients

Suprisingly, Rubinov defined *abstract supergradients* in [32, Definition 1.7] but not *abstract subgradients*. Here, we define abstract subgradients in a similar fashion.

Definition 4.3. *Let $\mathcal{W}$ be a set. Let $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ be a set of elementary functions. Let $w \in \mathcal{W}$. We call an elementary function $h \in H$* abstract subgradient *of the function $f : \mathcal{W} \to \overline{\mathbb{R}}$ at the point w, if the following inequality is satisfied:*

$$f(w) \leq h(w) \dot{+} \Big[-\Big(h(w') \dot{+} \big(-f(w') \big) \Big) \Big] \ , \ \ \forall w' \in \mathcal{W} \ . \tag{4.3}$$

We call the abstract subdifferential $\partial^H f(w) \subset H$ *of the function f at the point w the set of all abstract subgradients of the function f at the point w.*

Remark 4.4. *When the set of elementary functions is defined with a coupling $c : \mathcal{W} \times \mathcal{V}$ as c-affine functions, the Definition 4.3 of abstract subdifferential corresponds to the* upper-subdifferential *defined in Appendix A.8.*

4.1.2 Abstract convex algorithms

We present now the *abstract cutting plane method*, *abstract branch-and-bound method* and the *abstract tabu search*. We take their definition from Rubinov's book [32].

Abstract cutting plane method

Definition 4.5 ([32, §9.2.3])**.** *Let $\mathcal{W}$ be a set, $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ be a set of elementary functions and $f : \mathcal{W} \to \overline{\mathbb{R}}$ be a H-convex function. Then, the following algorithm is called* abstract cutting plane method*:*

1. *Set $k := 0$. Choose an arbitrary initial point $w_0 \in \mathcal{W}$.*
2. *Calculate an abstract subgradient $h_k \in \partial^H f(w_k)$. Let*
$$f_k(w) = \max_{i=0,\ldots,k} h_i(w) \ , \ \ \forall w \in \mathcal{W} \ .$$
3. *Calculate an optimal solution $\widehat{w} \in \min_{w\in\mathcal{W}} f_k(w)$.*
4. *Set $k := k+1$, $w_k = \widehat{w}$. Repeat from Step 2 until a stop condition is satisfied.*

A proof of the convergence of the method with weak assumptions on the elementary functions and the objective function is given in [28, Theorem 9.1.1].

Abstract branch-and-bound method

Definition 4.6 ([32, §9.2.4])**.** *Let $\mathcal{W}$ be a set, $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ be a set of elementary functions and $f : \mathcal{W} \to \overline{\mathbb{R}}$ be a H-convex function. Then, the following algorithm is called* abstract branch-and-bound method*:*

1. *Take an initial partition P_0 of the feasible set $\mathcal{W}$. Set $k := 0$.*
2. *Let $P_k = \{S_1^k, \ldots S_{N_k}^k\}$, where $\cup_{i=1}^{N_k} S_i^k \subset \mathcal{W}$.*
   For each $i = 1 \ldots, N_k$, choose a finite set of points $\mathcal{W}_i^k \subset S_i^k$.
3. *For each $i = 1, \ldots, N_k$, consider finite set $\mathcal{W}_i^k$ and compute the function h_i^k given by a finite maximum of subgradients:*
$$h_i^k = \max_{h\in\partial^H f(\mathcal{W}_i^k)} h \ .$$
4. *For each $i = 1 \ldots, N_k$, minimize each function h_i^k over the set S_i^k. Let the lower bound $\underline{h}_i^k$ be defined by*
$$\underline{h}_i^k = \min_{w\in S_i^k} h_i^k(w) \ .$$

5. *Compute* $j = \arg\min_{i=1,\ldots,N_k} \underline{h}_i^k$.

6. *Partition the set* S_j^k *such that*
$$S_j^k = \hat{S}^k \cup \tilde{S}^k \ .$$

7. *Set a new partition of* $\mathcal{W}$ *by*
$$P_{k+1} := P_k \setminus \{S_j^k\} \cup \{\hat{S}^k\} \cup \{\tilde{S}^k\} \ .$$

8. *Compute the upper bound* $\overline{f}$ *of the optimal value of* f *defined by*
$$\overline{f} = \min_{i=1,\ldots,N_k} \min_{w \in \mathcal{W}_i^k} f(w) \ .$$

9. *Delete all sets* S_i^k, *for which* $\underline{h}_i^k \geq \overline{f}$, *from the partition* P_{k+1}.

10. *Set* $k := k+1$. *Repeat from Step 2 until a stop condition is satisfied.*

Abstract tabu search

Definition 4.7 ([32, §9.2.5]). *Let* $\mathcal{W}$ *be a set,* $H \subset \overline{\mathbb{R}}^{\mathcal{W}}$ *be a set of elementary functions and* $f : \mathcal{W} \to \overline{\mathbb{R}}$ *be a* H*-convex function. At step* k*, suppose we know an upper bound* $\overline{f}^k$ *of the function* f *and that the set* $\mathcal{W}$ *is partitioned into* $P_k = \{S_1^k, \ldots S_{N_k}^k\}$*. For each* $i = 1, \ldots, N_k$*, we consider a finite number of points* $\left\{w_{ij}^k\right\}_j \subset S_i^k$ *and a finite number of subgradients* $h_{ij}^k \in \partial^H f(w_{ij}^k)$.

We define the following quantities:

- *the* gap *in* S_i^k: $G_i^k = \overline{f}^k - \min_{w \in S_i^k} \max_j h_{ij}^k(w)$;
- *the* precision of the approximation *in* S_i^k: $A_i^k = \sum_{w \in \mathcal{W}_i^k} |f(w_{ij}^k) - h_{ij}^k(w_{ij}^k)|$;
- *the volume* V_i^k *of* S_i^k;
- *the number* $\mathfrak{S}_i^k$ *(in a certain sense) of stationary points in* S_i^k;
- *an approximation of the minimum of the objective function* f *in* S_i^k: $M_i^k = \min_{w \in \mathcal{W}_i^k} f(w)$.

For given increasing functions $\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5 : \mathbb{R} \to \mathbb{R}$, *we define the* quality
$$Q_i^k = \Phi_1(G_i^k) + \Phi_2(A_i^k) + \Phi_3(A_i^k) + \Phi_4(\mathfrak{S}_i^k) - \Phi_5(M_i^k) \tag{4.4}$$
Then, the following algorithm is called abstract tabu search*:*

1. *Take an initial partition P_0 of the feasible set $\mathcal{W}$. Set $k := 0$.*
2. *Let $P_k = \{S_1^k, \dots S_{N_k}^k\}$, where $\cup_{i=1}^{N_k} S_i^k \subset \mathcal{W}$. For each $i = 1, \dots, N_k$, choose a finite set of points $\mathcal{W}_i^k \subset S_i^k$.*
3. *For each $i = 1, \dots, N_k$, consider the finite set $\mathcal{W}_i^k = \left\{w_{ij}^k\right\}_j$ and compute subgradients $h_{ij}^k \in \partial^H f(w_{ij}^k)$.*
4. *Set a quality threshold $\overline{Q}^k$. For each $i = 1, \dots, N_k$, compute the quality Q_i^k of the set $S_{i^*}^k$.*
5. *Delete all S_i^k which cannot contain a global minimum of f.*
6. *Choose a set $S_{i^*}^k$ such that its quality $Q_{i^*}^k$ is high enough: $Q_{i^*}^k \geq \overline{Q}^k$.*
7. *Partition the set $S_{i^*}^k$ into $S_{i^*}^k = \hat{S} \cup \tilde{S}$. Set a new partition $P_{k+1} = P_k \setminus \{Q_{i^*}^k\} \cup \{\hat{S}\} \cup \{\tilde{S}\}$. Repeat from Step 2 until a stop condition is satisfied.*

4.2 Proximal methods with OSL coupling

Here, we will deal with a generalization of proximal method to the special case of *One Sided Linear* (OSL) conjugacy. In §4.2.1, we systematically present OSL conjugacy. In §4.2.2, we define the OSL proximal operator.

4.2.1 One sided linear (OSL) conjugacy

First, we provide basic definitions for OSL conjugacy. Then, we sum up OSL conjugacy results in tables.

Basic definitions for OSL conjugacy

Definition 4.8 ([22, Definition 4.2.7])**.** *Let $\mathcal{W}$ be a set and $\mathcal{V}$ be a vector space. We say that a coupling $c : \mathcal{W} \times \mathcal{V} \to \overline{\mathbb{R}}$ is* one-sided *linear (OSL) if, for all $w \in \mathcal{W}$, the function $c(w, \cdot) : \mathcal{V} \to \overline{\mathbb{R}}$ is linear.*

Remark 4.9. *By the Definition 4.8 of OSL coupling, if $c : \mathcal{W} \times \mathcal{V} \to \overline{\mathbb{R}}$ is OSL, then c is* finite valued.

Definition 4.10 ([7, Definition 2.3])**.** *Let $\mathcal{U}$ and $\mathcal{V}$ be two vector spaces paired by a bilinear form $\langle , \rangle : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$. Let $\mathcal{W}$ be a set and $\theta : \mathcal{W} \to \mathcal{U}$ be a mapping. We define the* one-sided linear coupling *$\star_\theta$ (induced by the primal valued mapping θ between the set $\mathcal{W}$ and the vector space $\mathcal{V}$ by*

$$\star_\theta(w, v) = \langle \theta(w), \, v \rangle \; , \; \; \forall (w, v) \in \mathcal{W} \times \mathcal{V} \; . \tag{4.5a}$$

Similarly, we define the reverse OSL coupling $\star_\theta{}' : \mathcal{V} \times \mathcal{W}$ *by*

$$\star_\theta{}'(v,w) = \langle \theta(w),\ v \rangle \ , \ \forall (w,v) \in \mathcal{W} \times \mathcal{V} \ . \tag{4.5b}$$

Definition 4.11 ([7, Definition 2.4])**.** *Let $\mathcal{W}$ be a set and $\theta : \mathcal{W} \to \mathcal{U}$ be a mapping. Let $h : \mathcal{W} \to \overline{\mathbb{R}}$ be a function. We define the* conditional infimum *(of the function h knowing the mapping θ) as the function* $\inf \big[h \,\big|\, \theta\big] : \mathcal{U} \to \overline{\mathbb{R}}$ *given by*

$$\inf \big[h \,\big|\, \theta\big](u) = \inf \big\{ h(w) \big| w \in \mathcal{W} \ , \ \ \theta(w) = u \big\} \ , \ \ \forall u \in \mathcal{U} \ . \tag{4.6}$$

Systematic study of OSL coupling

Let $\mathcal{W}$, $\mathcal{U}$, $\mathcal{V}$ be three sets such that $\mathcal{U}$ and $\mathcal{V}$ are vector spaces paired by a scalar product $\langle \cdot, \cdot \rangle$. Let us consider $\theta : \mathcal{W} \to \mathcal{U}$ a mapping, $h : \mathcal{U} \to \overline{\mathbb{R}}$ a function to which we apply $\star_\theta$-conjugacy, $g : \mathcal{V} \to \overline{\mathbb{R}}$ a function to which we apply $\star'_\theta$-conjugacy, the complement $\theta(\mathcal{W})^{\mathrm{c}}$ of the set $\theta(\mathcal{W})$, a subset $W \subset \mathcal{W}$, and a subset $V \subset \mathcal{V}$.

conjugate	$h^{\star_\theta} = \big(\inf \big[h \mid \theta\big]\big)^{\star}$
biconjugate	$h^{\star_\theta \star_\theta{}'} = \big(h^{\star_\theta}\big)^{\star'} \circ \theta = \big(\inf \big[h \mid \theta\big]\big)^{\star\star'} \circ \theta$

Table 4.1: Summary of OSL conjugacy

h is such that...	Subdifferentials
$h : \mathcal{W} \to \overline{\mathbb{R}}$	$\partial_{\star_\theta} h \supset \partial\big(\inf \big[h \mid \theta\big]\big) \circ \theta$
$h = \inf \big[h \mid \theta\big]$	$\partial_{\star_\theta} h = \partial\big(\inf \big[h \mid \theta\big]\big) \circ \theta$
h is $\star_\theta$-convex	$\partial_{\star_\theta} h = \partial\big(\inf \big[h \mid \theta\big]\big)^{\star\star'} \circ \theta$
$h = f \circ \theta$	$\partial_{\star_\theta} h = \partial\big(f \dot{+} \delta_{\theta(\mathcal{W})}\big) \circ \theta \supset \partial f \circ \theta$

Table 4.2: OSL-subdifferentials

indicator conjugate	$\delta_W^{\star_\theta} = \sigma_{\theta(W)}$
indicator biconjugate	$\delta_W^{\star_\theta \star_\theta{}'} = \delta_{\overline{\mathrm{co}}(\theta(W))} \circ \theta$

Table 4.3: Summary of OSL conjugacy applied to the Indicator function $\delta_{\mathcal{W}}$

Systematic study of reverse OSL coupling

We provide the same summary for the reverse OSL conjugacy.

conjugate	$g^{\star_\theta{}'} = g^{\star'} \circ \theta$
biconjugate	$g^{\star_\theta{}' \star_\theta} = \big(g^{\star'} \dot{+} \delta_{\theta(\mathcal{W})}\big)^{\star}$

Table 4.4: Summary of reverse OSL conjugacy

g is such that ...	Subdifferentials
$g : \mathcal{V} \to \overline{\mathbb{R}}$	$\partial_{\star_\theta'} g = \theta^{-1} \circ \partial g$
g is $\star_\theta{}'$-convex	$\partial_{\star_\theta'} g = \theta^{-1} \circ \big(\partial(g^{\star'} \dotplus \delta_{\theta(\mathcal{W})})\big)^{-1}$
$g = \big(\psi \dotplus \delta_{\theta(\mathcal{W})}\big)^{\star}$	$\partial_{\star_\theta'} g = \theta^{-1} \circ \partial\big(\psi \dotplus \delta_{\theta(\mathcal{W})}\big)^{\star}$

Table 4.5: reverse OSL-subdifferentials

indicator conjugate	$\delta_V^{\star_\theta{}'} = \sigma_V \circ \theta$
indicator biconjugate	$\delta_V^{\star_\theta{}'\star_\theta}(\cdot) = \sup\limits_{u \in \theta(\mathcal{W}), v \in V} \langle u, \cdot - v \rangle$

Table 4.6: Summary of reverse OSL conjugacy applied to the indicator function

4.2.2 OSL proximal operator

We remind that proximal points methods are iterative methods, where a proximal operator is iteratively applied [3, Chapter 6]. Thus, we only define here a generalization of the proximal operator to OSL coupling

To do so, let us first define the usual proximal operator, then let us define Bregman distances, before defining the OSL proximal operator.

Usual proximal operator

Definition 4.12 ([26, Proposition 3a] [2, Definition 12.23]). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex lower semi-continuous function. Let $\|\cdot\|_2$ be the usual euclidean norm on $\mathbb{R}^n$. Then for all $u \in \mathbb{R}^n$,*

$$\inf_{\tilde{u} \in \mathbb{R}^n} \left\{ f(\tilde{u}) + \frac{1}{2} \|\tilde{u} - u\|_2^2 \right\} , \tag{4.7}$$

is uniquely attained. *We call* $\mathrm{Prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ *the* proximal operator of f *and we define it by*

$$\mathrm{Prox}_f(u) = \operatorname*{arg\,min}_{\tilde{u} \in \mathbb{R}^n} \left\{ f(\tilde{u}) + \frac{1}{2} \|\tilde{u} - u\|_2^2 \right\} , \ \ \forall u \in \mathbb{R}^n . \tag{4.8}$$

We refer the reader to [29, §3] for several interpretations of the proximal operator such as the *Moreau-Yosida regularization*, *the resolvent of subdifferential operator*, *the modified gradient step*.

Generalization of proximal operator with Bregman distances

Similarly to how the *mirror descent method* is designed [4], it is possible to generalize the proximal operator by replacing the function $\frac{1}{2}\|\cdot\|^2$ by a Bregman distance.

Definition 4.13 ([22, Definition 4.3.1])**.** *Let $\mathcal{W}$ and $\mathcal{V}$ be two sets, and let $c : \mathcal{W} \times \mathcal{V}$ be a coupling. Let $h : \mathcal{W} \to]-\infty, +\infty]$ be a Bregman function such that, for all $w \in \mathcal{W}$, the subdifferential $\partial^c h(w) : \mathcal{W} \rightrightarrows \mathcal{V}$ is* single valued*, meaning $\partial^c h(w) = \{\nabla h(w)\}$; and such that $\nabla h : \mathcal{W} \to \mathcal{V}$ is* injective.

We define the c-Bregman distance *associated with h as the function*

$$D^{c,h} : \mathcal{W} \times \mathcal{W} \to \overline{\mathbb{R}} \,, \tag{4.9}$$

given by

$$D^{c,h}(w, w', v') = h(w) \dotplus h^c(\nabla^c h(w')) \dotplus \big(-c(w', \nabla^c h(w'))\big) \,. \tag{4.10}$$

Thus, for a given Bregman function $h : \mathcal{W} \to]-\infty, +\infty]$ as in Definition 4.13, the OSL proximal operator of a function $f : \mathcal{W} \to \overline{\mathbb{R}}$, could be informally defined as

$$\mathrm{Prox}_f^{\star_\theta,h}(w') = \operatorname*{arg\,min}_{w \in \mathcal{W}} \big\{ f(w) \dotplus D^{\star_\theta,h}(w, w') \big\} \,,$$

if we suppose that the arg min exists. Furthermore, informally,

$$\begin{aligned}
\mathrm{Prox}_f^{\star_\theta,h}(w') &= \operatorname*{arg\,min}_{w \in \mathcal{W}} \big\{ f(w) \dotplus D^{\star_\theta,h}(w, w') \big\} \,, \\
&= \operatorname*{arg\,min}_{w \in \mathcal{W}} \big\{ f(w) \dotplus h(w) \dotplus h^{\star_\theta}\big(\nabla^{\star_\theta} h(w')\big) \dotplus \big(-\star_\theta(w, \partial^{\star_h}(w'))\big) \big\} \,, \\
&= \operatorname*{arg\,min}_{w \in \mathcal{W}} \big\{ f(w) \dotplus h(w) \dotplus \big(-\star_\theta(w, \nabla^{\star_\theta} h(w'))\big) \big\} \,, \\
&= \operatorname*{arg\,max}_{w \in \mathcal{W}} \big\{ \star_\theta(w, \nabla^{\star_\theta} h(w') \dot{+} \big(-f(w) \dotplus h(w)\big)\big) \big\} \,,
\end{aligned}$$

which is equivalent to

$$\partial^{\star_\theta}\big(f \dotplus h\big)\big(\mathrm{Prox}_f^{\star_\theta,h}(w')\big) \ni \nabla^{\star_\theta} h(w') \,,$$

leading to the formal definition of *OSL proximal operator*.

Definition 4.14. *Let $f : \mathcal{W} \to \overline{\mathbb{R}}$ be a function. Let $\theta : \mathcal{W} \to \mathcal{U}$ be a mapping. Let $\star_\theta : \mathcal{W} \times \mathcal{V} \to \overline{\mathbb{R}}$ be a OSL coupling. Let $h : \mathcal{W} \to]-\infty, +\infty]$ be a Bregman function as in Definition 4.13.*

Then, we call the OSL proximal operator *$Prox_f^{\star_\theta,h} : \mathcal{W} \rightrightarrows \mathcal{W}$ of the function f associated with h the function defined by*

$$Prox_f^{\star_\theta,h} = \big(\partial^{\star_\theta}\big(f \dotplus h\big)\big)^{-1} \circ \nabla^{\star_\theta} h \,. \tag{4.11}$$

Proposition 4.15. *Let us assume that the Bregman function $h : \mathcal{W} \to]-\infty, +\infty]$ can be decomposed as follows: $h = q \circ \theta$, where $q : \mathcal{U} \to]-\infty, +\infty]$. Then, we have that*

$$Prox_f^{\star_\theta,h}(w) = \theta^{-1} \circ \Big(\partial\big(f \dotplus q \dotplus \delta_{\theta(\mathcal{W})}\big)\Big)^{-1} \circ \nabla\big(q \dotplus \delta_{\theta(\mathcal{W})}\big) \circ \theta(w) \,, \forall w \in \mathcal{W} \tag{4.12}$$

Chapter 5

Numerical results for E-Capra convex problems

In this chapter, we apply the *abstract cutting plane* algorithm from [32, §9.2.30] to E-Capra convex problems that we present in §5.1. In §5.2, we provide a formulation of the abstract cutting plane algorithm adapted to E-Capra convex problems. Finally, in §5.3, we present the instances of the different problems and the numerical results of abstract cutting plane methods.

5.1 E-Capra convex optimization problems

First, we provide background on Capra conjugacy in §5.1.1. Then, we consider two E-Capra convex problems, the minimization of the ratio of the ℓ_1 norm over ℓ_2 norm on a blunt[1] closed convex cone, in §5.1.2; the minimization of the ℓ_0 pseudonorm on a blunt closed convex cone, in §5.1.3; and the spark of a matrix, in §5.1.4, which has a E-Capra convex objective function but not a E-Capra convex feasible set.

5.1.1 Capra convexity

First we define Capra conjugacy. Then, we provide characterization of Capra convex functions and Capra convex sets.

Capra conjugacy

Definition 5.1 (Mostly from [13, Definition 2.1])**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$. We define the* coupling *$¢ : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ between $\mathbb{R}^m$ and $\mathbb{R}^m$, that we call the* Capra coupling*, by*

$$\forall y \in \mathbb{R}^m \ , \ ¢(x, y) = \begin{cases} \frac{\langle x, y \rangle}{\|x\|} \ , & \text{if } x \neq 0 \ , \\ 0 \ , & \text{if } x = 0 \ . \end{cases} \tag{5.1a}$$

[1]meaning without the origin $0 \in \mathbb{R}^n$.

Similarly, we define the reverse coupling $\cent' : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ *between* $\mathbb{R}^m$ *and* $\mathbb{R}^m$*, that we call the* Capra reverse coupling*, by*

$$\cent'(y,x) = \cent(x,y) \ , \ \ \forall x, y \in \mathbb{R}^m \ . \tag{5.1b}$$

When the norm $\|\cdot\|$ *is the Euclidean norm on* $\mathbb{R}^n$*, we call that the coupling* $\cent$ the euclidean Capra (E-Capra) *coupling.*

Remark 5.2. $\cent$ *is a one-sided linear coupling for* $\theta = n$*, where* $n : \mathbb{R}^m \to \mathbb{R}^m$ *is defined by*

$$n(x) = \begin{cases} \frac{x}{\|x\|} \ , & \textit{if } x \neq 0 \ , \\ 0 \ , & \textit{if } x = 0 \ , \end{cases} \quad \forall x \in \mathbb{R}^m \ . \tag{5.2}$$

Capra convex functions and sets

Here, we present a characterization of Capra convex functions in the sense of the Definition A.6 of c-convexity. Then the definition and a characterization (in the case of a ℓ_p source norm) of Capra convex sets, in order to define Capra *convex problems*, which are defined as an optimization program where the objective function and the feasible set are Capra convex.

Proposition 5.3 ([7, Proposition 3.3])**.** *A function on* $\mathbb{R}^d$ *is* $\cent$*-convex if and only if it is the composition of a closed convex function on* $\mathbb{R}^d$ *with the normalization mapping* (5.2)*. More precisely, for any function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$*, we have the equivalences*

$$\begin{aligned} & h \textit{ is } \cent\textit{-convex} \ , \\ \iff & h = h^{\cent\cent'} \ , \\ \iff & h = (h^{\cent})^{\star'} \circ n \ , \qquad\qquad \text{(where } (h^{\cent})^{\star'} : \mathbb{R}^d \to \overline{\mathbb{R}} \text{ is a closed convex function)} \\ \iff & \textit{there exists a closed convex function } f : \mathbb{R}^d \to \overline{\mathbb{R}} \textit{ such that } h = f \circ n \ . \end{aligned}$$

Now, we present the definition and a characterization of Capra convex sets.

Definition 5.4 ([22, Definition 6.2.1])**.** *Let* $\|\cdot\|$ *be a source norm. Let* $\cent$ *be the corresponding* Capra *coupling, as in Definition 5.1. We say that the set* $D \subset \mathbb{R}^d$ *is* Capra convex *if the indicator function* δ_D *is a* Capra *convex function, meaning* $\delta_D = \delta_D^{\cent\cent'}$ *.*

Proposition 5.5 ([22, Proposition 6.2.6])**.** *Let the source norm be the* ℓ_p *norm defined by* $\ell_p(x) = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \forall x \in \mathbb{R}^n$*, where* $p \in]1, +\infty,[$*. Let* $\cent$ *be the corresponding Capra coupling, as in Definition 5.1, and let* $n : \mathbb{R}^d \to \mathbb{R}^d$ *be the corresponding normalization mapping. Let* $D \subseteq \mathbb{R}^d$ *be a nonempty set. Then, we have the equivalence*

$$D \textit{ is } \text{Capra} \textit{ convex} \iff \begin{cases} D \textit{ is a cone,} \\ D \cup \{0\} \textit{ is closed,} \\ D \cap \{0\} = \overline{\text{co}}\big(n(D)\big) \cap \{0\} \end{cases} . \tag{5.3}$$

5.1.2 Ratio of two norms

Here, we define the E-Capra convex problem of minimizing of the ratio of the ℓ_1 norm over ℓ_2 norm on a blunt closed convex cone. We remind that the norms $\ell_1 : \mathbb{R}^n \to \mathbb{R}_+$ and $\ell_2 : \mathbb{R}^n \to \mathbb{R}_+$ are defined by

$$\ell_1(x) = \sum_{i=1}^{n} |x_i| \ , \ \ \forall x \in \mathbb{R}^n \ , \tag{5.4a}$$

$$\ell_2(x) = \sqrt{\sum_{i=1}^{n} x_i^2} \ , \ \ \forall x \in \mathbb{R}^n \ . \tag{5.4b}$$

Proposition 5.6. *Let $\{u_1, \ldots, u_r\} \subset \mathbb{R}^n$ be a set of real vectors. Consider $C \subset \mathbb{R}^n$ the closed convex cone defined by*

$$C = \mathrm{cone}(u_1, \ldots, u_r) \ , \tag{5.5}$$

where $\mathrm{cone}(u_1, \ldots, u_r)$ are all the nonnegative combinations of the set $\{u_1, \ldots, u_r\}$.

Then, the minimization problem defined by

$$\inf_{x \in C \setminus \{0\}} \frac{\ell_1(x)}{\ell_2(x)} \tag{5.6}$$

is E-Capra convex.

Proof.

- The objective function $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}_+$ of the problem (5.6) is defined by

  $$f(x) = \frac{\ell_1(x)}{\ell_2(x)} \ , \ \ \forall x \in \mathbb{R}^n \ ,$$

  is E-Capra convex according to Proposition 5.3, as the function f satisfies $f = \ell_1 \circ \frac{\cdot}{\ell_2(\cdot)}$ and the ℓ_1 norm is a proper convex lsc function.

- The closed convex cone C in (5.5) satisfies

  $$\begin{cases} C \cup \{0\} \text{ is closed,} \\ C \cap \{0\} = \overline{\mathrm{co}}\big(n(C)\big) \cap \{0\} \ . \end{cases}$$

  Thus, we can apply Proposition 5.5 which yields the E-Capra convexity of C using Proposition 5.3, as C is a closed convex set.

Thus, the 'ratio of two norms over a blunt cone' problem (5.6) is an E-Capra convex problem.

□

5.1.3 Counting pseudonorm ℓ_0

Here, we define the E-Capra convex problem of minimizing the ℓ_0 pseudonorm on a blunt closed convex cone. We remind that the pseudonorm $\ell_0 : \mathbb{R}^n \to \{0, \dots, n\}$ is defined by

$$\ell_0(x) = |\{i : x_i \neq 0\}| \ , \ \ \forall x \in \mathbb{R}^n \ , \tag{5.7}$$

where $|\cdot|$ is the cardinality function.

Proposition 5.7. *Let $C \subset \mathbb{R}^n$ be the closed convex cone defined in* (5.5)*. Then, the minimization problem*

$$\inf_{x \in C \setminus \{0\}} \ell_0(x) \ , \tag{5.8}$$

*is E-*Capra *convex.*

Proof. The Capra convexity of ℓ_0 has been proven in [7]. We have proven that C is a E-Capra convex set in the proof of Proposition 5.6. □

5.1.4 Spark of a matrix

Here, we define the E-Capra convex problem of computing the spark of a matrix.

Definition 5.8. *Let $A \in \mathbb{R}^{m \times n}$ be a real matrix. We define the* spark *of the matrix A by*

$$\mathrm{spark}(A) = \min_{\substack{Ax=0 \\ x \in \mathbb{R}^n \setminus \{0\}}} \ell_0(x) \ . \tag{5.9}$$

The complexity of the computation of the spark for a given matrix have been studied in [36, §II]; the problem is actually NP-hard [36, Corollary 1].

Remark 5.9. *We can interpret the spark of a matrix A as the smallest number of* linearly dependent *columns of A. If we consider the matrix A as matroid, the spark is also the* size of the smallest circuit *on A [36, §II].*

Problems	Min of the ratio of two norms in a blunt cone	Min of ℓ_0 in a blunt cone	Spark of a matrix
Objective function	ℓ_1/ℓ_2	ℓ_0	ℓ_0
Feasible set	$\mathrm{cone}(u_1, \dots u_r) \setminus \{0\}$	$\mathrm{cone}(u_1, \dots u_r) \setminus \{0\}$	$\{x \in \mathbb{R}^n \setminus \{0\} : Ax = 0\}$
Objective E-Capra convex	✓	✓	✓
Feasible set E-Capra convex	✓	✓	

Table 5.1: Summary of three minimization problems with E-Capra objective function

5.2 Implementation

In §5.2.1, we present the E-Capra cutting plane method. In §5.2.2, we present the modified E-Capra cutting plane method for ℓ_0 that we will use for numerical tests in §5.3.

5.2.1 E-Capra cutting plane: linear program over the Euclidean sphere

First, we present a reformulation for E-Capra programs. Then, we present the E-Capra cutting plane algorithm, and finally formulas of E-Capra subgradients for particular cases.

Reformulation of the E-Capra problems with the E-Capra cutting plane method

To apply a cutting plane method to the three E-Capra minimization problems from Table 5.1.4, we need to rewrite them in a such a way that makes the E-Capra cutting planes apparent. We treat the three E-Capra convex problems at the same time here. So let us consider a general E-Capra minimization on a blunt closed convex cone.

Proposition 5.10. *Consider a E-*Capra *convex function $f : \mathbb{R}^n \to \mathbb{R}$, a closed convex cone $C \subset \mathbb{R}^n$ that satisfies the conditions in* (5.3)*, and the E-*Capra *minimization problem they define:*

$$\inf_{x \in C \setminus \{0\}} f(x) \; . \tag{5.10}$$

Then, the problem (5.10) *is equivalent to*

$$\inf_{\substack{x \in C \\ \ell_2(x) = 1}} \quad \sup_{y \in \mathbb{R}^n} \left\{ \langle x , \, y \rangle - f^{\cent}(y) \right\} \quad . \tag{5.11}$$

Proof.

$$
\begin{aligned}
& \inf_{x\in C\setminus\{0\}} f(x) \; , \\
\iff & \inf_{x\in C\setminus\{0\}} f^{¢¢'}(x) \; , && \text{(as } f = f^{¢¢'} \text{ by E-}\textsc{Capra}\text{ convexity of } f\text{)} \\
\iff & \inf_{x\in C\setminus\{0\}} \sup_{y\in\mathbb{R}^n} \Big\{ \Big\langle \frac{x}{\ell_2(x)}, \; y \Big\rangle - f^{¢}(y) \Big\} \; , && \text{(by definition of } ¢'\text{-conjugacy)} \\
\iff & \inf_{\substack{x/\ell_2(x)\in C \\ x\in\mathbb{R}^n\setminus\{0\}}} \sup_{y\in\mathbb{R}^n} \Big\{ \Big\langle \frac{x}{\ell_2(x)}, \; y \Big\rangle - f^{¢}(y) \Big\} \; , && \text{(as } C \text{ is a cone)} \\
\iff & \inf_{\substack{x\in C \\ \ell_2(x)=1 \\ x\in\mathbb{R}^n\setminus\{0\}}} \sup_{y\in\mathbb{R}^n} \Big\{ \langle x, \; y\rangle - f^{¢}(y) \Big\} \; , && \text{(by change of variable)} \\
\iff & \inf_{\substack{x\in C \\ \ell_2(x)=1}} \sup_{y\in\mathbb{R}^n} \Big\{ \langle x, \; y\rangle - f^{¢}(y) \Big\} \; , && \text{(as } \ell_2(x) = 1 \implies x \neq 0\text{)}
\end{aligned}
$$

□

E-Capra cutting plane algorithm

The E-Capra cutting plane method consists of successive approximations of an E-Capra convex problem formulated as in (5.11). Basically, in the formulation (5.11), we replace $\sup_{y\in\mathbb{R}^n}$ by $\max_{y\in\mathrm{Cuts}_k}$ at the iteration k, where $\mathrm{Cuts}_k \subset \partial^{¢} f$ is a set of subgradients of the E-Capra convex objective function f.

Definition 5.11. *We call the following algorithm* the E-Capra cutting plane method.

1. *Set* $k := 0$*. Find* $x_0 \in C$ *such that* $\ell_2(x_0) = 1$*.*
2. *Find* $y^k \in \partial^{¢} f(x^k)$*.*
3. *Find an optimal solution* x^{k+1} *of the* subproblem $\displaystyle \inf_{\substack{x\in C \\ \ell_2(x)=1}} \max_{h=0}^{k-1} \big\{ \langle x, \; y_h\rangle - f^{¢}(y_h) \big\}$.
4. *Set* $k := k+1$*. Repeat from Step 2 until a stop condition is satisfied.*

E-Capra subgradients of the objective functions

Here, we gather the formulas for the E-Capra subdifferential of ℓ_0 and ℓ_1/ℓ_2 that are used in the algorithm described in §5.2.1. First, we remind the definition of top-k norm.

Definition 5.12 (from [9, Definition 9]). *For all subset of indices $K \subset \{1,\dots,n\}$ and all real vector $x \in \mathbb{R}^n$, the real vector $x_K \in \mathbb{R}^n$ is defined by*

$$x_{K,i} = x_i \ , \ \ \forall i \in K \ , \tag{5.12a}$$
$$x_{K,i} = 0 \ , \ \ \forall i \in \{1,\dots,n\} \setminus K \ . \tag{5.12b}$$

For $k \in \{1,\dots,n\}$, we call top-k *norm associated with the source norm ℓ_2* (5.4b) *the norm defined by*

$$\|x\|^{\mathrm{tn}}_{k,2} = \sup_{K \subset \{1,\dots,n\}, |K| \leq k} \ell_2(x_K) \tag{5.12c}$$

Proposition 5.13. *Let $¢ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the E-*Capra *coupling defined by the source norm ℓ_2 and $x \in \mathbb{R}^n \setminus \{0\}$ be a real vector different from 0.*

- *Considering the permutation $\nu : \{1,\dots,n\} \to \{1,\dots,n\}$ such that $|x_{\nu(1)}| \geq \cdots \geq |x_{\nu(n)}|$ and considering the set* $\mathrm{supp}(x) = \Big\{ i \in \{1,\dots,n\} \big| x_i \neq 0 \Big\}$, *we have that [22, Proposition 5.4.7, §5.5.1]*

$$y \in \partial_{¢} \ell_0(x) \iff \begin{cases} \exists \lambda \in \mathbb{R}_+ \ , \ \ y_i = & \lambda x_i \ , \ \ \forall i \in \mathrm{supp}(x) \ , \\ |y_j| \leq & \min_{i \in \mathrm{supp}(x)} |y_i| \ , \ \ \forall j \notin \mathrm{supp}(x) \ , \\ |y_{\mathrm{supp}(k+1)}| \geq & (\|y\|^{\mathrm{tn}}_{k,2} + 1)^2 - (\|y\|^{\mathrm{tn}}_{k,2})^2 \ , \\ & \forall k \in \{0,\dots,\ell_0(x)-1\} \ , \\ |y_{\mathrm{supp}(\ell_0(x)+1)}| \leq & (\|y\|^{\mathrm{tn}}_{\ell_0(x),2} + 1)^2 - (\|y\|^{\mathrm{tn}}_{\ell_0(x),2})^2 \ , \\ & \forall k \in \{0,\dots,\ell_0(x)-1\} \ . \end{cases} \tag{5.13a}$$

- *Considering the component wise sign function* $\mathrm{sign} : \mathbb{R}^n \to \{-1,0,1\}^n$, *we also have that*

$$y \in \partial_{¢}(\ell_1/\ell_2)(x) \iff y = \mathrm{sign}(x) \ . \tag{5.13b}$$

Proof. The proof of (5.13b) is left as an exercise for the reader. □

5.2.2 Modified E-Capra cutting plane for the pseudonorm ℓ_0

Applying the E-Capra cutting plane presented in Definition 5.11 without any changes when the objective function is the pseudonorm ℓ_0 leads to two major weaknesses that we present here, before proposing solutions. Then, we present a modified E-Capra cutting plane for the pseudonorm ℓ_0.

Problems with the E-Capra cutting plane method

1. At Step 2 in Definition 5.11, during the computation of a E-Capra uppersubgradient $y^k \in \partial^{¢} f(x^k)$, the norm $\left\|y^k\right\|_2$ of subgradients tends to infinity when there are indices i_k such that $x^k_{i_k} \neq 0$ and $x^k_{i_k} \xrightarrow[k \to \infty]{} 0$.

2. The sphere constraint $\ell_2(x) = 1$ in the problem reformulation (5.11) is not a convex constraint, even if it is differentiable. Thus, when solving the subproblem at Step 3 in Definition 5.11 with a usual projected gradient descent, we are not guaranteed to find a minimizer, but only a local minimizer of the subproblem.

Proposed solutions

1. If $0 \neq |x^k_{i_k}| < \varepsilon$, where $\varepsilon > 0$ is a fixed threshold, we project x^k on the i_k-th axis before computing a E-Capra subgradient $y^k \in \partial^{\cent} f(x^k)$ at Step 2 in Definition 5.11.

2. To compensate for the nonoptimal resolution of the nonconvex subproblem at Step 3 in Definition 5.11, we use local search: we keep track of the minimal known value $\overline{\ell}_0^k$ at iteration k and we set the $1 + \overline{\ell}_0^k$ smallest components of the current solution x^k to 0 and check if the resulting vector belongs to the E-Capra convex set. If it does, we set $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k - 1$, otherwise we set $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k - 1$.

   Furthermore, as we have the implication $\ell_0(x) \leq k \implies \ell_1(x) \leq k$, we add the following admissible constraint

$$\ell_1(x) \leq \underbrace{\overline{\ell}_0^k}_{\text{minimal known value of } \ell_0 \text{ at step } k} ,$$

   to the nonconvex subproblem at Step 3, each time we improve the best known value $\overline{\ell}_0^k$. Doing so restrict the feasible space of the nonconvex subproblem in the hope of removing local minima of the subproblem, increasing the probability of finding the global minimum of the subproblem.

We present now the modified E-Capra cutting plane for the pseudonorm ℓ_0, that we will use for the numerical tests in §5.3.

Definition 5.14. *We call the following algorithm* the modified E-Capra cutting plane method for the pseudonorm ℓ_0*.*

1. *Set a threshold* $\varepsilon > 0$*. Set* $k := 0$*. Set the upper bound* $\overline{\ell}_0^k = n$*. Find* $x_0 \in C$ *such that* $\ell_2(x_0) = 1$*.*

2. *For each* $i \in \{1, \dots, n\}$*, if* $|x^k_i| < \varepsilon$*, set* $x^k_i := 0$*.*

3. *Find* $y^{k+1} \in \partial^{\cent} \ell_0(x^k)$*.*

4. *Find an optimal solution* x^k *of the* subproblem $\inf\limits_{\substack{x \in C \\ \ell_2(x) = 1 \\ \ell_1(x) \leq \overline{\ell}_0^k}} \max_{h=0}^{k-1} \left\{ \langle x,\, y_h \rangle - \ell_0^{\cent}(y_h) \right\}$.

5. *Set* $\hat{x} := x^{k+1}$. *Set the* $1 + \overline{\ell}_0^k$ *smallest components of* $\hat{x}$ *to* 0. *If* $\hat{x} \in C \setminus \{0\}$, *set* $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k - 1$. *Otherwise, set* $\overline{\ell}_0^{k+1} := \overline{\ell}_0^k$.

6. *Set* $k := k + 1$. *Repeat from Step 2 until a stop condition is satisfied.*

5.3 Numerical results

We aim to test the E-Capra *cutting plane method*, presented in Definition 5.11, with the following numerical tests on the three 'E-Capra convex' problems from Table 5.1.4. To our knowledge, no numerical tests of the *abstract cutting plane method* have been done so far.

In §5.3.1, we present the instances on which we apply the E-Capra cutting plane method. In §5.3.2, in §5.3.3 and in §5.3.4, we respectively present the numerical results of the tests for the minimization of the ratio of two norms, the minimization of the pseudonorm ℓ_0 and the spark of matrix. We discuss the results in §5.3.5.

5.3.1 Generated instances

The first goal of this numerical tests is to check if the E-Capra cutting plane method converges on simple instances. The second goal is estimate the influence of the dimension on the time of convergence. To do so, we have generated cones that contains a known optimal solution for the minimization of the ratio of two norms and the minimization of the pseudonorm ℓ_0. For the spark of a matrix, we have generated square matrices $A \in \mathbb{R}^{n\times 2}$, where $n \in 2\mathbb{N}$ is an even number, such that their spark equals $n/2$.

Instances for the ratio of two norms and for the pseudonorm ℓ_0

For each dimension $n \in \{3, \dots, 10, 20, \dots, 90, 100\}$ and each angle $\theta \in \frac{\pi}{2}\{\frac{1}{10}, \dots, \frac{5}{10}\}$, we have generated cones

$$K_{n,\theta} = \mathrm{cone}\{u^1, \dots, u^{2\lceil n/2\rceil}\} , \tag{5.14}$$

with an even number of generator $u^i \in \mathbb{R}^n$, where

$$\begin{gathered} \forall k \in \{1, \dots, n\} \ , \ u_1^k = \cos(\theta) \ , \\ \forall j \in \{1, \dots, \lceil n/2\rceil\}, \forall i \in \{2, \dots, n\} \ , \ \begin{cases} u_i^{2j-1} = & \varepsilon_i^j \frac{1}{\sqrt{n-1}} \sin(\theta) \ , \\ u_i^{2j} = & -\varepsilon_i^j \frac{1}{\sqrt{n-1}} \sin(\theta) \ . \end{cases} \end{gathered}$$

such that the vectors $\varepsilon^j \in \{-1, 1\}^{n-1}$ are $\lceil n/2\rceil$ distinct vectors.

It is easy to check that, for all dimension $n \in \mathbb{N}$ and all angles $\theta \in]0, \pi/2[$, the vector $(1, 0, \dots, 0)^T \in \mathbb{R}^n$ is included in $K_{n,\theta}$. Thus, we have generated cones centered on the first axis of $\mathbb{R}^n$ and we control their 'tightness' to the first axis thanks to the angle θ.

Instances for spark of a matrix

For each dimension $n \in \{6, 12, 14, \dots, 30, 50, 80, 100, 130, 150, 180, 200\}$, we have generated square matrices $A \in \mathbb{R}^{n \times n}$ such that their spark is equal to $s := n/2$. To do so, we used the following algorithm.

1. Randomly choose $s-1$ vectors $A_i \in \mathbb{R}^n$.
2. Randomly choose $s-1$ real numbers $\mu_i \in \mathbb{R}$.
3. Compute the vector $A_s = \sum_{i=1}^{s-1} \mu_i A_i$.
4. Randomly choose $n-s$ vectors $A_h \in \mathbb{R}^n$.
5. Set the matrix $A = \big(A_1, \dots, A_n\big)$.
6. Shuffle the columns of the matrix A.

5.3.2 Ratio of ℓ_1 over ℓ_2

Here, we measure the time and the number of iterations needed to attain the optimal solution for different dimensions and angle θ of the feasible cone defined in §5.3.1. We have run the *E*-Capra *cutting plane method* from Definition 5.11 using the nonlinear optimization solver *Ipopt* for the nonconvex subproblem of Step 3 with the following parameters: `tol = 1E-8`, `max_iter = 3000`. It is worth noticing that the results display a part of randomness as the subproblem Step 3 is nonconvex and we use a random starting point to solve it.

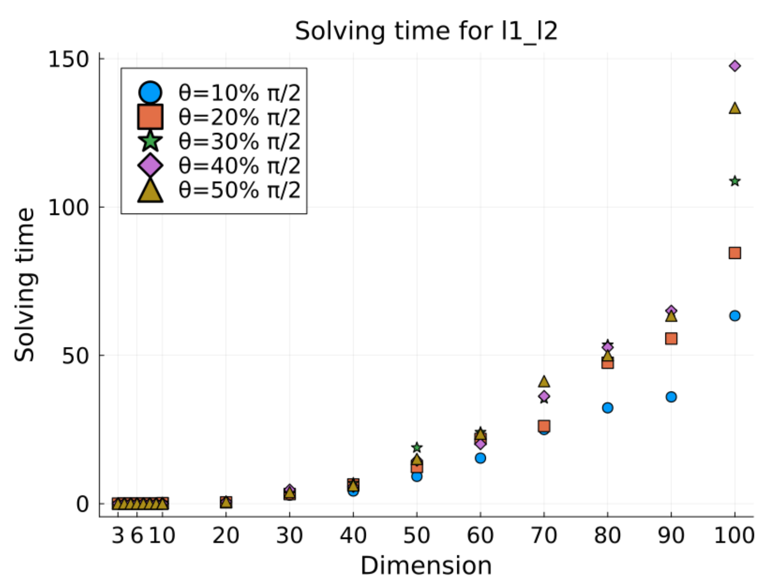

Figure 5.1: Solving time for the minimization of the ratio of two norms

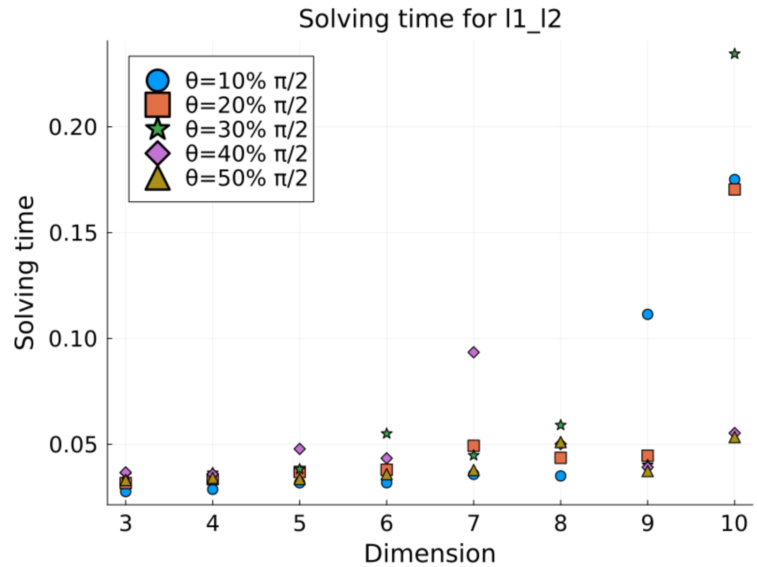

Figure 5.2: Solving time for the minimization of the ratio of two norms in low dimension

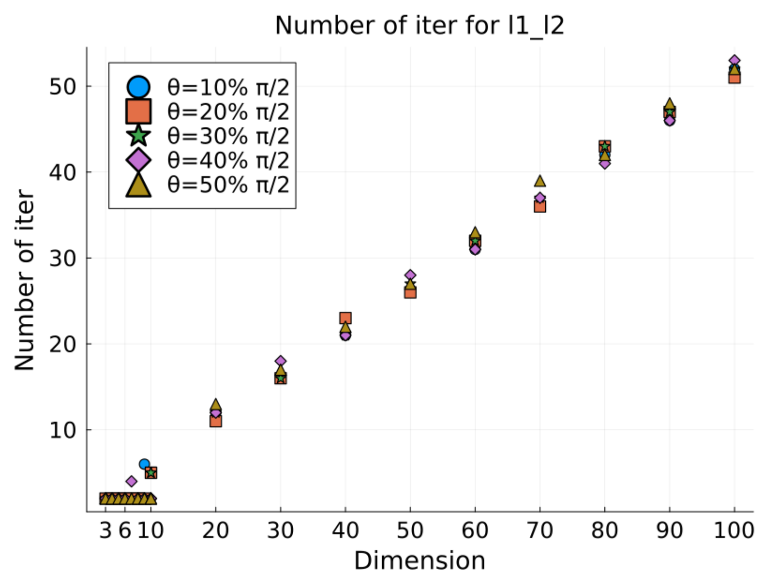

Figure 5.3: Number of iterations for the minimization of the ratio of two norms

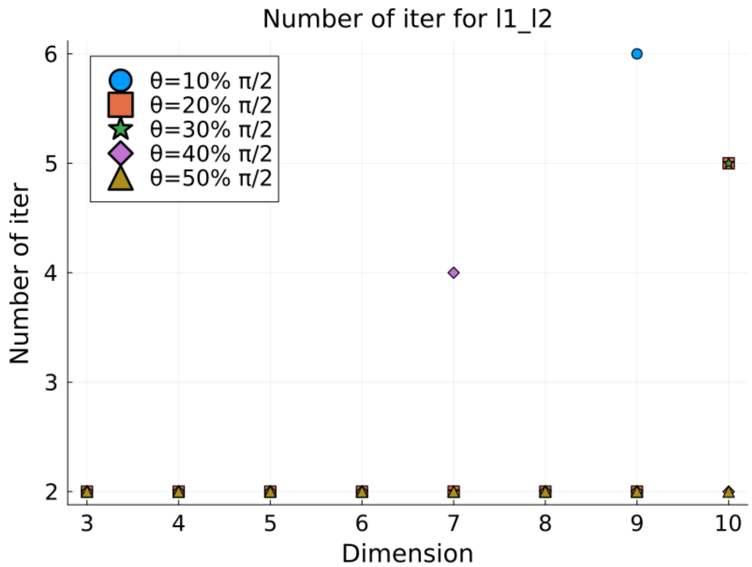

Figure 5.4: Number of iterations for the minimization of the ratio of two norms in low dimension

We see in Figure 5.3.2 that the number of iterations to solve the instances in low dimension does not depend on the angle nor on the dimension. Beyond the dimension 10 in Figure 5.3.2, the number of iterations seems to grow linearly with the dimension and not to depend on the angle θ.

We see in Figure 5.3.2 and Figure 5.3.2 that the time for solving the instances grows exponentially with the dimension and increases faster for greater angle θ.

5.3.3 Counting pseudonorm ℓ_0

Here, for a given time budget of 200 seconds, we measure the relative gap $\frac{\text{best found value}-1}{\text{dim}}$ between the best value found by the *modified E-*Capra *cutting plane method* from Definition 5.14 and the optimal value of instances. We use the nonlinear optimization solver *Ipopt* for the nonconvex subproblem of Step 4 with the following parameters: `tol = 1E-8`, `max_iter = 3000`.

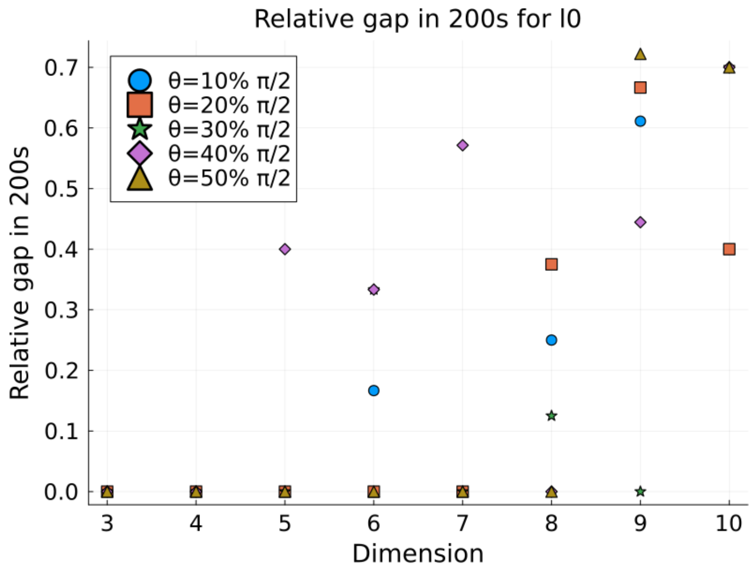

Figure 5.5: Relative gap for the minimization of ℓ_0 with a time budget = 200s

We see in Figure 5.3.3 that from dimension 3 to 5 the modified E-Capra cutting plane method generally attains optimality, from dimension 6 to 8 attains a relative gap smaller than 40%, and for dimension 9 and 10 a relative gap greater than 40%.

5.3.4 Spark of a matrix

Here, we test *modified E-*Capra *cutting plane method* from Definition 5.14 to compute the spark of the matrices described in §5.3.1. We compare the time needed, on the one hand, to compute the spark of the instances by the modified E-Capra cutting plane method to on the time needed by a 'bruteforce' method — where we select all n choose k columns of the matrix for $k = 1, \ldots, n$ and stopping at the first k such that we find k linearly dependent columns. We also look at the time needed to compute the spark for a matrix by the modified E-Capra cutting plane method in higher dimension.

We use the nonlinear optimization solver *Ipopt* for the nonconvex subproblem of Step 4 of the modified E-Capra cutting plane method with the following parameters: `tol = 1E-8`, `max_iter = 3000`.

Dim.	6	12	14	16	18	20	22
BF time (s)	2.93E-4	2.20E-2	6.80E-2	2.79E-1	1.60E0	8.10E0	3.84E1
CP time (s)	8.40E-2	1.59E-1	1.83E-1	3.06E-1	2.59E-1	3.66E-1	3.89E-1

Table 5.2: Solving time comparison the between brute force (BF) and the cutting plane (CP) methods for spark

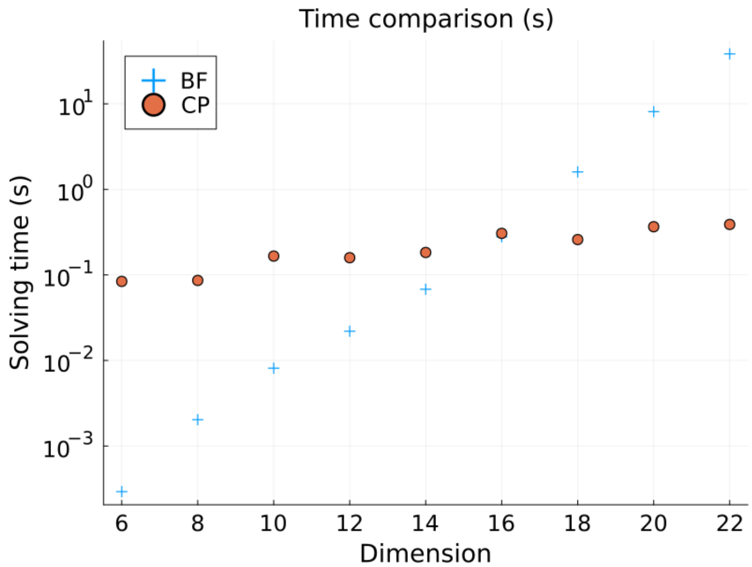

Figure 5.6: Solving time for spark comparison bewteen brute force and E-Capra cutting plane method

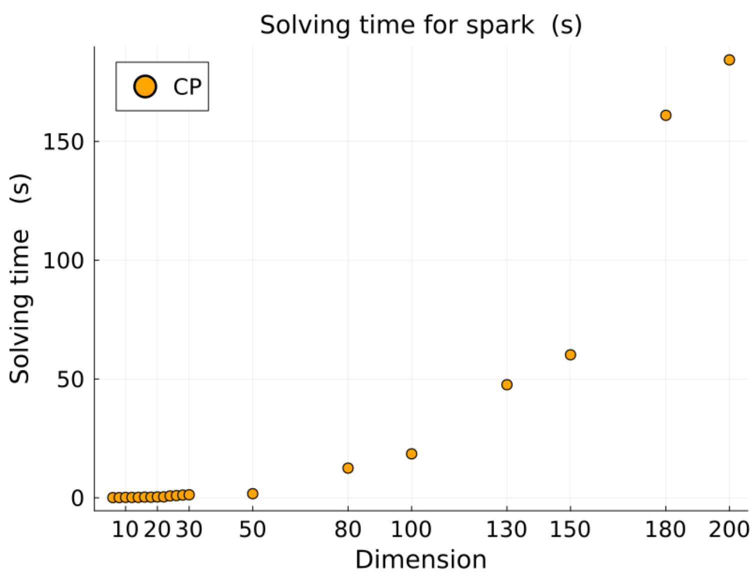

Figure 5.7: Solving time for spark

In Table 5.3.4 and in Figure 5.3.4, we see that the modified E-Capra cutting plane if 100 times slower than the bruteforce method at dimension 6 and 100 times faster than bruteforce method at dimension 22. The tipping point where the modified E-Capra cutting plane becomes better is at dimension 16.

In Figure 5.3.4, we see that the time needed to compute the spark seems to grow exponentially with dimension.

5.3.5 Discussion

We first compare the numerical results for the three problems, then, we provide perspective for further numerical tests.

Comparison of the results for the three problems

If we were to rank the three considered problems by order of difficulty, we would come up with this ranking according to the numerical results from § 5.3.2, § 5.3.3 and § 5.3.4 :

1. the minimization of the ratio of the ℓ_1 norm over the ℓ_2 norm;
2. the computation of the spark of a square matrix;
3. the minimization of the ℓ_0 pseudonorm in a blunt convex cone.

The fact that the minimization of the ratio of the ℓ_1 norm over the ℓ_2 norm is the easiest problem of the three comes as no surprise, as it satisfies the assumptions of the convergence theorem of the cutting plane method [28, Theorem 9.1.1], while the two other problems do not (as the pseudonorm ℓ_0 is not continuous). However, it is more surprising to see that the minimization of the ℓ_0 pseudonorm in a blunt convex cone is harder than the computation of the spark of a square matrix. Indeed, the computation of the spark of a matrix is not a E-CAPRA convex problem, as we can see in Table 5.1.4. Maybe this observation has a link with the fact that the constraints of minimization problem for the computation of the spark is given by an equation $Ax = 0$, while the constraints of the minimization of the ℓ_0 pseudonorm in a blunt convex cone is given by a convex combination $x = \sum_{i=1}^{r} \lambda_i u_i \ , \ \sum_{i=1}^{r} \lambda_i = 1 \ , \ \lambda_i \geq 0$, of the generators $u_1, \dots u_r$ of the cone.

It is worth noting that using the *modified E-*CAPRA *cutting plane method* from Definition 4 has been decisive for the convergence of the tests in the case of the minimization of ℓ_0 and for the computation of sparks. With the *E-*CAPRA *cutting plane method* from Definition 5.11, *Ipopt* was unable to find any feasible point of the nonconvex subproblem at Step 3.

Perspective for future numerical tests

- Points of a polyhedral cone K can be expressed, in a primal form, as a convex combination of generators of the cone K and, in a dual form, as an intersection of hyperplanes $Ax \leq 0$ according to Minkowski–Wey's theorem [10, Theorem 3.11]. As a consequence, *the minimization of the ℓ_0 pseudonorm in a blunt convex cone* should be tested with the cone constraints expressed as an intersection of hyperplanes of the form $Ax \leq 0$.

- For the minimization of the ratio of two norms, the minimization problem $\inf_{x \neq 0} \ell_1(x)/\ell_2(x)$ was a toy example. Tests should be conducted with ratio of more relevant norms such that the minimization problem $\inf_{x \neq 0} \|Ax\| \,/\, \|x\|$, which is used to compute the singular values of a the matrix A.

- A branch-and-cut method with E-CAPRA cuts for the minimization of the ℓ_0 pseudonorm in a blunt convex cone and for the computation of the spark of a matrix should be tested and compared to the cutting plane method.

Conclusion

In this report, we have explored what insights the perturbation duality scheme has to offer on PILP, in the first part, and we have tested the efficiency of a cutting plane method for E-Capra convex problems (while highlighting other global optimization methods and a possible definition of proximal operator for OSL conjugacy), in the second part.

In Chapter 2, we have understood that Jeroslow's subadditive dual problem for a PILP corresponds to the dual problem we obtained by perturbation-duality with the subadditive coupling $c_{\mathcal{S}}$, but by restricting the dual space of subadditive functions to subadditive functions which coincides with the value function at the anchor and that are $c_{\mathcal{S}}$-uppersubgradients. Perspectives of new dual problems for PILP involving subsets of the Chvátal functions have been given in §2.4 and in Chapter 3. In particular, we obtain a strong duality result between PILP and the dual problems by perturbing only a part of the right-hand side of the constraints and by coupling the perturbation with affine Chvátal functions as in Definition 3.5.

In Chapter 5, we have observed promising results for the abstract cutting plane applied to E-Capra convex problems, especially for the ratio of the norm $\ell_1$over the norm ℓ_2 and the computation of the spark of square matrix. The first results suggest that global optimization methods could be efficient for solving E-Capra convex problems. For the ratio of two norms, other ratio should be tested, such as $\|A\cdot\| \,/\, \|\cdot\|$, whose minimization is useful to find the singular values of the matrix A. For E-Capra convex problems with the pseudonorm ℓ_0, which has a natural combinatorial structure, branch-and-bound and branch-and-cuts methods could be tested.

Part III

Appendix

Appendix A

Generalities

A.1 Convex analysis conventions and definitions

We remind several conventions and definitions of convex analysis that are used throughout this report.

A.1.1 $\inf_{\varnothing}, \sup_{\varnothing}$ conventions

We use the following convention:

$$\inf_{\varnothing} \cdot = +\infty \ , \tag{A.1a}$$

$$\sup_{\varnothing} \cdot = -\infty \ . \tag{A.1b}$$

A.1.2 Moreau additions

The Moreau additions extend the usual addition to $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$.

Definition A.1. *We call* lower addition *the operator which extends the usual addition by*

$$(+\infty) \underset{\cdot}{+} (-\infty) = (-\infty) \underset{\cdot}{+} (+\infty) = -\infty \ . \tag{A.2a}$$

We call upper addition *the operator which extends the usual addition by*

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty \ . \tag{A.2b}$$

A.1.3 Properness of functions

Definition A.2. *Let $f : \mathcal{W} \to \overline{\mathbb{R}}$. Then*

$$\mathrm{dom} f = \big\{ w \in \mathcal{W} \mid f(w) < +\infty \big\} \ , \tag{A.3}$$

is called its effective domain*.*

Remark A.3. *Let $f : \mathcal{W} \to \overline{\mathbb{R}}$. If $f|_{\mathrm{dom} f} = -\infty$ we say f is a* valley function*.*

A.2 Background on J. J. Moreau lower and upper additions

When we manipulate functions with values in $\overline{\mathbb{R}} = [-\infty, +\infty]$, we adopt the following Moreau *lower addition* or *upper addition*, depending on whether we deal with sup or inf operations. We follow [27]. In the sequel, u, v and w are any elements of $\overline{\mathbb{R}}$.

Moreau lower addition

The Moreau *lower addition* extends the usual addition with

$$(+\infty) \underset{\cdot}{+} (-\infty) = (-\infty) \underset{\cdot}{+} (+\infty) = -\infty \; . \tag{A.4a}$$

With the *lower addition*, $(\overline{\mathbb{R}}, \underset{\cdot}{+})$ is a convex cone, with $\underset{\cdot}{+}$ commutative and associative. The lower addition displays the following properties:

$$u \leq u' \; , \;\; v \leq v' \Rightarrow u \underset{\cdot}{+} v \leq u' \underset{\cdot}{+} v' \; , \tag{A.4b}$$

$$(-u) \underset{\cdot}{+} (-v) \leq -(u \underset{\cdot}{+} v) \; , \tag{A.4c}$$

$$(-u) \underset{\cdot}{+} u \leq 0 \; , \tag{A.4d}$$

$$\sup_{a \in \mathbb{A}} f(a) \underset{\cdot}{+} \sup_{b \in \mathbb{B}} g(b) = \sup_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \underset{\cdot}{+} g(b) \big) \; , \tag{A.4e}$$

$$\inf_{a \in \mathbb{A}} f(a) \underset{\cdot}{+} \inf_{b \in \mathbb{B}} g(b) \leq \inf_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \underset{\cdot}{+} g(b) \big) \; , \tag{A.4f}$$

$$t < +\infty \Rightarrow \inf_{a \in \mathbb{A}} f(a) \underset{\cdot}{+} t = \inf_{a \in \mathbb{A}} \big(f(a) \underset{\cdot}{+} t \big) \; . \tag{A.4g}$$

Moreau upper addition

The Moreau *upper addition* extends the usual addition with

$$(+\infty) \dot{+} (-\infty) = (-\infty) \dot{+} (+\infty) = +\infty \; . \tag{A.5a}$$

With the *upper addition*, $(\overline{\mathbb{R}}, \dot{+})$ is a convex cone, with $\dot{+}$ commutative and associative. The upper addition displays the following properties:

$$u \leq u' \; , \;\; v \leq v' \Rightarrow u \dot{+} v \leq u' \dot{+} v' \; , \tag{A.5b}$$

$$(-u) \dot{+} (-v) \geq -(u \dot{+} v) \; , \tag{A.5c}$$

$$(-u) \dot{+} u \geq 0 \; , \tag{A.5d}$$

$$\inf_{a \in \mathbb{A}} f(a) \dot{+} \inf_{b \in \mathbb{B}} g(b) = \inf_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \dot{+} g(b) \big) \; , \tag{A.5e}$$

$$\sup_{a \in \mathbb{A}} f(a) \dot{+} \sup_{b \in \mathbb{B}} g(b) \geq \sup_{a \in \mathbb{A}, b \in \mathbb{B}} \big(f(a) \dot{+} g(b) \big) \; , \tag{A.5f}$$

$$-\infty < t \Rightarrow \sup_{a \in \mathbb{A}} f(a) \dot{+} t = \sup_{a \in \mathbb{A}} \big(f(a) \dot{+} t \big) \; . \tag{A.5g}$$

Joint properties of the Moreau lower and upper addition

We obviously have that

$$u \mathbin{\underset{\cdot}{+}} v \leq u \mathbin{\dot{+}} v \; . \tag{A.6a}$$

The Moreau lower and upper additions are related by

$$-\,(u \mathbin{\dot{+}} v) = (-u) \mathbin{\underset{\cdot}{+}} (-v) \; , \;\; -(u \mathbin{\underset{\cdot}{+}} v) = (-u) \mathbin{\dot{+}} (-v) \; . \tag{A.6b}$$

They satisfy the inequality

$$(u \mathbin{\dot{+}} v) \mathbin{\underset{\cdot}{+}} w \leq u \mathbin{\dot{+}} (v \mathbin{\underset{\cdot}{+}} w) \; . \tag{A.6c}$$

with

$$(u \mathbin{\dot{+}} v) \mathbin{\underset{\cdot}{+}} w < u \mathbin{\dot{+}} (v \mathbin{\underset{\cdot}{+}} w) \iff \begin{cases} u = +\infty \;\text{ and }\; w = -\infty \; , \\ \;\text{or} \\ u = -\infty \;\text{ and }\; w = +\infty \;\text{ and }\; -\infty < v < +\infty \; . \end{cases} \tag{A.6d}$$

Finally, we have that

$$u \mathbin{\underset{\cdot}{+}} (-v) \leq 0 \iff u \leq v \iff 0 \leq v \mathbin{\dot{+}} (-u) \; , \tag{A.6e}$$

$$u \mathbin{\underset{\cdot}{+}} (-v) \leq w \iff u \leq v \mathbin{\dot{+}} w \iff u \mathbin{\underset{\cdot}{+}} (-w) \leq v \; , \tag{A.6f}$$

$$w \leq v \mathbin{\dot{+}} (-u) \iff u \mathbin{\underset{\cdot}{+}} w \leq v \iff u \leq v \mathbin{\dot{+}} (-w) \; . \tag{A.6g}$$

A.3 Background on Fenchel-Moreau conjugacy with respect to a coupling

Let be given two sets $\mathcal{U}$ and $\mathcal{V}$. Consider a *coupling* function $c : \mathcal{U} \times \mathcal{V} \to [-\infty, +\infty]$. We also use the notation $\mathcal{U} \overset{c}{\leftrightarrow} \mathcal{V}$ for a coupling, so that

$$\mathcal{U} \overset{c}{\leftrightarrow} \mathcal{V} \iff c : \mathcal{U} \times \mathcal{V} \to [-\infty, +\infty] \; . \tag{A.7}$$

Definition A.4. *The* Fenchel-Moreau conjugate *of a function $f : \mathcal{U} \to [-\infty, +\infty]$, with respect to the coupling c in* (A.7)*, is the function $f^c : \mathcal{V} \to [-\infty, +\infty]$ defined by*

$$f^c(v) = \sup_{u \in \mathcal{U}} \Big(c(u, v) \mathbin{\underset{\cdot}{+}} \big(-f(u) \big) \Big) \; , \;\; \forall v \in \mathcal{V} \; . \tag{A.8}$$

We associate with the coupling c the coupling $c' : \mathcal{V} \times \mathcal{U} \to [-\infty, +\infty]$ defined by $c'(v, u) = c(u, v)$. The Fenchel-Moreau biconjugate *is the function $f^{cc'} : \mathcal{U} \to [-\infty, +\infty]$ defined by*

$$f^{cc'}(u) = (f^c)^{c'}(u) = \sup_{v \in \mathcal{V}} \Big(c(u, v) \mathbin{\underset{\cdot}{+}} \big(-f^c(v) \big) \Big) \; , \;\; \forall u \in \mathcal{U} \; . \tag{A.9}$$

The following property is well known.

Proposition A.5. *For any function $f : \mathcal{U} \to [-\infty, +\infty]$, we have that*

$$f^{cc'}(u) \leq f(u) \ . \tag{A.10}$$

Proof. We prove (A.10) as follows.

$$\begin{aligned}
f^{cc'}(u) \dotplus \big(-f(u)\big) &= \sup_{v \in \mathcal{V}} \Big(c(u,v) \dotplus \big(-f^c(v)\big)\Big) \dotplus \big(-f(u)\big) && \text{(by (A.9) and (A.4e))} \\
&= \sup_{v \in \mathcal{V}} \bigg(\Big(c(u,v) \dotplus \big(-f^c(v)\big)\Big) \dotplus \big(-f(u)\big)\bigg) && \text{(by (A.4e))} \\
&= \sup_{v \in \mathcal{V}} \bigg(c(u,v) \dotplus \big(-f^c(v)\big) \dotplus \big(-f(u)\big)\bigg) && \text{(by associativity of } \dotplus) \\
&= \sup_{v \in \mathcal{V}} \bigg(c(u,v) \dotplus \big(-f(u)\big) \dotplus \big(-f^c(v)\big)\bigg) && \text{(by commutativity of } \dotplus) \\
&\leq \sup_{v \in \mathcal{V}} \bigg(\sup_{u \in \mathcal{U}} \Big(c(u,v) \dotplus \big(-f(u)\big)\Big) \dotplus \big(-f^c(v)\big)\bigg) && \text{(by (A.4b))} \\
&= \sup_{v \in \mathcal{V}} \Big(f^c(v) \dotplus \big(-f^c(v)\big)\Big) && \text{(by (A.8))} \\
&\leq 0 \ . && \text{(by (A.4d))}
\end{aligned}$$

We have obtained that $f^{cc'}(u) \dotplus \big(-f(u)\big) \leq 0$. Now, using (A.6e), we obtain (A.10). This ends the proof. □

Definition A.6.

- *Let $u \in \mathcal{U}$. We say that the function $f : \mathcal{U} \to \overline{\mathbb{R}}$ is c-convex in u, if*

$$f^{cc'}(u) = f(u) \ . \tag{A.12}$$

- *We say that the function f is c-convex on $\mathcal{U}' \subset \mathcal{U}$ if f is c-convex in u' , $\forall u' \in \mathcal{U}'$.*
- *We say that the function f is c-convex if f is c-convex on $\mathcal{U}$.*

The following properties are easy to establish.

Proposition A.7. *For any family $\{f_u\}_{u \in \mathbb{U}}$ of functions $f_u : \mathcal{U} \to [-\infty, +\infty]$, we have that*

$$\big(\inf_{u \in \mathbb{U}} f_u\big)^c(u) = \sup_{u \in \mathbb{U}} f_u^c(u) \tag{A.13a}$$

$$-\big(\inf_{u \in \mathbb{U}} f_u\big)^c(u) = \inf_{u \in \mathbb{U}} \big(-f_u^c(u)\big) \ . \tag{A.13b}$$

We also define the c-uppersubgdifferential.

Definition A.8. *Let $f : \mathcal{U} \to \overline{\mathbb{R}}$. Then its c-uppersubdifferential $\partial^c f(u) \subset \mathcal{V}$ at the point $u \in \mathcal{U}$, is defined by*

$$v \in \partial^c \ f(u) \iff f(u) = c(u,v) \dotplus (-f^c(v)) \ . \tag{A.14}$$

Bibliography

[1] E. J. Balder. An extension of duality-stability relations to nonconvex optimization problems. *SIAM Journal on Control and Optimization*, 15(2):329–343, 1977.

[2] Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, New York, second edition, 2017.

[3] Amir Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[4] Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[5] Charles E Blair and Robert G Jeroslow. The value function of an integer program. *Mathematical programming*, 23(1):237–273, 1982.

[6] Mikhail A. Bragin. Survey on Lagrangian relaxation for MILP: importance, challenges, historical review, recent advancements, and opportunities. *Annals of Operations Research*, July 2023.

[7] Jean-Philippe Chancelier and Michel De Lara. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.

[8] Jean-Philippe Chancelier and Michel De Lara. Capra-convexity, convex factorization and variational formulations for the l_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022.

[9] Jean-Philippe Chancelier and Michel De Lara. Orthant-strictly monotonic norms, generalized top-k and k-support norms and the l0 pseudonorm. *Journal of Convex Analysis (to appear)*, 2022.

[10] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. *Integer Programming*. Springer Cham, October 2014.

[11] Michel De Lara. Duality between Lagrangians and Rockafellians. *Journal of Convex Analysis*, 30(3):887–896, 2023.

[12] Michael Fekete. Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Mathematische Zeitschrift*, 17(1):228–249, 1923.

[13] Adrien Le Franc, Jean-Philippe Chancelier, and Michel De Lara. The capra-subdifferential of the l0 pseudonorm. *Optimization*, pages 1–23, 2022. Accepted for publication.

[14] Arthur M. Geoffrion. Lagrangean relaxation and its uses in integer programming. *Math. Programming Study*, 2:82–114, 1974.

[15] Jean Charles Gilbert. Fragments d'Optimisation Différentiable - Théories et Algorithmes. Lecture, March 2021.

[16] P. C. Gilmore and R. E. Gomory. The theory and computation of knapsack functions. *Operations Research*, 14(6):1045–1074, 1966.

[17] M Güzelsoy, Ted K Ralphs, and J Cochran. Integer programming duality. In *Encyclopedia of Operations Research and Management Science*, pages 1–13. Wiley Hoboken, NJ, USA, 2010.

[18] R. Jeroslow. Cutting-plane theory: Algebraic methods. *Discrete Mathematics*, 23(2):121–150, 1978.

[19] Robert G Jeroslow. Minimal inequalities. *Mathematical programming*, 17(1):1–15, 1979.

[20] Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, 01 2004.

[21] Frank R. Kschischang. The subadditivity lemma, November 2009. `https://www.comm.utoronto.ca/~frank/notes/subadditivity.pdf` Accessed: 17-9-2023.

[22] Adrien Le Franc. *Subdifferentiability in convex and stochastic optimization applied to renewable power systems*. Theses, École des Ponts ParisTech, December 2021.

[23] S Martello and P Toth. *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, 1990.

[24] J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In Schaible S. Hadjisavvas N., Komlósi S., editor, *Handbook of Generalized Convexity and Generalized Monotonicity. Nonconvex Optimization and Its Applications*, volume 76, pages 237–292. Springer-Verlag, 2005.

[25] J. J. Moreau. Fonctionnelles convexes. *Séminaire Jean Leray*, 2:1–108, 1966-1967.

[26] Jean-Jacques Moreau. Proximité et dualité dans un espace Hilbertien. *Bulletin de la Société mathématique de France*, 93:273–299, 1965.

[27] Jean Jacques Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math. Pures Appl. (9)*, 49:109–154, 1970.

[28] Diethard Pallaschke and Stefan Rolewicz. *Foundations of mathematical optimization*, volume 388 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997.

[29] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

[30] R. Tyrrell Rockafellar. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1974.

[31] Johannes O. Royset and Roger J-B Wets. *An Optimization Primer*. Springer International Publishing, Cham, 2021.

[32] Alexander Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.

[33] A. Schrijver. *Theory of Linear and Integer Programming*. Wiley-Interscience Series in Discrete Mathematics and Optimization, New York. Wiley, 1986.

[34] Ivan Singer. *Abstract Convex Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1997.

[35] Ivan Singer. *Duality for Nonconvex Approximation and Optimization*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2006.

[36] Andreas M. Tillmann and Marc E. Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2014.