École des Ponts
ParisTech

THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

Multistage stochastic optimization and polyhedral geometry

Optimisation stochastique multi-étapes et géométrie polyédrale

École doctorale Mathématiques, Sciences et Technologies de l'Information et de la Communication

Mathématiques appliquées

Thèse préparée au laboratoire CERMICS, École des Ponts
au sein de l'équipe Optimisation
et au CMAP, École polytechnique au sein de l'équipe INRIA Tropical

Thèse soutenue le 14 décembre 2022, par
Maël FORCIER

Composition du jury:

Jesús DE LOERA Pr, University of California, Davis	*Rapporteur*
Wolfram WIESEMANN Pr, Imperial College of London	*Rapporteur*
Francesca MAGGIONI Pr, Università degli Studi di Bergamo	*Examinatrice*
Nadia OUDJANE Pr, Électricité de France	*Examinatrice*
Lionel POURNIN Pr, Université Paris 13	*Examinateur*
Jean-Philippe CHANCELIER Pr, École des Ponts	*Directeur de thèse*
Stéphane GAUBERT Pr, École polytechnique - INRIA	*Co-directeur de thèse*
Vincent LECLÈRE Dr, École des Ponts	*Encadrant de thèse*

Résumé

Dans cette thèse, nous utilisons les outils de la géométrie polyédrale pour appréhender la structure de problèmes stochastiques. Plus précisément, lorsque les variables aléatoires de problèmes stochastiques linéaires multiétapes (MSLP) peuvent être remplacées par des variables aléatoires à support fini sans changer les fonctions valeurs, on parle de discrétisation exacte. On qualifie une discrétisation exacte de locale si elle s'applique à un point particulier de l'espace d'état, d'uniforme si elle ne dépend pas de l'état et d'universelle si elle est indépendente de la distribution.

Notre but est de donner des conditions pour obtenir des discrétisations exactes universelles, locale ou uniforme.

Grâce à la notion d'éventail normal, nous établissons une discrétisation exacte locale et universelle pour les problèmes stochastiques linéaires à 2 étapes (2SLP), qui permet ensuite d'obtenir une discrétisation exacte, uniforme et universelle, à l'aide de l'équivalence normale des fibres du polyèdre couplant sur les cellules d'un certain complexe polyédral appelé *chamber complex*. En construisant par programmation dynamique, des complexes polyédraux universels, où la fonction des coûts futurs espérés est affine par morceaux, nous prouvons une discrétisation exacte uniforme et universelle pour les MSLP avec distribution de coût générale. De plus, nous donnons une interprétation duale à l'aide d'une généralisation du polyèdre de fibres pondéré adapté au 2SLP que l'on étend aux MSLP au moyen de polyèdres de fibres imbriqués. Ces discrétisations nous permettent alors de déduire des résultats de complexité pour les MSLP en montrant qu'ils deviennent résolubles en temps polynomial pour toute distribution régulière, lorsque certains paramètres sont fixés.

Nous nous intéressons ensuite aux 2SLP dont la matrice et le second membre des contraintes ont des distributions générales. Grâce à une discrétisation exacte et locale, nous étendons la portée des méthodes de partitions adaptatives (APM) en donnant un oracle géométrique pour obtenir une partition adaptée, c'est à dire fournissant une discrétisation exacte locale. Nous donnons également une condition nécessaire et suffisante pour la correction des algorithmes APM, ainsi que des bornes de complexité.

Enfin, nous introduisons une classe d'algorithmes, appelée Programmation Dynamique par Suivi de Trajectoire, *Trajectory Following Dynamic Programming* en anglais, qui affine successivement des approximations des fonctions des coûts futurs espérés d'un problème stochastique multi-étapes avec des variables aléatoires indépendantes. Ce cadre algorithmique englobe la plupart des variantes de l'algorithme *Stochastic Dual Dynamic Programming* (SDDP). En supposant le caractère Lipschitz de la fonction valeur, nous donnons une nouvelle preuve de convergence et de complexité qui autorise les variables aléatoires avec des supports infinis. En particulier, nous en déduisons des nouveaux résultats de complexité pour plusieurs algorithmes.

Abstract

In this manuscript we study how the tools from polyhedral geometry enlighten the structure of multistage stochastic programming. More precisely, when the arbitrary random variables of a Multistage Stochastic Linear Problem (MSLP) can be replaced by finitely supported random variables without changing the value functions, we say that there exists an exact quantization. We call an exact quantization local if it applies at a particular state, uniform if it does not depend on the state and universal if it is independent of the noise distribution.

Our aim is to provide conditions to obtain local or uniform, universal exact quantization of MSLP, and algorithms exploiting these conditions.

Through the notion of normal fans, we show a local and uniform exact quantization for 2-stage linear problems (2SLP) whose cost has a non necessarily finite distribution, which then provides a universal and uniform exact quantization thanks to the property of normal equivalence on a chamber complex. By constructing, through dynamic programming, universal chamber complexes, where the expected cost-to-go function is piecewise affine, we prove a uniform and universal exact quantization for MSLP with general cost distributions. Further, we give a dual interpretation of this result by defining new objects which extend of the notion of fiber polytope to general distributions. These quantizations allow us to derive new complexity results for MSLP showing that with fixed parameters, MSLP becomes polynomial for every regular distribution.

We then focus on 2SLP with generally distributed matrix and right-hand side constraints. Thanks to a local exact quantization, we extend the scope of Adaptive Partition-based methods (APM) by providing a geometric oracle to obtain an adapted partition. We also provide necessary and sufficient conditions of correcteness of APM, as well as convergence speed result.

Finally, we introduce a class of algorithms, called Trajectory Following Dynamic Programming (TFDP) algorithms, that iteratively refines approximations of expected cost-to-go functions of multistage stochastic problems with independent random variables. This framework encompasses most variants of the Stochastic Dual Dynamic Programming algorithm. Leveraging a Lipschitz assumption on the expected cost-to-go functions, we provide a new convergence and complexity proof that allows random variables with non-finitely supported distributions. In particular, this leads to new complexity results for numerous known algorithms.

Remerciements

Mes premiers et plus vifs remerciements vont bien sûr à Vincent Leclère et Stéphane Gaubert, pour tout leur suivi, leur aide et la culture mathématique qu'ils m'ont apportés. Merci Vincent d'avoir été là dès que j'en avais besoin pendant toute la thèse, d'avoir toujours été à l'écoute, d'avoir cru en moi et de m'avoir toujours poussé à m'améliorer et enfin merci pour ton aide et ta rigueur qui m'ont tant de fois remotivé lorsqu'il fallait rédiger des articles. Merci à toi Stéphane de m'avoir redonné à tant de reprises le goût de la beauté des mathématiques, que ce soit pendant tes cours qui m'ont donné l'envie de faire une thèse ou pendant toutes ces discussions intéressantes et passionnées, et merci pour tout ce que tu m'as appris durant ces 3 ans.

Je voudrais également remercier Jesús De Loera et Wolfram Wiesemann d'avoir accepté d'être les rapporteurs de ma thèse et Francesca Maggioni, Nadia Oudjane et Lionel Pournin d'avoir bien voulu être examinateurs dans mon jury de thèse.

L'environnement de travail du laboratoire CERMICS de l'École des Ponts a été essentiel pour moi durant cette thèse. Un grand merci à Jean-Philippe Chancelier pour son aide et sa disponibilité dès que j'avais une question mathématique ou informatique. Merci également à l'ensemble des chercheurs du CERMICS, notamment Frédéric Meunier et Axel Parmentier, qui avec Vincent, m'ont permis de découvrir l'enseignement mais aussi Pierre Carpentier et Michel de Lara. Je remercie aussi chaleureusement Isabelle Simunic et Stéphanie Bonnel qui ont été d'une efficacité incroyable dans le traitement de toutes mes démarches administratives.

J'ai aussi passé une partie de ma thèse au CMAP que je voulais remercier pour son accueil et en particulier Marianne Akian, Xavier Allamigeon et Hanadi Dib de l'équipe tropical. Dû aux circonstances particulières, je n'ai pas pu aller aux labos tous les jours et je souhaitais donc remercier toutes les personnes qui m'ont accueilli chez elles ou dans leurs bureaux pour télétravailler pendant les déconfinements et couvre-feus successifs.

Un immense merci à l'ensemble des doctorants du CERMICS et à la bonne humeur et ambiance qu'ils créaient à chaque fois que j'allais au labo. Ce climat de travail m'a énormément aidé et remotivé tout au long de ces trois ans. Merci à Cyrille, Benoît, Sébastien, Ézechiel, Thomas et Thomas, Adrien, Clément, Michel, William, Guillaume, Zoé, Louis, Emanuele, Dylan, Coco, Julien, Nerea, Eduardo, Roberta, Hélène, Léo, Henri, François, à ceux du troisième et à tous ceux que j'ai oubliés.

Enfin, merci infiniment à Alice, mon équipe, dont le soutien quotidien a été indispensable, et à ma famille et mes amis qui m'ont toujours supporté depuis le début.

Contents

Notations

As a general guideline ***bold*** letters denote random variables, normal scripts their realizations. Capital letters denote matrices or sets, calligraphic (e.g. $\mathcal{N}$) denote collections of sets. Depending on the context, overline are either used for subcollections of maximal elements (as in $\overline{\mathcal{C}}$) or for upper-approximation (as in $\overline{V}$). Similarly, underline are either used for subcollections of minimal elements or lower-approximation. Hats (as in $\hat{V}$) represent functions parametrized by ξ, and their counterpart without hat (as in V) being their expectation.

- The (tropical) indicator function $\mathbb{I}_{\mathtt{P}}$ takes value 0 if P is true and $+\infty$ otherwise.
- The indicator function $\mathbb{1}_{\mathtt{P}}$ takes value 1 if P is true and 0 otherwise.
- $[k] := \{1, \dots, k\}$.
- $\sharp E$ is the cardinal of a set E.
- $\mathrm{ri}(E)$ is the relative interior of a set E
- A_I is the submatrix of a matrix A, composed of the rows of indices $i \in I$.
- $A_{\cdot,J}$ is the submatrix of a matrix A composed of the columns of indices $j \in J$.
- $\mathrm{Cone}(A) := A\mathbb{R}^n_+$ the conic hull of the columns of A.
- $x \leqslant y$ is the standard partial order, given by $\forall i, x_i \leqslant y_i$.
- $F \subset G$ if F is a subface of G.
- $\mathcal{P} \preccurlyeq \mathcal{Q}$ if $\mathcal{P}$ is a refinement of the polyhedral complex $\mathcal{Q}$.
- $\mathcal{P} \wedge \mathcal{Q}$ is the common refinement of $\mathcal{P}$ and $\mathcal{Q}$.
- $E \subset_{\mathbb{P}} F$ if E is "$\mathbb{P}$-almost surely included" in F,*i.e.* $\mathbb{P}\big[E \cap F\big] = \mathbb{P}\big[E\big]$.
- $E \sim_{\mathbb{P}} F$ if E is "$\mathbb{P}$-almost surely equal" to F, *i.e.* $\mathbb{P}\big[E \cap F\big] = \mathbb{P}\big[E\big] = \mathbb{P}\big[F\big]$.
- $\operatorname{supp} \mathcal{C} := \bigcup_{C \in \mathcal{C}} E$ is the support of a collection of sets $\mathcal{C}$
- $E = X \sqcup Y$ means that $E = X \cup Y$ and $X \cap Y = \emptyset$.
- $\overline{\mathcal{C}}$ is the subcollection of maximal elements of a collection of sets $\mathcal{C}$.
- $\underline{\mathcal{C}}$ is the subcollection of minimal elements of a collection of sets $\mathcal{C}$.
- $\mathrm{rc}(P)$ is the recession cone of a polyhedron P.
- $\mathcal{F}(P)$ the set of faces of P.

- $\mathcal{F}_{\text{low}}(P)$ the set of lower faces of P.
- $\text{Vert}(P)$ the set of vertices of P.
- $\text{Ray}(P)$ a set with vectors each representing one extreme rays (for example the normalized extreme rays).
- P^{ψ} is the face of P given by $\arg\min_{x \in P} \psi^{\top} x$.
- $N_P(x)$ is the normal cone of P at x
- $\mathcal{N}(P)$ the normal fan of P.
- K° is the polar $\{\phi \,|\, \phi^{\top} x \leqslant 0, \forall x \in K\}$ of a cone K.
- $I_{A,b}(x) := \{i \mid A_i x = b_i\}$ the set of active constraints in x for an H-representation $\{z \mid Az \leqslant b\}$.
- $\mathcal{I}(A, b)$ the collection of these sets $\{I_{A,b}(x) \mid Ax \leqslant b\}$.
- $\mathcal{B}$ represents backward Bellman operators.
- $\mathcal{F}$ represents forward Bellman operators.
- ξ is a noise.
- ℓ_t is the loss function at time t.
- $\hat{V}_t$ is the cost-to-go function at time t.
- V_t is the expected cost-to-go function at time t.
- L denotes a Lipschitz constant.
- L denotes a diameter.
- X a set of state (e.g. X_t^r is the set of reachable state) and $\mathcal{X}$ a set-valued application representing set of states (e.g. $\mathcal{X}_t(x, \xi)$ is the set of admissible next state, $\mathcal{X}^{\sharp}_{\gamma,t}(\widetilde{V})(x, \xi)$ is the set of γ-optimal next state...).

Acronyms

APDP Adaptive Partition-based Dynamic Programming

APM Adaptive Partition-based method

CMS Computational Management Science

DSDDP Dual Stochastic Dual Dynamic Programming

DP Dynamic Programming

ECSO European Conference on Stochastic Optimization

EDDP Explorative Dual Dynamic Programming

GAPDP Generalized Adaptive Partition-based Dynamic Programming

GAPM Generalized Adaptive Partition-based method

GKZ Gelfand Kapranov Zelevinsky (vector or coordinate)

LP Linear Problem

ISDDP Inexact Stochastic Dual Dynamic integer Programming

MIDAS Mixed Integer Dynamic Approximation Scheme

MSCP Multistage Stochastic Convex Problem

MSLP Multistage Stochastic Linear Problem

MSP Multistage Stochastic Problem

NDDP Non-sequential Dual Dynamic Programming

PLP Parametric Linear Problem

RDDP Robust Dual Dynamic integer Programming

SAA Sample Average Approximation

SDDiP Stochastic Dual Dynamic integer Programming

SDDP Stochastic Dual Dynamic Programming

SDP Stochastic Dynamic Programming (should not be confused with Semi Definite Programming)

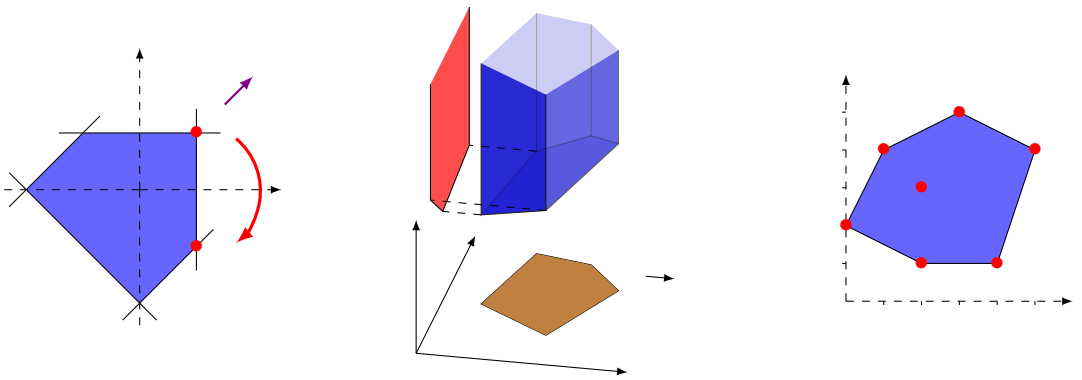
SLDP Stochastic Lipschitz Dynamic Programming

TDP Tropical Dynamic Programming

TFDP Trajectory Following Dynamic Programming

2SLP Multistage Stochastic Linear Problem

1 Introduction (version française)



Contents

L'optimisation est le domaine des mathématiques qui étudie les minima d'une fonction, appelée fonction objectif, sur un ensemble donné. Dans une perspective plus appliquée, l'optimisation modélise diverses situations où une personne doit choisir la meilleure décision, selon un certain objectif, en satisfaisant différentes contraintes. Au XVIIème siècle, le calcul différentiel de Leibniz et Newton a permis l'émergence de méthodes pour trouver des solutions aux problèmes d'optimisation sans constrainte avec des fonctions lisses. Des mathématiciens du XVIIIème siècle comme Euler et Lagrange ont ensuite développé des méthodes pour construire des solutions de problèmes d'optimisation sous contraintes. Depuis le XXème siècle, les progrès de l'informatique ont permis d'implémenter des algorithmes pratiques pour résoudre divers problèmes d'optimisation avec des fonctions non lisses. Par exemple, la programmation linéaire est un domaine de l'optimisation qui trouve de nombreuses applications et que l'on peut résoudre très efficacement en pratique. Dantzig [DT03] a développé une formulation générale pour l'optimisation linéaire et inventé la méthode du simplexe, qui est maintenant mieux comprise grâce à la géométrie polyédrale. Plus généralement, en optimisation convexe non lisse, des mathématiciens comme Fenchel et Rockafellar [Roc15] ont découvert de nouveaux liens entre optimisation et géométrie. Ainsi, les ponts entre l'optimisation et la géométrie ont permis d'améliorer l'efficacité d'algorithmes déjà existants et d'en inventer de nouveaux. Dans cette thèse, nous regarderons plus précisément les liens entre l'optimisation stochastique et la géométrie polyédrale.

Dans ce chapitre, nous présentons d'abord, dans la section 1.1 l'optimisation stochastique en rappelant les principes de la programmation dynamique et différentes méthodes d'approximation par discrétisation. Après avoir donné un aperçu des notions de géométrie polyédrale qui seront utilisées tout au long de cette thèse, et des raisons pour lesquelles ces notions ont été définies dans la section 1.2, nous introduisons le problème de discrétisation exacte dans la section 1.3, qui est la question centrale traitée dans ce travail. Les contributions principales de la thèse sont présentées dans la section 1.4 et son plan dans la section 1.5.

1.1 Optimisation stochastique

En optimisation sous incertitude, on doit prendre une décision sans connaître précisément certains paramètres essentiels du problème. L'incertitude peut provenir d'un manque d'information, d'une imprécision dans les mesures ou d'une difficulté à modéliser le problème. Le cas le plus commun est lorsque le décideur doit faire un choix ici et maintenant mais que le critère qu'il veut optimiser dépend d'un paramètre qu'il ne pourra observer que dans le futur.

Il existe diverses approches de la décision dans l'incertain. Il y a deux représentations principales de l'incertitude qui est soit modélisée comme antagoniste, soit stochastique. En optimisation robuste [BTEGN09, BBC11], le décideur doit choisir une solution optimale en considérant le pire cas d'un ensemble d'incertitude. En apprentissage par renforcement [KLM96, SB18], ou dans les problèmes de bandits manchots [BCB$^+$12], nous supposons que l'incertitude est aléatoire mais que le décideur ne connaît pas sa distribution et doit alors l'apprendre.

L'optimisation stochastique considère des problèmes où le décideur minimise une fonction à valeurs réelles, parfois appelée mesure de risque, qui dépend de paramètres incertains modélisés comme des variables aléatoires. Une partie de l'optimisation stochastique traite de problèmes d'optimisation distributionnellement robuste [WKS14, RM19, ZKW21] où l'on doit considérer la pire distribution dans un ensemble de distributions donné. Les problèmes d'optimisation stochastique avec aversion au risque avec mesure de risque cohérente [ADEH99] sont équivalents aux problèmes d'optimisation distributionnellement robuste. L'objectif à minimiser le plus naturel et répandu est l'espérance d'un coût. Lorsque l'on choisit l'espérance comme mesure de risque à optimiser, nous parlons de risque neutre. Dans cette thèse, nous considèrerons essentiellement[1] des problèmes avec risque neutre.

[1]Nous discuterons seulement le cas de l'aversion au risque dans la section Section 6.4.

1.1.1 Optimisation stochastique multi-étapes

Les problèmes d'optimisation stochastique multi-étapes, *multistage stochastic problems* (MSP) en anglais, constituent une classe importante des problèmes d'optimisation stochastique où les décisions x_t sont prises successivement à chaque étape t. Entre les étapes t et $t+1$, une partie de l'incertitude ξ_t est révélée.

$$x_0 \rightsquigarrow \xi_1 \rightsquigarrow x_1 \rightsquigarrow \xi_2 \rightsquigarrow \cdots \rightsquigarrow x_{T-1} \rightsquigarrow \xi_T \rightsquigarrow x_T$$

Les paramètres d'incertitude sont modélisés comme des variables aléatoires, et nous les noterons en caractères gras tout au long de cette thèse. Nous présentons maintenant informellement[2] un cadre général pour les MSP.

$$\begin{aligned} \min_{(\boldsymbol{x}_t)_{t\in[T]}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{x}_t, \boldsymbol{\xi}_t)\right] && && \text{(MSP)} \\ s.c. \quad & \boldsymbol{x}_0 = x_0 && p.s. && \\ & \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) && p.s. && \forall t \in [T] \\ & \sigma(\boldsymbol{x}_t) \subset \sigma(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_t) && && \forall t \in [T] \end{aligned}$$

À chaque étape t, le décideur choisit une solution $\boldsymbol{x}_t$ pour maximiser le coût actuel ℓ_t, qui dépend du choix $\boldsymbol{x}_t$ et d'une réalisation d'une variable aléatoire $\boldsymbol{\xi}_t$, auquel on ajoute l'espérance des coûts futurs. La contrainte $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ modèlise le fait que la décision $\boldsymbol{x}_t$ doit satisfaire certaines contraintes qui dépendent de la décision passée $\boldsymbol{x}_{t-1}$ et du paramètre d'incertitude actuel $\boldsymbol{\xi}_t$. La dernière contrainte, appelée contrainte de non-anticipativité, assure le fait que le décideur ne peut baser sa décision que sur les paramètres passés et présents observés et non sur les paramètres futurs pas encore observés. Ici, σ fait réference aux tribus, ou σ-algèbres, induites par des variables aléatoires. Pour la cohérence des notations, nous supposons que la première variable aléatoire $\boldsymbol{\xi}_1$ est fixée presque sûrement à un certain paramètre déterministe ξ_1 et que l'état x_0 est également fixé.

Exemple 1.1 (Modélisation de la gestion d'une centrale de stockage hydraulique par un MSP). Nous considérons une station de transfert d'énergie par pompage-turbinage. Les paramètres d'incertitude sont la demande en électricité et le flux d'eau provenant de la pluie et du ruissellement. Le décideur choisit une quantité d'eau turbinée chaque jour pour optimiser un certain coût, en respectant des contraintes physiques, par exemple en s'assurant que le barrage ne va pas déborder. Pour modéliser ce problème, nous pouvons définir le MSP suivant:

$$\begin{aligned} \min_{(\boldsymbol{u}_t)_t, (\boldsymbol{v}_t)_t} \quad & \mathbb{E}\left[\sum_{t=1}^{T} c_t(\boldsymbol{d}_t - \boldsymbol{u}_t)\right] \\ s.c. \quad & \boldsymbol{x}_0 = x_0 && p.s. && \\ & 0 \leqslant \boldsymbol{u}_t \leqslant \boldsymbol{d}_t && p.s. && \forall t \in [T] \\ & \boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{w}_t - \boldsymbol{u}_t - \boldsymbol{v}_t && p.s. && \forall t \in [T] \\ & 0 \leqslant \boldsymbol{x}_t \leqslant \overline{x} && p.s. && \forall t \in [T] \\ & 0 \leqslant \boldsymbol{v}_t && p.s. && \forall t \in [T] \\ & \sigma(\boldsymbol{u}_t) \subset \sigma(\boldsymbol{w}_1, \boldsymbol{d}_1 \cdots, \boldsymbol{w}_t, \boldsymbol{d}_t) && && \forall t \in [T], \end{aligned}$$

où $\overline{x}$ est la capacité du réservoir, T est le nombre d'étapes, ici nombre de jours, dans la période que nous considérons, et pour chaque étape t, $\boldsymbol{x}_t$ représente la quantité d'eau dans le réservoir,

[2] Une définition plus formelle sera donnée dans la section 6.1

$\boldsymbol{u}_t$ la quantité d'eau turbinée, $\boldsymbol{v}_t$ la quantité d'eau évacuée en ouvrant la vanne, $\boldsymbol{d}_t$ la demande, $\boldsymbol{w}_t$ la quantité d'eau apportée par la pluie et le ruissellement et c_t le coût d'une unité de demande non fournie.

1.1.2 Programmation dynamique

Les problèmes d'optimisation stochastique multi-étapes sont notoirement difficiles à résoudre et obtenir des solutions, même approximées, dans un temps raisonnable est hors de portée dans le cas général. En effet, sans hypothèse supplémentaire, la solution optimale est une fonction des bruits passés ce qui amène après discrétisation à un nombre de variables de décisions exponentiel en l'horizon. Une hypothèse usuelle, permettant une décomposition d'une part et de condenser l'information d'autre part, est l'indépendance des paramètres d'incertitude. Plus formellement, nous supposons que $(\boldsymbol{\xi}_t)_{t\in[T]}$ est une suite de variables aléatoires indépendantes.

Pour le décomposer, nous plongeons le problème (MSP) dans une famille de problèmes d'optimisation obtenus en faisant varier l'état initial x_{t-1}:

$$\begin{aligned} V_t(x_{t-1}) := \min_{(\boldsymbol{x}_\tau)_{t\leqslant\tau\leqslant T}} \quad & \mathbb{E}\left[\sum_{\tau=t}^{T} \ell_\tau(\boldsymbol{x}_\tau, \boldsymbol{\xi}_\tau)\right] \\ s.c. \quad & \boldsymbol{x}_{t-1} = x_{t-1} && p.s. \\ & \boldsymbol{x}_\tau \in \mathcal{X}_\tau(\boldsymbol{x}_{\tau-1}, \boldsymbol{\xi}_\tau) && p.s. \quad \forall t \leqslant \tau \leqslant T \\ & \sigma(\boldsymbol{x}_\tau) \subset \sigma(\boldsymbol{\xi}_t, \cdots, \boldsymbol{\xi}_\tau) && \forall t \leqslant \tau \leqslant T. \end{aligned}$$

où V_t est appelée la fonction de coût futur espéré au temps t.

En s'appuyant sur l'hypothèse d'indépendance, nous avons pour tout $t \in [T]$, l'équation suivante, appelée équation de Bellman ou équation de programmation dynamique (voir [Bel66, Ber12]) qui lie les fonctions de coût futur espéré aux étapes t et $t+1$:

$$V_t(x_{t-1}) = \mathbb{E}\Big[\hat{V}_t(x_{t-1}, \boldsymbol{\xi}_t)\Big] \tag{1.2a}$$

$$\hat{V}_t(x_{t-1}, \xi_t) := \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \ell_t(x_t, \xi_t) + V_{t+1}(x_t) \tag{1.2b}$$

où l'on pose $V_{T+1} :\equiv 0$ pour la cohérence des notations.

Notons enfin que l'on peut réécrire le problème MSP dans la forme imbriquée suivante:

$$\min_{x_1 \in X_1(x_0, \xi_1)} \ell_1(x_1, \xi_1) + \mathbb{E}\left[\min_{x_2 \in X_2(x_1, \boldsymbol{\xi}_2)} \ell_2(x_2, \boldsymbol{\xi}_2) + \mathbb{E}\Big[\cdots + \mathbb{E}\big[\min_{x_T \in \mathcal{X}_T(x_{T-1}, \boldsymbol{\xi}_T)} \ell_T(x_T, \boldsymbol{\xi}_T)\big]\Big]\right].$$

Le principe de programmation dynamique nous apprend l'existence d'une solution optimale donnée comme une fonction de l'état et du paramètre d'incertitude actuels. Pour mémoire, rappelons que, sans l'hypothèse d'indépendance, la solution optimale dépend de toute l'histoire des bruits passés. Ainsi, la programmation dynamique permet de condenser l'information nécessaire dans l'état.

De plus, la programmation dynamique permet également de décomposer un gros problème (voir le problème MSP) à T étapes en une suite de T problèmes d'intégration (1.2a) et de T problèmes (1.2b) paramétriques et déterministes à une étape. Plus précisément, pour calculer $V_t(x_{t-1})$ pour tout x_{t-1}, nous devons d'abord calculer $\hat{V}_t(x_{t-1}, \xi_t)$ en résolvant des problèmes (1.2b) d'optimisation déterministe à une étape chacun paramétré par l'état x_{t-1} et le paramètre d'incertitude ξ_t. Dans un deuxième temps, nous devons d'abord calculer une intégrale car $V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}(x_{t-1}, \boldsymbol{\xi}_t)\big]$. Nous expliquons maintenant comment effectuer de tels calculs lorsque les nombres d'états et de paramètres incertains possibles sont finis.

Lorsqu'à la fois les variables de décisions x_t et les paramètres d'incertitude ξ_t vivent dans des ensembles finis[3] X_t et Ξ_t, nous pouvons calculer exactement V_t par récurrence rétrograde.

[3]Ce cadre est aussi appelé processus de décision Markovien.

En supposant que $V_{t+1}(x_t)$ est connu pour tout $x_t \in X_t$, nous itérons sur tous les états possibles $x_{t-1} \in X_{t-1}$, et les paramètres incertains $\xi_t \in \Xi_t$ pour calculer $\hat{V}_t(x_{t-1}, \xi_t)$ comme $\min_{x_t \in \mathcal{X}_t(x_{t-1})} \ell_t(x_t, \xi_t) + V_{t+1}(x_t)$. Le minimum est calculé par exploration exhaustive, et $V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}(x_{t-1}, \boldsymbol{\xi}_t)\big]$ est une moyenne finie des $\hat{V}_t(x_{t-1}, \xi_t)$. Cet algorithme est appelé programmation dynamique stochastique (SDP), voir l'algorithme 1.1.

```
1  V_{T+1} ≡ 0 ;
2  for t from T to 1 do
3  |  for x_{t-1} ∈ X_{t-1} do
4  |  |  v ← 0;
5  |  |  for ξ_t ∈ Ξ_t do
6  |  |  |  m ← +∞;
7  |  |  |  for x_t ∈ X_t do
8  |  |  |  |  if x_t ∈ X_t(x_{t-1}, ξ_t) then
9  |  |  |  |  |  m ← min(m, ℓ_t(x_t, ξ_t) + V_{t+1}(x_t));
10 |  |  |  |  end
11 |  |  |  end
12 |  |  |  v ← v + m × P[ξ_t = ξ_t];
13 |  |  end
14 |  |  V_t(x_{t-1}) ← v;
15 |  end
16 end
```

Algorithm 1.1: Programmation dynamique stochastique (SDP) pour MSP dans le cas fini.

Néanmoins, dans beaucoup d'applications, l'espace d'état où x_t vit est continu. Nous considérons alors un cadre plus général où x_t peut prendre un nombre infini de valeurs. Une première idée intéressante est de discrétiser. Au lieu de résoudre directement le MSP, nous résolvons par l'algorithme SDP, une version approximée du MSP où les variables de décision x_t appartiennent à un ensemble fini X_t. La précision de l'approximation augmente avec le nombre de points calculés. En particulier, la taille des ensembles discrétisés nécessaire pour obtenir une certaine précision, est exponentielle en la dimension de l'espace d'état. Plusieurs approches ont été développées pour résoudre les MSP tout en évitant ce problème appelé malédiction de la dimension.

Au lieu de réaliser une seule récurrence rétrograde en calculant $V_t(x_{t-1})$ pour tout x_{t-1} appartenant à une grille discrétisée préalablement, nous pouvons approximer $V_t(x_{t-1})$ le long de certaines trajectoires choisies au fur et à mesure de l'algorithme. L'algorithme *Stochastic Dual Dynamic Programming* (SDDP), introduit par Pereira et Pinto dans [PP91] pour gérer le système de production hydro-électrique Brésilien, réalise successivement des phases directes pour calculer des trajectoires et des phases rétrogrades pour mettre à jour des estimations des fonctions de coût futur espéré, le long de ces trajectoires, via la théorie de la dualité. Il existe de nombreuses variantes de l'algorithme SDDP que nous présentons et commentons dans le chapitre 6, nous englobons toutes ces variantes sous le nom d'algorithmes de programmation dynamique par suivi de trajectoire, *Trajectory Following Dynamic Programming algorithm* (TFDP) en anglais. Ainsi, les algorithmes TFDP permettent de considérer des espaces de décisions X_t infinis.

Cependant, dans la littérature actuelle, les algorithmes TFDP ne fonctionnent que lorsque les variables $\boldsymbol{\xi}_t$ sont à support fini. En effet, sans l'hypothèse de support fini, nous ne pouvons pas espérer calculer $\hat{V}_t(x_{t-1}, \xi_t)$ car il faudrait pour cela résoudre un nombre infini de problème (1.2b) à une étape, paramétré par le paramètre incertain ξ_t. Pour éviter cette difficulté, nous pouvons remplacer la distribution, a priori générale, de $\boldsymbol{\xi}_t$ par une distribution à support fini. Les différentes approches dans la littérature pour réduire les paramètres stochastiques ξ_t à un nombre fini sont appelées méthodes de discrétisation.

1.1.3 Discrétisation en optimisation stochastique

Nous présentons maintenant différentes méthodes de discrétisation fréquemment utilisées en optimisation stochastique pour approximer des MSP avec des distributions générales.

Sample Average Approximation Une première idée consiste à remplacer les variables aléatoires $\boldsymbol{\xi}_t$ à distribution générale par N scénarios $(\check{\xi}_{t,k})_{k\in[N]}$ tirés indépendamment et identiquement distribués selon la distribution de $\boldsymbol{\xi}_t$. Comme dans les méthodes de Monte-Carlo, nous remplaçons alors la distribution de probabilité réelle $\mathbb{P}$ par la distribution de probabilité empirique $\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{k=1}^{N} \delta_{\check{\xi}_{t,k}}$. L'espérance $\mathbb{E}\big[f(\boldsymbol{\xi}_t)\big]$ est alors remplacée par la moyenne d'un nombre fini d'échantillons $\frac{1}{N_t}\sum_{k=1}^{N} f(\check{\xi}_{t,k})$. Par exemple, l'équation de programmation dynamique (1.2) devient

$$V_t(x_{t-1}) = \frac{1}{N}\sum_{k=1}^{N} \Big(\underbrace{\min_{x_t \in \mathcal{X}_t(x_{t-1},\check{\xi}_{t,k})} \ell_t(x_t,\check{\xi}_{t,k}) + V_{t+1}(x_t)}_{:=\hat{V}(x_{t-1},\check{\xi}_{t,k})} \Big).$$

Nous obtenons ensuite un nouveau MSP avec des variables aléatoires à support fini suivant leurs distributions empiriques. Ce nouveau problème est appelé *Sample Average Approximation* (SAA), voir le chapitre 5 de [SDR14] pour une analyse exhaustive du sujet. Par la loi des grands nombres et sous hypothèses faibles de régularité, le problème SAA converge vers le MSP original, dans plusieurs sens, par exemple convergence de la valeur ou de l'ensemble des solutions optimales. En supposant l'existence d'une variance, par le théorème de la limite centrale, on peut aussi obtenir des intervalles de confiance pour la valeur du MSP.

Le premier avantage de la méthode SAA est qu'elle ne nécessite que de faibles hypothèses et fonctionne pour une très grande classe de problèmes d'optimisation stochastique. Deuxièmement, SAA est très facile à implémenter en pratique, dès que l'on a une méthode pour tirer facilement des réalisations des variables aléatoires. Cependant, les résultats de convergence sont soit asymptotiques, soit en $\frac{1}{\sqrt{N}}$ tout en restant probabilistes. Notons que la valeur v_N^{SAA} du problème SAA est une variable aléatoire (dépendant de l'échantillon tiré), et peut ainsi être vu comme un estimateur statisque de la valeur réelle du problème. Si cet estimateur est cohérent (asymptotiquement presque sûrement convergent), il est biaisé (l'espérance de v_N^{SAA} est plus petite que la valeur réelle), mais nous ne pouvons pas garantir qu'une réalisation donnée fournit une borne inférieure de la valeur réelle.

Approximations basées sur les inégalités d'Edmundson-Madansky et de Jensen Nous introduisons maintenant différentes méthodes d'approximation, sous des hypothèses de convexité, exploitant les inégalités d'Edmundson-Madansky et de Jensen. Nous présentons plus formellement ces inégalités dans la section 6.2.4 et renvoyons à [Kuh06, EZ94b, EZ94a, KM$^+$76] pour des études plus détaillées sur le sujet. Lorsqu'une fonction f est convexe en un paramètre stochastique $\boldsymbol{\xi}$, l'inégalité de Jensen nous dit que $f(\mathbb{E}[\boldsymbol{\xi}]) \leqslant \mathbb{E}\big[f(\boldsymbol{\xi})\big]$. Cette inégalité fournit alors des bornes inférieures aux valeurs de problèmes d'optimisation stochastique convexe multi-étapes, *Multistage Stochastic Convex Programs* (MSCP) en anglais, en remplaçant les va-riables aléatoires $\boldsymbol{\xi}_t$ par des variables déterministes égales à leurs espérances $\mathbb{E}\big[\boldsymbol{\xi}_t\big]$. Le nouveau problème est appelé le problème de la valeur espérée (voir par exemple [MAB14]) qui donne une borne inférieure pour la valeur réelle du MSCP. Cependant, cette borne inférieure peut être très éloignée de la valeur réelle. Pour améliorer cette borne inférieure, nous pouvons partitionner l'ensemble Ξ où la variable aléatoire $\boldsymbol{\xi}$ vit et appliquer l'inégalité de Jensen à chaque ensemble P de la partition $\mathcal{P}$. Nous pouvons alors considérer un nouveau problème approximé où la variable aléatoire $\boldsymbol{\xi}$ est remplacée par une variable à support fini $\check{\xi}$ avec les scénari $\check{\xi}_P := \mathbb{E}\big[\boldsymbol{\xi} \,|\, \boldsymbol{\xi} \in P\big]$ associés au probabilités $\check{p}_P := \mathbb{P}\big[\boldsymbol{\xi} \in P\big]$. Dans ce cas, plus la partition $\mathcal{P}$ est fine, meilleure est l'approximation. De plus, quand le diamètre des ensembles $P \in \mathcal{P}$ tend

vers 0, la valeur approximée converge vers la valeur du MSCP réel. Lorsque les valeurs des variables aléatoires appartiennent à un polyèdre Q, nous pouvons obtenir une borne supérieure sur l'espérance $\mathbb{E}\big[f(\boldsymbol{\xi})\big]$ en prenant une combinaison convexe des valeurs $f(v)$ où v décrit les sommets du polyèdre Q. Cette inégalité, appelée inégalité d'Edmundson-Madansky (voir [Mad59]), permet d'obtenir de la même façon des schémas d'approximations avec des partitions pour déduire des bornes supérieures convergentes. Enfin, lorsque les fonctions de coûts sont des fonctions selles, c'est-à-dire qu'elles sont convexes dans une variable aléatoire $\boldsymbol{\xi}$ et concave dans une autre variable aléatoire $\boldsymbol{\eta}$, en s'appuyant sur les idées des deux inégalités, nous pouvons obtenir des bornes supérieures et inférieures pour les MSCP. L'avantage de ce genre d'approximations est la garantie d'avoir des bornes supérieures ou inférieures. En revanche, ces approximations sont plus difficiles à implémenter que SAA, et nécessitent une hypothèse de convexité.

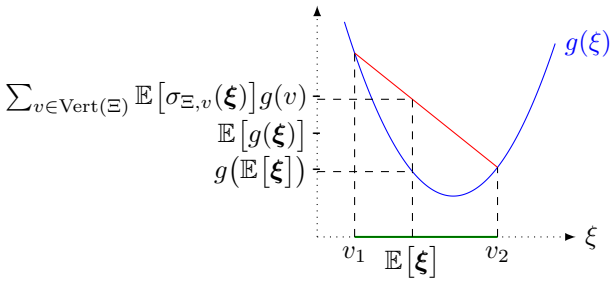


Figure 1.2: Une illustration des inégalités de Jensen et d'Edmundson-Madansky, voir aussi Proposition 6.7

Approximations basées sur la dépendance Dans le cas où les variables aléatoires sont dépendantes entre elles, nous pouvons approximer ces dépendances. L'idée consiste à relâcher les contraintes de non-anticipativité, par exemple en considérant que le décideur peut voir le futur, ou en rassemblant différents scénari en les représentant par leur moyenne. Par exemple, l'algorithme *progressive hedging* présenté par Rockafellar et Wets dans [RW91] est basé sur ces idées. Dans le cas convexe, en exploitant les inégalités de Jensen et d'Edmundson-Madansky, nous pouvons obtenir des bornes supérieures et inférieures grâce à des arbres de scénari barycentriques (voir [Fra96, CS05, Kuh06]). Dans [PP12], Pflug and Pichler ont défini une distance, appelée distance imbriquée, *nested distance* en anglais, entre des suites de variables aléatoires dépendantes. Cette distance imbriquée permet de borner l'écart, notamment de valeur, entre un MSP et ses problèmes approximés et ainsi fournir de bonnes approximations du MSP.

Notons qu'en optimisation stochastique avec risque neutre, nous supposons en général que les distributions des variables aléatoires sont connues par le décideur. Cette hypothèse rend les problèmes d'optimisation stochastique très dépendants du choix de la distribution lorsque l'on modélise un problème pratique. L'impact de ce choix est étudié dans le domaine de l'analyse de sensibilité [Dup90, RS91] qui consiste à comprendre comment la valeur d'un problème stochastique varie avec de petites perturbations de la distribution. En pratique, il n'est souvent pas nécessaire de résoudre un problème d'optimisation stochastique de manière très précise ou exacte car l'erreur de modélisation peut supplanter les erreurs d'optimisation ou d'approximation. En revanche, d'un point de vue théorique, fournir des renseignements mathématiques sur les solutions optimales d'un problème avec distributions de probalités connues, ouvre la porte à une meilleure compréhension de la structure du problème et ainsi à l'amélioration d'algorithmes existants et à des idées pour en développer de nouveaux.

Dans cette thèse, en lieu et place de ces approches, nous développons des méthodes de discrétisation exacte en s'appuyant sur des notions de géométrie polyédrale.

1.2 Géométrie polyédrale, combinatoire et complexité

L'optimisation a révélé des liens féconds avec les domaines de la géométrie et de la combinatoire. Nous rappelons d'abord quelques connexions standards entre les problèmes d'optimisation linéaire et la géométrie polyédrale. Nous présentons ensuite brièvement des outils de géométrie polyédrale "d'ordre supérieur", comme l'éventail secondaire qui sera un objet essentiel dans cette thèse, et leurs liens avec les problèmes d'optimisation linéaire paramétrée. Ces notions plus fondamentales sont à l'origine motivées par l'étude des systèmes d'équations polynomiales et la théorie de l'élimination, un sujet qui ne sera pas examiné dans cette thèse mais que l'on évoque brièvement dans la section Section 1.2.3. Enfin, après avoir introduit rapidement la théorie de la complexité, nous rappelons des résultats de complexité connus en géométrie polyédrale et en optimisation stochastique.

1.2.1 Optimisation linéaire et polyèdres

L'optimisation linéaire est le domaine de l'optimisation qui considère des fonctions objectif linéaires et des contraintes définies comme des inégalités affines (voir [MG07]). Un polyèdre est un ensemble de points qui correspond à une intersection d'un nombre fini de demi-espaces. Nous donnons une définition plus formelle de la notion de polyèdre dans la section 3.1 et nous renvoyons à la lecture des livres [Zie12, Grü13] pour une introduction plus complète à la théorie des polyèdres. Ainsi, l'ensemble des solutions admissibles d'un problème linéaire (LP) est un polyèdre. Un résultat essentiel en optimisation linéaire est le fait que le minimum d'une fonction linéaire sur un polyèdre, est atteint sur au moins un sommet, ou point extrême, lorsque le polyèdre admet au moins un sommet et que l'infimum est fini. La méthode du simplexe est l'un des algorithmes les plus efficaces et communément utilisés pour résoudre les LP. Le simplexe revient à suivre un chemin le long des sommets du polyèdre des points admissibles en faisant décroître la fonction objectif.

$$
\begin{aligned}
\min_{x\in\mathbb{R}^2} \quad & c^\top x := x_1 + x_2 & \\
s.c. \quad & x_1 \leqslant 1 & (1)\\
& x_1 - x_2 \leqslant 2 & (2)\\
& -x_1 - x_2 \leqslant 2 & (3)\\
& -x_1 + x_2 \leqslant 2 & (4)\\
& x_2 \leqslant 1 & (5)
\end{aligned}
$$

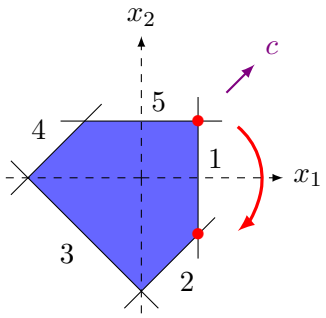


Figure 1.3: Illustration d'une étape de l'algorithme du simplexe pour un problème d'optimisation linéaire. Nous pivotons de la base $\{1, 5\}$ à la base $\{1, 2\}$.

Pour implémenter la méthode du simplexe dans un ordinateur, nous avons besoin de notions plus combinatoires. Une base est un ensemble d'indices, représentant les lignes ou les colonnes d'une matrice, qui induit une sous-matrice inversible. Les sommets d'un polyèdres peuvent être représentés grâce à des bases. Ainsi, d'un point de vue combinatoire, la méthode du simplexe passe de base en base grâce à une règle de pivot, voir la figure 1.3. Les ponts entre la géométrie polyédrale et l'optimisation linéaire sont largement étudiés, par exemple pour améliorer les performances de la méthode du simplexe. Des problèmes fondamentaux concernant la complexité de l'optimisation linéaire tels que le 9ème problème de Smale [Sma98] sont ainsi fortement liés à des conjectures fondamentales de géométrie polyédrale comme la conjecture de Hirsch [San12].

1.2.2 Notions polyédrales d'ordre supérieur

L'optimisation linéaire paramétrée étudie le comportement des problèmes d'optimisation linéaire lorsque des paramètres comme le vecteur de coût, la matrice des contraintes ou le second membre des contraintes varient. Une approche combinatoire de l'optimisation linéaire paramétrée a été développée par Walkup et Wets [WW69]. Leur théorème de décomposition en bases décrit comment la valeur d'un problème linéaire sous forme standard varie avec le coût et le second membre des contraintes. Une approche plus récente par Sturmfels et Thomas [ST97] a connecté ce théorème à la théorie des subdivisions régulières.

Les subdivisions régulières sont des objets combinatoires qui décrivent la géométrie d'un nombre fini de points (resp. vecteurs) appelé configuration de points (resp. vecteurs.) Nous définissons formellement les subdivisions régulières dans la définition 3.17 et renvoyons à la monographie de De Loera, Rambau et Santos [DLRS10] pour une analyse plus exhaustive de cette théorie. Informellement, une subdivision régulière d'une configuration de vecteurs $A = (a_i)_{i\in[n]}$ associée au vecteur de hauteurs $\omega \in \mathbb{R}^n$ est la collection des ensembles d'indices $I_F = \{i \in [n] \,|\, a_i \in F\}$ quand F décrit les faces visibles par dessous du cone $\mathrm{Cone}(a_i, \omega_i)_{i\in[n]}$, voir la figure 1.4. Un objet géométrique fondamental appelé éventail[4] secondaire et noté Σ-fan(A) offre une meilleure compréhension de la structure des subdivisions régulières. Plus précisément, 2 vecteurs de hauteur donnent la même subdivision régulière si et seulement s'ils appartiennent au même intérieur relatif d'un cone secondaire (i.e. un cone de l'éventail secondaire).

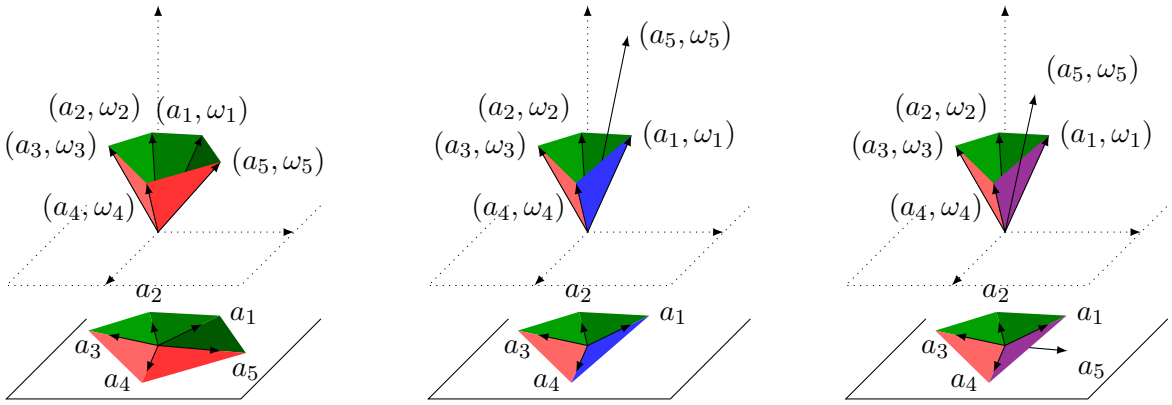


Figure 1.4: 3 configurations de vecteurs relevées et les projections des faces visibles par dessous voirSection 3.3.1 pour plus de détails.

Cet éventail est en fait l'éventail normal d'un certain polyèdre appelé polytope secondaire et noté $\Sigma(A)$. Les sommets de $\Sigma(A)$ peuvent être retrouvés avec les subdivisions régulières de A vu comme une configuration de vecteurs, grâce à ce qu'on appelle les coordonnées GKZ. Dans [BS92], Billera et Sturmfels ont défini un autre objet fondamental appelé le polytope des fibres, *fiber polytope* en anglais, pour mieux comprendre la structure d'une classe plus générale de subdivisions. Ils ont aussi montré que le polytope secondaire était un cas particulier de polytope des fibres.

1.2.3 Motivation par la théorie de l'élimination et les équations polynomiales

Ces notions récentes et avancées d'ordre supérieur, utilisées tout au long dans cette thèse dans le contexte de l'optimisation stochastique, sont motivées, dans un tout autre contexte, par l'étude des systèmes d'équation polynomiales.

En théorie de l'élimination, le *resultant* $R(P, Q)$, ou déterminant de Sylvester, de 2 polynômes P et Q à une indéterminée, est un scalaire qui vaut 0 si et seulement si P et Q ont une racine

[4] Un éventail est un complexe polyédrale dont les cellules sont des cones.

commune. Le résultant de P et Q est défini comme le déterminant d'une matrice qui dépend des coefficients de P et Q. Ainsi, la fonction résultant $R_d : (P, Q) \mapsto R(P, Q)$, restreinte à l'ensemble des polynômes de degrés plus petit que d, peut être vue comme un polynôme multivarié dans les coefficients de P et Q. Le *discriminant* $\Delta_d(P)$ d'un polynôme univarié P est défini comme le résultant $R(P, P')$ de P et du polynôme dérivé P'. Tout comme le résultant, la fonction discriminant Δ_d peut également être vue comme un polynôme multivarié. Ces objets sont fondamentaux en théorie de l'élimination pour résoudre des systèmes polynomiaux.

Nous considérons à présent le cas de polynômes à plusieurs variables. Soit $A \subset \mathbb{N}^n$ un ensemble fini de vecteurs d'entiers naturels. Les notions de résultants et de discriminants peuvent être étendus au polynômes multivariés, nous parlons alors de A-résultant et A-discriminant, notés respectivement R_A et Δ_A (voir [GKZ94, Chapitre 8 et 9]). Enfin, au polynôme multivarié $P := \sum_{a \in A} c_a X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$ dont les exposants sont dans A, nous pouvons associer son *polytope de Newton* noté $\mathrm{Nw}(P) := \mathrm{Conv}\{a \in A \,|\, c_a \neq 0\}$, voir Fig. 1.5. Le polytope de Newton est une notion essentielle pour étudier le comportement des polynômes multivariés (voir [EK08, MTY19]). Dans [GKZ90] (voir aussi [GKZ94, Chapter 13]), Gelfand, Kapranov et Zelevinsky ont montré que le discriminant Δ_A avait des liens profonds avec la théorie des subdivisions régulières. Plus précisément, ils ont prouvé que le polytope de Newton $\mathrm{Nw}(\Delta_A)$ du discriminant Δ_A était égal, à un facteur scalaire près, au polytope secondaire $\Sigma(A)$. Sturmfels a ensuite prouvé dans [Stu94] que le polytope de Newton du résultant $\mathrm{Nw}(R_A)$ divise, au sens de la somme de Minkowski, un polytope de fibres particulier.

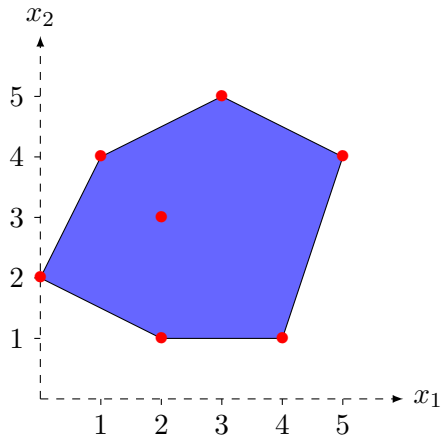


Figure 1.5: The Newton polytope $\mathrm{Nw}(P)$ of the polynomial $P := X_1^3X_2^5 + 3X_1^5X_2^4 - 7X_1^4X_2 + X_1^2X_2 - 8X_2^2 + X_1X_2^4 - 4X_1^2X_2^3$

1.2.4 Théorie de la complexité

La théorie de la complexité classifie les problèmes mathématiques selon leurs difficultés à être résolus. On peut prouver qu'un problème de décision D est au moins aussi dur qu'un problème D' en montrant que chaque instance de D' est équivalente à une instance de D. On dit alors que D' peut être réduit à D. A l'inverse, pour prouver qu'un problème D peut être résolu en temps raisonnable, on peut étudier la complexité de n'importe quel algorithme qui résoud D. Nous renvoyons à la monographie d'Arora et Barak [AB09] pour une introduction exhaustive sur la théorie de la complexité.

Dans [Kha80], Khachiyan a prouvé que la méthode de l'ellipsoïde résolvait les problèmes d'optimisation linéaire en temps polynomial, démontrant ainsi que l'optimisation linéaire était dans la classe de complexité P. Nous renvoyons au livre [GLS12] de Grötschel, Lovász et Schrijver pour une introduction complète sur la complexité de la programmation linéaire. En revanche, lorsque l'on rajoute certaines contraintes, comme le fait d'avoir des solutions à coordonnées

entières, ou lorsque l'on modifie la structure du problème, par exemple avec de l'aléatoire, la classe de complexité du problème change radicalement. En effet, la programmation linéaire en nombres entiers est un problème NP-difficile (voir [vzGS78]). Quand un problème est dans un classe de complexité aussi difficile, nous pouvons montrer que des restrictions du problème avec des paramètres fixés sont résolubles en temps polynomial, *i.e.* sont dans la classe P. Cette approche est assez commune et souvent la meilleure que l'on puisse espérer. Par exemple, Lenstra a démontré dans [LJ83] que, lorsque la dimension de l'ensemble des points admissibles est fixée, la programmation linéaire en nombres entiers devient résoluble en temps polynomial. De même, si l'on peut prouver que compter le nombre de points à coordonnées entières d'un polytope est un problème ♯P-complet, Barvinok a en revanche montré dans [Bar94] (voir aussi [Bar08]) que lorsque la dimension est fixée, compter le nombre de points entiers dans un polytope devient un problème polynomial. De manière peut-être plus surprenante, le problème du calcul du volume d'un polytope est comparable à celui du comptage de nombre de points entiers, c'est-à-dire ♯P-complet en général [DF88] et polynomial lorsque que la dimension est fixée.

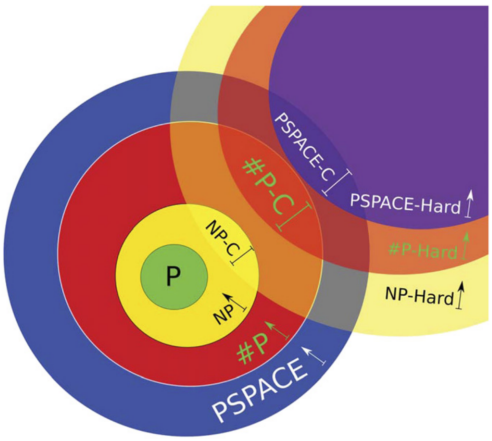


Figure 1.6: Différentes classes de complexité. L'image provient de [IJCN15].

La complexité de l'optimisation stochastique a elle aussi été beaucoup étudiée. Dans [DS06], Dyer et Stougie ont montré qu'un cas particulier de l'optimisation stochastique multi-étapes est PSPACE-difficile, et que l'optimisation stochastique à 2 étapes (2SP) avec des distributions discrètes est ♯P-difficile en réduisant le problème de fiabilité des graphes, *graph reliability* en anglais, au problème 2SP. Hanasusanto, Kuhn et Wiesemann [HKW16] ont ensuite démontré que résoudre, avec une précision assez élevée, des problèmes stochastiques linéaires à 2 étapes (2SLP) avec distributions continues était aussi ♯P-difficile, en s'appuyant sur le caractère ♯P-complet du calcul des volumes du polytope du sac à dos et de l'*order polytope*. Plusieurs articles comme [Sha06] ont étudié la complexité de 2SLP et MSLP en analysant la convergence des méthodes SAA. Shapiro et Nemirovski [SN05] ont prouvé en particulier que les problèmes 2SLP peuvent être résolus avec la méthode SAA, avec une grande probabilité et une précision ε, en temps pseudo-polynomial, *i.e.* polynomial en $1/\varepsilon$ et dans la taille de l'entrée du problème, lorsque les dimensions des espaces de décisions sont fixées. Ceci implique que les petits problèmes 2SLP peuvent être facilement résolus avec une précision moyenne. Enfin, Lan [Lan20] et Zhang et Sun [ZS19] ont analysé indépendemment la complexité de l'algorithme SDDP. Leurs résultats impliquent que les MSLP avec des variables aléatoires à support fini peuvent être résolus approximativement en temps pseudo-polynomial en l'erreur d'approximation ε lorsque toutes les dimensions sont fixées.

1.3 Le problème de discrétisation exacte

Dans cette thèse, nous nous demandons comment traiter les problèmes d'optimisation stochastique avec des distributions générales (et donc avec des supports non-nécessairement finis). La question principale est la suivante:

> Comment résoudre de manière *exacte* des problèmes d'optimisation stochastique multi-étapes avec des *distributions générales* ?

En d'autres termes, existe-t-il des formules analytiques et des algorithmes pour calculer des solutions exactes au problèmes d'optimisation stochastique avec des distributions générales ?

Nous avons vu dans la section 1.1.3 que nous pouvons approximer un MSP, soit par SAA soit avec les inégalités de Jensen et Edmundson-Madansky, en remplaçant les variables aléatoires continues par des variables aléatoires à support finis, voir la table 1.1. Ces techniques peuvent être comprises comme des méthodes de discrétisations approximées, nous définissons maintenant la notion de discrétisation exacte.

Définition 1.1 (Discrétisation exacte)**.** *Nous disons qu'un MSP (avec indépendance des variables aléatoires entre les étapes) admet une* discrétisation exacte locale *au temps t en x_{t-1} s'il existe une variable aléatoire $(\check{\boldsymbol{\xi}}_t)_{t\in[T]}$ à support fini qui produit les mêmes fonctions de coût futur espéré, c'est-à-dire tel que*

$$V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}_t(x_{t-1}, \boldsymbol{\xi}_t)\big] = \mathbb{E}\big[\hat{V}_t(x_{t-1}, \check{\boldsymbol{\xi}}_t)\big].$$

En particulier, nous avons $V(x_{t-1}) = \sum_{\xi\in\mathrm{supp}(\check{\boldsymbol{\xi}}_t)} \mathbb{P}\big[\boldsymbol{\xi}_t = \xi\big]\hat{V}(x_{t-1}, \xi)$.

*Nous qualifions une discrétisation d'*uniformément exacte *si elle est localement exacte pour tout $x_{t-1} \in \mathbb{R}^{n_{t-1}}$ et tout $t \in [T]$.*

Nous disons qu'une discrétisation est induite par une partition *s'il existe une partition $\mathcal{P}_{t,x_{t-1}}$ (mesurable et finie) de l'ensemble d'incertitude Ξ_t telle que pour tout $P \in \mathcal{P}_{t,x_{t-1}}$, on a*

$$\mathbb{P}\big[\check{\boldsymbol{\xi}}_t = \check{\xi}_{t,P}\big] = \check{p}_{t,P} \quad \text{avec} \quad \check{p}_{t,P} := \mathbb{P}\big[\boldsymbol{\xi}_t \in P\big], \quad \check{\xi}_{t,P} := \mathbb{E}\big[\boldsymbol{\xi}_t \,|\, \boldsymbol{\xi}_t \in P\big]^5.$$

En particulier, la discrétisation induite par une partition s'écrit

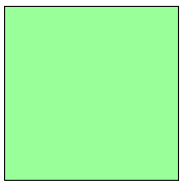
$$V_t(x_{t-1}) = \sum_{P\in\mathcal{P}_{t,x_{t-1}}} \check{p}_{t,P}\hat{V}(x_{t-1}, \check{\xi}_{t,P}).$$

Enfin, s'il existe une partition $\mathcal{P}_{t,x_{t-1}}$ telle que la discrétisation induite est exacte au temps t en x_{t-1}, pour toutes les distributions de $(\boldsymbol{\xi}_\tau)_{\tau\in[T]}$, on dit que la discrétisation exacte est universelle.

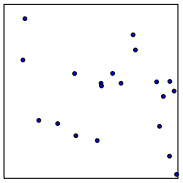
	Hypothèse	Approximations
SAA	Général	Intervalle de confiance
Jensen/Edmundson-Madansky	Convexe	Bornes supérieures et inférieures
Discrétisation exacte	Linéaire	Valeurs exactes

Table 1.1: Comparaison des avantages et inconvénients des méthodes de discrétisations approximées ou exactes

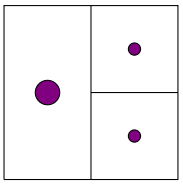
[5] Lorsque $\check{p}_{t,P} := \mathbb{P}\big[\boldsymbol{\xi}_t \in P\big]$ est égal à 0, $\mathbb{E}\big[\boldsymbol{\xi}_t \,|\, \boldsymbol{\xi}_t \in P\big]$ n'est pas bien défini. Nous choisissons alors un $\check{\xi}_{t,P}$ arbitraire dans P. Ce choix n'a pas d'importance car $\check{\xi}_{t,P}$ n'apparaîtra dans les formules qu'après le facteur nul $\check{p}_{t,P} = 0$.



$\boldsymbol{\xi}_t$ continu



$(\check{\xi}_{t,k})_{k=N}$ pour SAA avec $N = 20$



$(\check{\xi}_{t,P})_{P\in\mathcal{P}}$ pour une discrétisation induite par $\mathcal{P}$

Figure 1.7: Comparaisons des scénari choisis selon la discrétisation.

Dans cette thèse, nous nous intéressons uniquement aux discrétisations induites par des partitions. Avant de rentrer précisément dans les théorèmes de discrétisation exacte abordés dans le coeur de la thèse, nous expliquons pourquoi nous ne pouvons pas espérer avoir un théorème général de discrétisation exacte sans linéarité. Nous nous concentrons ensuite sur l'optimisation stochastique linéaire. En donnant des contre-exemples, nous montrons qu'il n'existe ni discrétisation locale ni discrétisation uniforme exacte et induite par une partition lorsque la matrice de recours est stochastique avec support infini, et qu'il n'existe pas de discrétisation uniforme lorsque les contraintes sont stochastiques avec support infini.

1.3.1 Cas non-linéaire

Si $\mathcal{P}_{t,x_{t-1}}$ est une partition de Ξ_t, la loi des espérances totales donne

$$V_t(x_{t-1}) = \sum_{P\in\mathcal{P}_{t,x_{t-1}}} \check{p}_{t,P}\mathbb{E}\big[\hat{V}_t(x_{t-1},\boldsymbol{\xi}_t)\,|\,\boldsymbol{\xi}_t\in P\big].$$

Si nous supposons que la fonction $\hat{V}_t$ de coût futur paramétré est convexe dans le paramètre incertain ξ_t, par l'inégalité de Jensen nous avons $\hat{V}_t(x_{t-1},\check{\xi}_{t,P}) = \hat{V}_t\Big(x_{t-1},\mathbb{E}\big[\boldsymbol{\xi}_t\,|\,\boldsymbol{\xi}_t\in P\big]\Big) \leqslant \mathbb{E}\Big[\hat{V}_t(x_{t-1},\boldsymbol{\xi}_t)\,|\,\boldsymbol{\xi}_t\in P\Big]$, pour tout $P\in\mathcal{P}_{t,x_{t-1}}$. En particulier, nous avons une discrétisation exacte induite par cette partition si et seulement si nous avons l'égalité dans l'inégalité de Jensen pour tout $P\in\mathcal{P}_{t,x_{t-1}}$ tel que $\check{p}_{t,P}>0$. Lorsque l'on considère des fonctions strictement convexes, l'inégalité de Jensen est généralement stricte, on dit aussi que le saut de Jensen est strictement positif (voir par exemple [Wal14]). Ainsi, une discrétisation exacte induite par une partition est très improbable si $\hat{V}_t$ est strictement convexe en ξ_t. Nous pouvons faire la même analyse lorsque $\hat{V}_t$ est strictement concave en ξ_t en inversant l'inégalité. Même si l'on pourrait trouver certains cas particuliers qui fonctionnent, nous ne pouvons espérer obtenir un résultat général de discrétisation exacte si $\hat{V}_t$ n'est pas affine par morceaux en ξ_t.

Nous nous concentrons donc sur le cas linéaire.

1.3.2 Contre-exemples en optimisation stochastique linéaire

Les problèmes d'optimisation stochastique linéaire multi-étapes (MSLP) sont des cas particuliers des MSP où les fonctions de coûts instantanés sont linéaires et les contraintes sont affines. Pour simplifier, nous considérons un problème stochastique linéaire à 2 étapes (2SLP) qui est un cas particulier de MSLP avec horizon 2. Dans les 2SLP, une seule fonction des coûts futurs espérés est étudiée $V := V_2$. Nous notons aussi $n := n_1$ et $m := n_2$ les dimensions des variables d'états $x := x_1$ et $y := x_2$ et $\Xi := \mathrm{supp}(\boldsymbol{\xi}) \subset \mathbb{R}^m\times\mathbb{R}^{\ell\times n}\times\mathbb{R}^{\ell\times m}\times\mathbb{R}^\ell$ le support de la variable aléatoire $\boldsymbol{\xi} = (\boldsymbol{q},\boldsymbol{W},\boldsymbol{T},\boldsymbol{h}) := \boldsymbol{\xi}_2 = (\boldsymbol{c}_2,\boldsymbol{A}_2,\boldsymbol{B}_2,\boldsymbol{b}_2)$.

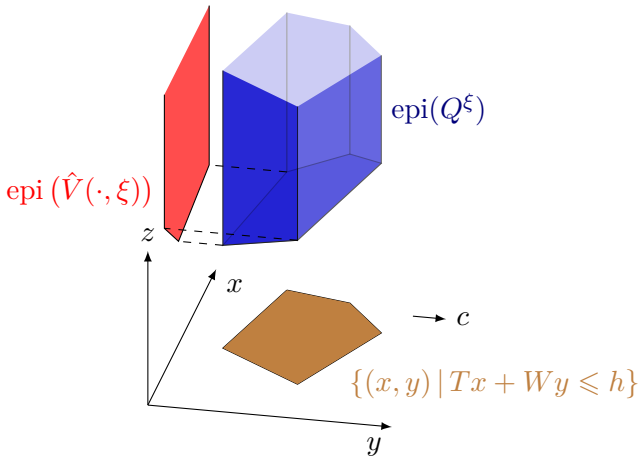


Figure 1.8: epi $\left(\hat{V}(\cdot,\xi)\right)$ is the projection of epi(Q^ξ)

Nous considérons ainsi la fonction des coûts futurs espérés suivante

$$\hat{V}:\begin{cases}\mathbb{R}^n\times\Xi & \to\mathbb{R}\cup\{-\infty,+\infty\}\\ (x,\xi) & \mapsto\begin{cases}\min_{y\in\mathbb{R}^m} & q^\top y\\ \text{s.t.} & Tx+Wy\leqslant h\end{cases}\end{cases}$$

et nous définissons

$$V(x):=\mathbb{E}[\hat{V}(x,\boldsymbol{\xi})].$$

Nous montrons maintenant que dans le cas général, il n'existe ni de discrétisation locale ni de discrétisation uniforme exacte et induite par une partition lorsque la matrice de recours $\boldsymbol{W}$ est stochastique à support infini:

Exemple 1.2 ($\boldsymbol{W}$ stochastique). Soit $m=\ell=1$, $\boldsymbol{q}=(1)$, $\boldsymbol{T}=(0)$, $\boldsymbol{W}=(-\boldsymbol{u})$, $\boldsymbol{h}=(-1)$ où $\boldsymbol{u}$ est une variable aléatoire uniforme sur $[1,2]$. Nous avons alors pour tout $x\in\mathbb{R}^n$:

$$\hat{V}(x,\xi)=\begin{array}{ll}\min\limits_{y\in\mathbb{R}} & y\\ \text{s.t.} & uy\geqslant 1\end{array}=\frac{1}{u}$$

La fonction $\hat{V}$ est donc constante en x et strictement convexe en ξ. Ainsi, pour toute partition $\mathcal{P}$ de Ξ:

$$\sum_{P\in\mathcal{P}}\check{p}_P\hat{V}(x,\check{\xi}_P)<V(x)$$

où $\check{p}_P=\mathbb{P}[\boldsymbol{\xi}\in P]$, $\check{\xi}_P=\mathbb{E}[\boldsymbol{\xi}\,|\,\boldsymbol{\xi}\in P]$. Pour tout $x\in\mathbb{R}^n$, il n'y a donc pas de discrétisation exacte locale en x, et a fortiori uniforme, quand $\boldsymbol{W}$ est à support infini. Nous détaillons la preuve dans l'annexe A.

Une condition nécessaire pour avoir une discrétisation exacte et uniforme est le fait qu'à chaque temps t, les fonctions V_t sont des fonctions polyédrales, c'est-à-dire qu'elle prennent leurs valeurs dans $\mathbb{R}\cup\{+\infty\}$ et que leurs épigraphes sont des polyèdres (possiblement vides). En effet, pour tout $\xi\in\text{supp}(\boldsymbol{\xi})$, la fonction $Q^\xi:(x,y)\to\ q^\top y+\mathbb{I}_{Tx+Wy\leqslant h}$ est polyédrale. Ainsi, $\hat{V}(\cdot,\xi)=\min_{y\in\mathbb{R}^m}Q^{c,\xi}(\cdot,y)$ est polyédrale car epi $(\hat{V}(\cdot,\xi))$ est une projection d'epi (Q^ξ) (voir la figure Fig. 1.8 et [JKM08]). Enfin, V est aussi polyédrale comme combinaison linéaire positive de fonctions polyédrales. Les exemples suivants montrent alors que si les contraintes n'ont pas de distributions à support fini, nous ne pouvons pas espérer avoir un théorème général de discrétisation exacte et uniforme.

Exemple 1.3 ($\boldsymbol{T}$ stochastique). Ici, et dans l'exemple suivant, $\boldsymbol{u}$ est une variable aléatoire uniforme sur $[0,1]$.

$$V(x) = \mathbb{E}\begin{bmatrix} \min\limits_{y\in\mathbb{R}^m} & y \\ \text{s.t.} & \boldsymbol{u}x \leqslant y \\ & 1 \leqslant y \end{bmatrix} = \mathbb{E}\big[\max(\boldsymbol{u}x, 1)\big] = \begin{cases} 1 & \text{if } x \leqslant 1 \\ \frac{x}{2} + \frac{1}{2x} & \text{if } x \geqslant 1 \end{cases}$$

Exemple 1.4 ($\boldsymbol{h}$ stochastique).

$$V(x) = \mathbb{E}\begin{bmatrix} \min\limits_{y\in\mathbb{R}^m} & y \\ \text{s.t.} & \boldsymbol{u} \leqslant y \\ & x \leqslant y \end{bmatrix} = \mathbb{E}\big[\max(x, \boldsymbol{u})\big] = \begin{cases} \frac{1}{2} & \text{if } x \leqslant 0 \\ \frac{x^2+1}{2} & \text{if } x \in [0,1] \\ x & \text{if } x \geqslant 1 \ . \end{cases}$$

En revanche, nous montrerons un théorème de discrétisation exacte uniforme et universelle sans aucune restriction lorsque le coût est stochastique. De plus, nous présenterons également un résultat de discrétisation locale lorsque $\boldsymbol{T}$ et $\boldsymbol{h}$ sont stochastiques, voir la table Table 1.2.

1.4 Contributions

Nous synthésisons maintenant les contributions et résultats majeurs de cette thèse.

Nous montrons tout au long de la thèse comment des notions de géométrie polyédrale peuvent éclairer la structure des problèmes d'optimisation stochastique linéaire multi-étapes. Dans le chapitre 3, nous reformulons des résultats de géométrie polyédrale et de combinatoire avec des notations utiles pour les chercheurs en optimisation stochastique. Même si les résultats de ce chapitre peuvent sembler déjà connus en géométrie polyédrale, nous pensons d'une part qu'ils sont utiles pour introduire le sujet aux non spécialistes et, d'autre part, cela permet d'expliciter des preuves formelles qui peuvent être compliquées à trouver dans la littérature.

1.4.1 Comprendre les MSLP grâce à la géométrie polyédrale

Tout d'abord, nous construisons un pont entre la géométrie polyédrale et l'optimisation stochastique.

Contribution 1. *Nous montrons que les fonctions des coûts futurs espérés sont affines par morceaux sur des complexes polyédraux explicites et universels, lorsque les contraintes sont à supports finis et que les coûts ont des distributions générales.*

En effet, nous prouvons dans le théorème 4.9 que dans le cas à 2 étapes, la fonction du coût futur espéré est affine sur chaque cellule du *chamber complex*, un complexe polyédral universel qui dépend exclusivement des contraintes. Cela nous permet de déduire des formules combinatoires pour la fonction du coût futur espéré et son sous-différentiel dans le théorème 4.9. Nous étendons ensuite ce résultat au MSLP dans le théorème 4.14 en introduisant par récurrence rétrograde des complexes polyédraux universels où les fonctions de coûts futurs espérés sont affines par morceaux, indépendamment des distributions des coûts.

Contribution 2. *Nous donnons une caractérisation explicite des fonctions de coûts futurs espérés à l'aide des fonctions d'appui des polyèdres de fibres imbriqués et pondérés, lorsque les contraintes sont déterministes et que les coûts ont des distributions générales.*

Nous définissons une généralisation du polytope de fibres, *fiber polytope* en anglais, de Billera et Sturmfels [BS92], que l'on nomme polyèdre de fibres pondéré, *weighted fiber polyhedron*, et nous montrons, pour les 2SLP, dans le théorème 4.5 que la fonction du coût futur espéré est, à une application affine près, égale à la fonction d'appui du polyèdre de fibres pondéré. Ce résultat s'étend au MSLP dans le théorème 4.17 en définissant des polyèdres de fibres imbriqués par programmation dynamique.

1.4.2 Résultats de discrétisation exacte

Dans cette thèse, nous donnons également plusieurs résultats de discrétisations exactes que nous résumons ici.

Contribution 3. *Nous montrons une discrétisation exacte locale et universelle pour les 2SLP avec recours fixé.*

En effet, nous montrons dans le théorème 4.2 que tous les raffinements de l'opposé de l'éventail normal $\mathcal{N}(P_x)$, de l'ensemble P_x des recours admissibles à x, induit une discrétisation exacte *locale et universelle* lorsque que le coût $\boldsymbol{c}$ a une distribution générale. En considérant le problème dual, nous montrons également, dans le théorème 5.3, une discrétisation *locale et universelle* pour les 2SLP avec $\boldsymbol{T}$ et $\boldsymbol{h}$ stochastiques. Nous étendons ensuite cette discrétisation exacte *locale et universelle* aux 2SLP où $\boldsymbol{q}$, $\boldsymbol{T}$ et $\boldsymbol{h}$ sont tous stochastiques dans le théorème 5.15.

	Locale et universelle	Uniforme et universelle
$\boldsymbol{W}$	×: Contre-exemple 1.2	×: Contre-exemple 1.2
$(\boldsymbol{T}, \boldsymbol{h})$	✓: $\mathcal{R}_x$ dans Thm. 5.3	×: Contre-exemples 1.3 et 1.4
$\boldsymbol{q}$	✓: $-\mathcal{N}(P_x)$ dans Thm. 4.2	✓: $\bigwedge_{\sigma\in\mathcal{C}(P,\pi)} -\mathcal{N}_\sigma$ dans Thm. 4.3
$(\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})$	✓: $\mathcal{R}_x$ dans Thm. 5.15	×: Contre-exemples 1.3 et 1.4

Table 1.2: Existence ou non de discrétisation exacte, uniforme ou locale, induite par une partition et universelle avec les contre-exemples ou les partitions associées.

Contribution 4. *Nous montrons une discrétisation exacte uniforme et universelle pour les MSLP avec des distributions des coûts générales et des contraintes à supports finis.*

Lorsque les contraintes sont déterministes, la discrétisation exacte et locale pour les 2SLP est valide pour tous les états x appartenant au même intérieur relatif d'une cellule du *chamber complex*. En prenant le raffinement commun à toutes les cellules du *chamber complex*, nous déduisons dans le théorème 4.3 une discrétisation exacte *uniforme et universelle* pour les coûts $\boldsymbol{c}$ avec distribution générale. En propageant les *chamber complexes* par programmation dynamique, nous étendons la discrétisation exacte *uniforme et universelle* au MSLP avec des contraintes à support fini dans le théorème 4.15.

1.4.3 Nouveaux algorithmes

Les nouvelles perspectives de la géométrie polyédrale et de la discrétisation exacte permettent de définir de nouveaux algorithmes, voir la table 1.3, pour résoudre des problèmes d'optimisation stochastique.

Contribution 5. *Nous présentons une méthode de partitions adaptatives généralisée (GAPM) pour résoudre les 2SLP avec $\boldsymbol{T}$ et $\boldsymbol{h}$ stochastiques et donnons une extension avec $\boldsymbol{q}$ stochastique.*

Les méthodes de partitions adaptatives (APM) ont été introduites par Song et Luedtke [SL15] pour réduire la taille des 2SLP avec support fini en rassemblant des scénari grâce à des partitions que l'on met à jour. Ramirez-Pico et Moreno ont montré dans [RPM21] que les méthodes APM pouvaient être adaptées aux distributions générales, appelant ainsi le nouvel algorithme GAPM, si l'on respecte une condition analytique suffisante. Nous présentons un algorithme, voir algorithme 5.1, qui permet d'implémenter une méthode GAPM pour toute distribution en appliquant une discrétisation exacte locale. Nous donnons aussi des résultats de

	2SLP	MSLP
Locale $(\boldsymbol{T}, \boldsymbol{h})$	GAPM dans Chap. 5	GAPDP dans Chap.6
Locale $\boldsymbol{q}$	Simplexe[6]dans Chap. 7	GAPDP dans Chap.6
Uniforme $\boldsymbol{q}$	GAPM étendu dans Chap. 5	Chamber complex dans Chap.4

Table 1.3: Algorithmes et outils pour utiliser les discrétisations exactes dans les 2SLP et MSLP

convergence et complexité pour les méthodes APM. Grâce à une implémentation en Julia, nous comparons GAPM avec les méthodes SAA en donnant des résultats numériques.

Contribution 6. *Nous fournissons un cadre algorithmique appelé Programmation Dynamique par Suivi de Trajectoire,* Trajectory Following Dynamic Programming *(TFDP) en anglais, qui englobe au moins 14 algorithmes, (voir Table 6.1) en incluant SDDP et ses variantes. Nous expliquons comment adapter tous ces algorithmes, à l'origine développés pour des distributions à support fini, dans le cas de distributions générales.*

Dans le chapitre 6, nous montrons comment toutes ces variantes rentrent dans le cadre TFDP, voir algorithme 6.1, en discutant le choix d'approximations supérieures et inférieures appelées coupes, le choix de la trajectoire dépendant du paramètre incertain, *i.e.* la sélection du noeud, et le problème d'optimisation résolu dans la phase directe.

1.4.4 Résultats de complexité

Nous présentons plusieurs résultats concernant la complexité de l'optimisation stochastique linéaire multi-étapes.

Contribution 7. *Nous montrons que les MSLP avec distributions de coûts générales peuvent être résolus approximativement en temps polynomial en* $\log(1/\varepsilon)$*, pour une grande classe de fonctions de densités régulières, lorsque* T*,* $n_2, \cdots, n_T$ *sont fixés.*

En particulier, la dimension n_1 de l'espace du premier état n'est pas fixé. La preuve repose sur la théorie de l'optimisation linéaire avec oracles [GLS12] et sur les théorèmes de bornes supérieures de McMullen [McM70] and Stanley [Sta75] sur les nombre de sommets et taille de triangulation d'un polyèdre. Ce théorème de complexité, voir corollaire 4.30, s'applique dans de nombreux cas car les distributions des coûts peut être supposées essentiellement arbitraires; nous supposons uniquement qu'elle est donnée implicitement grâce à un oracle approximé (voir la définition 4.27), cela s'applique en particulier à toutes les distributions avec une densité régulière par rapport à la mesure de Lebesgue. De plus, nous obtenons un résultat similaire de complexité polynomiale dans le modèle exact de calcul (ou modèle de Turing) pour des classes de distributions appropriées, comme les distributions uniformes ou exponentielles dans le corollaire 4.25. Pour comparaison, à notre connaissance, toutes les bornes de complexité précédentes étaient en temps pseudo-polynomial, c'est-à-dire polynomial en $1/\varepsilon$.

En outre, nous donnons des bornes de complexité sur le nombre d'itérations des méthodes APM et pour les algorithmes rentrant dans le cadre TFDP. Nous considérons un MSLP tel que, à chaque étape, nous avons un recours relativement complet, les ensembles de solutions admissibles sont de diamètres finis au plus D et de dimensions au plus d, et les fonctions des coûts futurs espérés sont Lipschitziennes de paramètre L. Soit γ l'erreur d'approximation accumulée pendant une itération.

[6]Nous donnons des pistes pour développer un simplexe d'ordre supérieur pour les 2SLP avec distribution de coût générale dans la conclusion de ce manuscrit.

Contribution 8. *Les algorithmes rentrant dans le cadre TFDP trouvent une solution ε-optimale en au plus $\left(\frac{2DL}{\varepsilon-\gamma}\right)^n (T-1)^{n+1}$ itérations, lorsqu'ils sont implémentés avec les sélections de noeud problem child ou explorative, et en au plus $(T-1)\left(\frac{4DL(T-1)}{\varepsilon-\gamma}\right)^{n+2(T-1)}$ itérations quand ils sont implémentés avec une sélection de noeud aléatoire.*

En particulier, cela fournit de nouveaux résultats de complexité, voir le corollaire 6.17 et le théorème 6.21, pour au moins 10 algorithmes dans la littérature (voir 6.1) Pour les 2SLP, nous montrons dans le théorème 5.11 que les méthodes APM, dont GAPM, trouvent une solution ε-optimale en au plus $(1+\frac{LD}{\varepsilon})^n$ itérations.

1.5 Plan de la thèse

Le reste de la thèse est rédigé en langue anglaise et suit le plan suivant : Le chapitre 2 est une traduction en langue anglaise de cette introduction.

Le chapitre 3 introduit la géométrie polyédrale pour les spécialistes d'optimisation stochastique. Il rappelle des définitions de notions standards comme les polyèdres et leurs représentations, les éventails normaux et les *chamber complexes*. Nous donnons également des preuves formelles pour les bijections entre les faces d'un polyèdre, ses cônes normaux et ses ensembles de contraintes actives. Nous définissons les notions plus combinatoires de subdivision régulière et d'éventail secondaire et montrons comment elles sont reliées au concept de contraintes actives. Nous présentons enfin une adaptation du théorème de décomposition par bases de Walkup et Wets [WW69] avec les notations utilisées dans cette thèse.

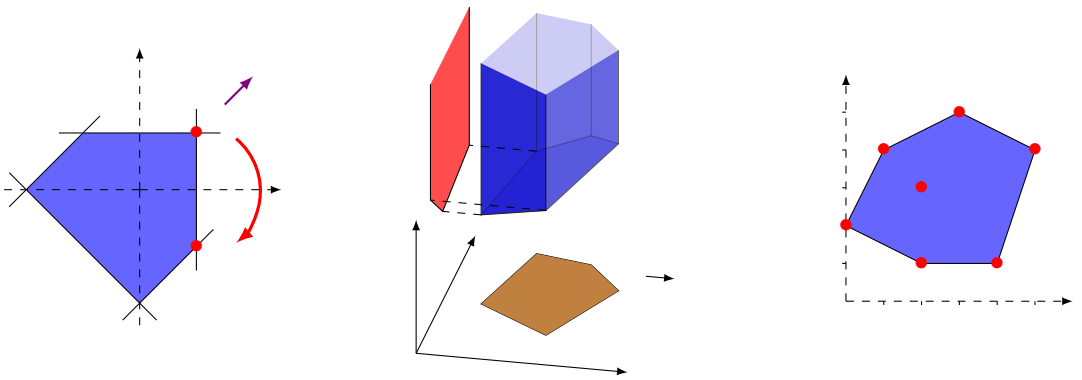
Dans le chapitre 4, nous étudions la structure polyédrale des MSLP avec distribution de coût générale. Nous montrons d'abord une discrétisation exacte uniforme et universelle pour les 2SLP qui nous permet de donner une caractérisation de la fonction du coût futur espéré. Nous donnons également une interprétation duale en définissant un nouvel objet appelé polyèdre des fibres pondéré. Ces résultats sont ensuite étendus aux MSLP. En construisant, par programmation dynamique, des complexes polyédraux universels où les fonctions des coûts futurs espérés sont affines par morceaux, nous montrons une discrétisation exacte uniforme et universelle pour les MSLP avec distributions de coûts générales. Nous en déduisons finalement des résultats de complexité pour les MSLP en montrant qu'en fixant certains paramètres, les MSLP deviennent résolubles en temps polynomiaux pour toutes distributions des coûts régulières. La plupart des résultats du chapitre 4 proviennent du papier [FGL21], qui a été récompensé du prix du meilleur papier étudiant à la conférence ECSO-CMS 2022. De nouveaux résultats complémentaires sont également ajoutés.

Le chapitre 5 traite des méthodes de partitions adaptatives (APM), qui sont des méthodes numériques qui résolvent en particulier des 2SLP. Nous disons qu'une partition de l'espace des incertitudes est *adaptée* à l'état de la première étape $\check{x}$ si elle induit une discrétisation exacte locale en $\check{x}$. L'idée centrale des APM est de construire itérativement une partition adaptée à tous les états que l'on a déjà testés. En s'appuyant sur l'éventail normal de l'ensemble des solutions duales admissibles, nous présentons une discrétisation exacte et locale lorsque la matrice $\boldsymbol{T}$ et le second membre des contraintes $\boldsymbol{h}$ ont des distributions générales. Cela nous permet de déduire une condition nécessaire et suffisante pour qu'une partition soit adaptée même dans le cas des distributions à support infini, et de fournir un oracle géométrique pour construire une partition adaptée. En outre, en montrant les liens entre les APM et l'algorithme L-shaped, nous prouvons la convergence et donnons des bornes de complexité pour les méthodes APM. Nous présentons des résultats numériques comparant GAPM au méthodes SAA. Pour finir, nous étendons l'algorithme GAPM au cas où le coût a une distribution générale. La majorité de ce chapitre peut être trouvé dans l'article [FL22b], publié dans Operations Research Letters, mais l'extension aux coûts à distributions générales est écrite dans la section 3 du papier [FL22a].

Dans le chapitre 6, nous introduisons une classe d'algorithmes appelée Programmation Dynamique par Suivi de Trajectoires, *Trajectory Following Dynamic Programming* (TFDP) en anglais, qui raffine itérativement des approximations des fonctions des coûts futurs espérés pour les MSP avec variables aléatoires indépendantes. Ce cadre algorithmique englobe la plupart des variantes de l'algorithme SDDP. En supposant le caractère Lipschitz des fonctions des coûts futurs espérés, nous fournissons une nouvelle preuve de convergence et de complexité qui permet de considérer des variables aléatoires à supports infinis. En particulier, nous en déduisons de nouveaux résultats de complexité pour de nombreux algorithmes connus. De plus, nous détaillons comment les algorithmes TFDP peuvent être implémentés sans l'hypothèse de support fini, soit grâce à des approximations soit grâce à des discrétisations exactes. Ce chapitre présente les résultats du papier [FL22a], à l'exception de la section 3 qui est présentée dans le chapitre précédent.

Nous concluons et donnons quelques perspectives pour des travaux futurs dans le chapitre Chapter 7.

2 Introduction



Contents

Optimization is the field of mathematics that studies the minimum of a function, known as the objective, over a set. From an applied point of view, optimization model many situations where a decision maker has to choose the best solution, with regard to some objective, while satisfying various constraints. In the 17th century, the differential calculus of Leibniz and Newton opened the door to methods to find solutions to unconstrained and smooth optimization problems. Then, mathematicians of the 18th century such as Euler and Lagrange provided methods to derive solutions of constrained optimization problems. Since the 20th century, the rise of computer science has allowed implementing practical algorithms to solve nonsmooth optimization problems. For instance, linear programming is a subfield of optimization which is widely applicable in practice and can be solved very efficiently. Dantzig [DT03] developed a general linear programming formulation and invented the simplex method, which is now better understood thanks to polyhedral geometry. More generally, in nonsmooth convex optimization, mathematicians such as Fenchel and Rockafellar [Roc15] discovered new links between optimization and geometry. Thus, the bridges between optimization and geometry led to improvements on the efficiency of existing algorithms and to the inventions of new algorithms. In this thesis, we will wonder what are the links between stochastic optimization and polyhedral geometry.

In this chapter, we first introduce the field of stochastic optimization by recalling dynamic programming principles and different approximate quantization methods in Section 2.1. After giving an overview of the polyhedral geometry notions used throughout this thesis and their original motivation in Section 2.2, we introduce the exact quantization problem in Section 2.3 which is the main question tackled in this work. The main contributions of this thesis are presented in Section 2.4 and its structure in Section 2.5.

2.1 Stochastic optimization

In optimization under uncertainty, one has to make a decision without knowing precisely some key parameters of the problem. The uncertainty can come from a lack of information, an imprecision in measures or a difficulty to model a problem. The most common case is when the decision maker has to make a choice here and now but the criterion she wants to optimize depends on a parameter she will only observe in the future.

There exists various approaches to decision under uncertainty. Typically we either assume that the uncertainty is adversarial or stochastic. In robust optimization [BTEGN09, BBC11], the decision maker has to choose the optimal solution while considering the worst case in an uncertainty set. In reinforcement learning [KLM96, SB18] or bandit optimization [BCB$^+$12], we assume that the uncertainty is random but that the decision maker does not know the distribution and has to learn it.

Stochastic optimization considers problems where the decision maker minimizes a real-valued function, sometimes called risk measure, depending on the uncertain parameters modeled as random variables. A part of stochastic optimization deals with distributionally robust optimization problems [WKS14, RM19, ZKW21] where we consider the worst possible distribution over a set of distributions. Risk averse stochastic programs using a coherent risk measures [ADEH99] are equivalent to distributionally robust optimization problems. The most common and natural criterion to optimize is the expectation of the cost. When we choose the expectation as the risk measure to optimize, we say that we are in a risk neutral setting. In this thesis, we will consider essentially[1] risk neutral problems.

Thus, stochastic programming is a powerful modeling paradigm that has found many applications in energy, logistics or finance (see e.g. [Dup02, WZ05, GZ13] and references therein). We refer to the books [SDR14, RW21, BL11, KM$^+$76] for a more complete presentation on stochastic optimization.

[1] We only discuss risk averse extensions in Section 6.4

2.1.1 Multistage stochastic optimization

Multistage stochastic problems (MSP) constitute an important class of stochastic programs where decisions x_t are taken sequentially during stages. Between stages t and $t+1$, some part of the uncertainty ξ_t is revealed.

$$x_0 \rightsquigarrow \xi_1 \rightsquigarrow x_1 \rightsquigarrow \xi_2 \rightsquigarrow \cdots \rightsquigarrow x_{T-1} \rightsquigarrow \xi_T \rightsquigarrow x_T$$

The uncertainty parameters are modeled as random variables, and we will denote them as bold letters along this thesis. We now present informally[2] a general setting for MSP.

$$\begin{aligned} \min_{(\boldsymbol{x}_t)_{t\in[T]}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{x}_t, \boldsymbol{\xi}_t)\right] && && \text{(MSP)} \\ s.t. \quad & \boldsymbol{x}_0 = x_0 && a.s. & \\ & \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) && a.s. & \forall t \in [T] \\ & \sigma(\boldsymbol{x}_t) \subset \sigma(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_t) && & \forall t \in [T] \end{aligned}$$

At each time step t, the decision maker chooses the solution $\boldsymbol{x}_t$ in order to minimize the actual loss function ℓ_t, which depends on the choice $\boldsymbol{x}_t$ and on a realization of a random variable $\boldsymbol{\xi}_t$, plus the expected future losses. The first constraint $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ models that the decision $\boldsymbol{x}_t$ must satisfy some constraints which depends on the past decision $\boldsymbol{x}_{t-1}$ and the actual random parameter $\boldsymbol{\xi}_t$. The last constraint, called non-anticipativity constraint, ensures that the decision maker can only base her decision on the past and present observed parameters and not on the unobserved future parameters. Here, σ refers to the σ-algebra induced by a random variable. For notational consistency, we assume that the first random variable $\boldsymbol{\xi}_1$ is fixed to a certain deterministic value ξ_1 almost surely and that the state x_0 is fixed.

Example 2.1 (MSP modeling hydroelectric energy storage management). We consider a hydroelectic storage. The uncertain parameters are the demand in electricity and the water inflow from rain and runoff. The decision maker chooses the quantity of water hustled each day to optimize some cost, under physical constraints, for example ensuring that the dam must not overflow. To model this problem, we can define the following MSP

$$\begin{aligned} \min_{(\boldsymbol{u}_t)_t, (\boldsymbol{v}_t)_t} \quad & \mathbb{E}\left[\sum_{t=1}^{T} c_t(\boldsymbol{d}_t - \boldsymbol{u}_t)\right] && && \\ s.t. \quad & \boldsymbol{x}_0 = x_0 && a.s. & \\ & 0 \leqslant \boldsymbol{u}_t \leqslant \boldsymbol{d}_t && a.s. & \forall t \in [T] \\ & \boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{w}_t - \boldsymbol{u}_t - \boldsymbol{v}_t && a.s. & \forall t \in [T] \\ & 0 \leqslant \boldsymbol{x}_t \leqslant \overline{x} && a.s. & \forall t \in [T] \\ & 0 \leqslant \boldsymbol{v}_t && a.s. & \forall t \in [T] \\ & \sigma(\boldsymbol{u}_t) \subset \sigma(\boldsymbol{w}_1, \boldsymbol{d}_1 \cdots, \boldsymbol{w}_t, \boldsymbol{d}_t) && & \forall t \in [T] \end{aligned}$$

where $\overline{x}$ is the capacity of the reservoir, T is the number of time steps in the period we consider, and at step t, $\boldsymbol{x}_t$ represents the quantity of water in the reservoir, $\boldsymbol{u}_t$ the quantity of water hustled, $\boldsymbol{v}_t$ quantity of water evacuated by opening the valve, $\boldsymbol{d}_t$ the demand, $\boldsymbol{w}_t$ the quantity of water coming from rain and runoff and c_t the cost of unmet demand.

[2] This is formally presented in Section 6.1

2.1.2 Dynamic programming

Multistage stochastic problems are notoriously difficult to solve and obtaining solutions, even approximate, in reasonable time, is out of reach in the general case. Indeed, without further assumptions, optimal solution are given as a function of past noises which leads, after quantization, to a number of variables exponential in the horizon. A usual assumption that allows information compression and decomposition, is the stagewise independence of uncertainty parameters. More formally, we assume that $(\boldsymbol{\xi}_t)_{t\in[T]}$ is a sequence of independent random variables.

To decompose it, we see Problem MSP as an element in a family of parametrized optimization problems given by

$$\begin{aligned} V_t(x_{t-1}) := \min_{(\boldsymbol{x}_\tau)_{t\leqslant\tau\leqslant T}} \quad & \mathbb{E}\left[\sum_{\tau=t}^{T} \ell_\tau(\boldsymbol{x}_\tau, \boldsymbol{\xi}_\tau)\right] \\ s.c. \quad & \boldsymbol{x}_{t-1} = x_{t-1} && a.s. \\ & \boldsymbol{x}_\tau \in \mathcal{X}_\tau(\boldsymbol{x}_{\tau-1}, \boldsymbol{\xi}_\tau) && a.s. \quad \forall t \leqslant \tau \leqslant T \\ & \sigma(\boldsymbol{x}_\tau) \subset \sigma(\boldsymbol{\xi}_t, \cdots, \boldsymbol{\xi}_\tau) && \forall t \leqslant \tau \leqslant T. \end{aligned}$$

where V_t is known as the *expected cost-to-go* function at time t.

Leveraging the stagewise independence assumption, we have, at all $t \in [T]$, the so-called Bellman or dynamic programming equations (see [Bel66, Ber12]) that relate the cost-to-go functions at time t and $t+1$:

$$V_t(x_{t-1}) = \mathbb{E}\Big[\hat{V}_t(x_{t-1}, \boldsymbol{\xi}_t)\Big] \tag{2.2a}$$

$$\hat{V}_t(x_{t-1}, \xi_t) := \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \ell_t(x_t, \xi_t) + V_{t+1}(x_t) \tag{2.2b}$$

where we set $V_{T+1} :\equiv 0$ for notational consistency.

Finally, note that, we can also rewrite Problem (MSP) in the following nested form:

$$\min_{x_1 \in X_1(x_0, \xi_1)} \ell_1(x_1, \xi_1) + \mathbb{E}\left[\min_{x_2 \in X_2(x_1, \boldsymbol{\xi}_2)} \ell_2(x_2, \boldsymbol{\xi}_2) + \mathbb{E}\Big[\cdots + \mathbb{E}\big[\min_{x_T \in \mathcal{X}_T(x_{T-1}, \boldsymbol{\xi}_T)} \ell_T(x_T, \boldsymbol{\xi}_T)\big]\Big]\right].$$

The dynamic programming principle tells us that there exists an optimal solution given as a function of the current state and noise. For comparison, recall that, without the independence assumption, the solution depends on the whole history of past noises. Thus, dynamic programming allows compressing sufficient information in the state.

Futhermore, it decomposed one big T-stage problem (2.1) into a sequence of T integration problems and T parametric deterministic 1-stage problems. More precisely, to compute $V_t(x_{t-1})$ for all x_{t-1}, we would first need to compute $\hat{V}_t(x_{t-1}, \xi_t)$ by solving deterministic 1-stage problems (2.2b) each parametrized by the state x_{t-1} and the uncertainty parameter ξ_t. Secondly, we would need to compute an integral since $V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}(x_{t-1}, \boldsymbol{\xi}_t)\big]$. We now explain how to do so when the number of possible states and noises is finite.

When both the decision variables x_t and the uncertainty parameters ξ_t live in finite sets X_t and Ξ_t, a framework also known as finite Markov decision processes, we can exactly compute V_t by backward induction. Assuming that $V_{t+1}(x_t)$ is known for all $x_t \in X_t$, we loop over each possible $x_{t-1} \in X_{t-1}$, and $\xi_t \in \Xi_t$ to compute $\hat{V}_t(x_{t-1}, \xi_t)$ as $\min_{x_t \in \mathcal{X}_t(x_{t-1})} \ell_t(x_t, \xi_t) + V_{t+1}(x_t)$. The minimimum is done by brute force, and $V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}(x_{t-1}, \boldsymbol{\xi}_t)\big]$ is a finite average of $\hat{V}_t(x_{t-1}, \xi_t)$. This algorithm is called Stochastic Dynamic Programming (SDP) see Algorithm 2.1.

However, in many applications, the state space where x_t lives is continuous. We then consider a more general setting where x_t can take an infinite number of values. Since this general problem is difficult to solve, a first interesting idea is to discretize. Instead of solving directly the MSP, we solve by applying SDP, an approximate version of MSP where the decision variables x_t belongs

$V_{T+1} \equiv 0$;
for t *from* T *to* 1 **do**
  for $x_{t-1} \in X_{t-1}$ **do**
    $v \leftarrow 0$;
    for $\xi_t \in \Xi_t$ **do**
      $m \leftarrow +\infty$;
      for $x_t \in X_t$ **do**
        if $x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)$ **then**
          $m \leftarrow \min\Big(m, \ell_t(x_t, \xi_t) + V_{t+1}(x_t)\Big)$;
        end
      end
      $v \leftarrow v + m \times \mathbb{P}\big[\boldsymbol{\xi}_t = \xi_t\big]$;
    end
    $V_t(x_{t-1}) \leftarrow v$;
  end
end

Algorithm 2.1: Stochastic Dynamic Programming (SDP) for discretized version of MSP.

in a finite set X_t. The precision of the approximation increases with the number of computed points. In particular, the size of the discretized sets, needed to obtain a certain precision, is exponential in the dimension of the initial space. Different approaches were developed to solve MSP while avoiding this phenomenon called curse of dimensionality.

Instead of doing one backward induction computing $V_t(x_{t-1})$ for all x_{t-1} in a precomputed discretized grid, we can approximate $V_t(x_{t-1})$ along some adaptive trajectories. The Stochastic Dual Dynamic Programming (SDDP) algorithm, introduced by Pereira and Pinto in [PP91] to manage the hydroelectric system in Brazil, iterates over forward pass to compute trajectories and backward path to update the estimation of expected cost-to-go functions along said trajectories through duality results. There are many variants of SDDP that we present and discuss in Chapter 6, we encompass all these variants under the name Trajectory Following Dynamic Programming algorithm (TFDP). Thus, TFDP algorithms allows considering infinite sets of decision X_t.

However, in the current literature, TFDP algorithms work when the random variables $\boldsymbol{\xi}_t$ have a finite support. Indeed, without the finite support assumption for random variables, we cannot hope to compute $\hat{V}_t(x_{t-1}, \xi_t)$ because it would require to solve an infinite number of 1-stage problems (2.2b) parametrized by the uncertainty parameter ξ_t. To avoid this difficulty, we can replace the general distribution of $\boldsymbol{\xi}_t$ by a finitely supported distribution. The different approaches in the literature to reduce the stochastic parameters ξ_t to a finite number of scenarios are known as quantization methods.

2.1.3 Quantization in stochastic optimization

We now present different quantization methods commonly used in stochastic optimization to approximate MSP with general distributions.

Sample Average Approximation The first idea is to replace the general random variables $\boldsymbol{\xi}_t$ by a finite number of scenarios $(\check{\xi}_{t,k})_{k\in[N]}$ sampled independently and identically distributed according to the distribution of $\boldsymbol{\xi}_t$. As in Monte-Carlo methods, we replace the true unknown probability distribution $\mathbb{P}$ by the empirical probability $\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{k=1}^{N} \delta_{\check{\xi}_{t,k}}$. Then, the expectation $\mathbb{E}\big[f(\boldsymbol{\xi}_t)\big]$ is replaced by the finite mean $\frac{1}{N_t}\sum_{k=1}^{N} f(\check{\xi}_{t,k})$. For example, the dynamic

programming equation (2.2) becomes

$$V_t(x_{t-1}) = \frac{1}{N}\sum_{k=1}^{N}\Big(\underbrace{\min_{x_t \in \mathcal{X}_t(x_{t-1},\check{\xi}_{t,k})} \ell_t(x_t,\check{\xi}_{t,k}) + V_{t+1}(x_t)}_{:=\hat{V}(x_{t-1},\check{\xi}_{t,k})}\Big).$$

We then obtain a new MSP with finitely supported random variables following their empirical distributions. This new problem is called Sample Average Approximation (SAA), see [SDR14, Chapter 5] for a comprehensive review of the subject. By the Law of Large Numbers and under some mild regularity assumptions, the SAA problem converges to the true MSP, in many senses, for example the convergence of the value or of the set of optimal solutions. Assuming the existence of a variance, by the Central Limit Theorem, confidence intervals for the value of the MSP are also derived.

The first advantage of SAA is that it only requires mild assumptions and work on a large class of stochastic optimization setting. Secondly, SAA is very easy to implement in practice, as long as we know how to sample the random variables. However, these convergence results are either asymptotic or converging in $\frac{1}{\sqrt{N}}$ while staying probabilistic. Note that the value v_N^{SAA} of the SAA problem is a random variable (depending on the sample), and thus can be seen as a statistical estimator of the value of the true problem. If this estimator is consistent (asymptotically almost surely converging), it is biased (the expectation of v_N^{SAA} is lower than the true value), but we cannot guarantee that a given realization provides a lower bound for the true value. In the same idea, there also exist quasi Monte-Carlo methods which generate samples through deterministic methods in the hope to reduce variance while conserving asymptotic convergence.

Edmundson Madansky and Jensen based approximations We now present different approximations method, under convexity assumptions, leveraging inequalities known as Jensen's and Edmundson-Madansky's. We present more formally this inequalities in section 6.2.4 and refer to [Kuh06, EZ94b, EZ94a, KM$^+$76] for a detailed introduction on the topic. When a function f is convex in a stochastic parameter $\boldsymbol{\xi}$, the Jensen inequality tells us that $f(\mathbb{E}[\boldsymbol{\xi}]) \leqslant \mathbb{E}\big[f(\boldsymbol{\xi})\big]$. This inequality then provides a lower bound on the value of Multistage Stochastic Convex Programs (MSCP) by replacing the random variables $\boldsymbol{\xi}_t$ by deterministic variables equal to their expectations $\mathbb{E}[\boldsymbol{\xi}_t]$. This new problem known as the expected value problem (see e.g. [MAB14]) then gives a lower bound on the true value of MSCP. However, this lower bound can be very far from the real value. To improve the lower bound, we can partition the set Ξ where the random variable $\boldsymbol{\xi}$ lives and leverage the Jensen inequality at each set P of the partition $\mathcal{P}$. We then consider another approximated problem where the random variable $\boldsymbol{\xi}$ is replaced by a finitely supported random variable $\check{\xi}$ with scenarios $\check{\xi}_P := \mathbb{E}[\boldsymbol{\xi} \,|\, \boldsymbol{\xi} \in P]$ associated to the probabilities $\check{p}_P := \mathbb{P}[\boldsymbol{\xi} \in P]$. In that case, the finer is the partition $\mathcal{P}$, the better is the approximation. Moreover, when the diameters of the sets $P \in \mathcal{P}$ goes to 0, the approximated value converges toward the value of the true MSCP. When the random variables belongs to a polyhedron Q, we have an upper bound on the expectation $\mathbb{E}\big[f(\boldsymbol{\xi})\big]$ by taking a convex combination of the values $f(v)$ where v describes the vertices of the polyhedron Q. This inequality, called Edmundson-Madansky inequality (see [Mad59]), allows deriving to the same kind of approximation schemes with partitions to get converging upper bounds. Finally, when the loss functions are saddle functions, i.e. both convex in a random variable $\boldsymbol{\xi}$ and concave in another random variable $\boldsymbol{\eta}$, leveraging the ideas of both inequalities, we can derive upper and lower approximations of the problem MSCP. The advantage of these kinds of approximations is the guarantee to have lower and upper bounds. However, these approximations are harder to implement than SAA, and need a convexity assumption.

Dependence based approximations In the case where the random variables have dependence, we can approximate these dependencies. The idea consists in relaxing the non-

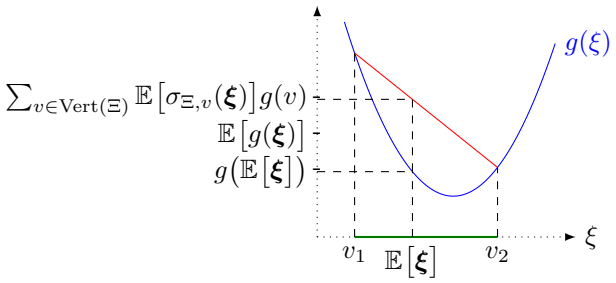


Figure 2.2: An illustration of Jensen and Edmundson-Madansky inequalities, see also Proposition 6.7

anticipativity constraints, for example by considering that the decision maker can see in the future, or in gathering different scenarios together and representing them by their mean. For example, the progressive hedging algorithm presented by Rockafellar an Wets in [RW91] is based on these ideas. In the convex case, leveraging Jensen and Edmundson-Madansky inequalities, we can show lower and upper bounds based on barycentric scenario trees (see [Fra96, CS05, Kuh06]). In [PP12], Pflug and Pichler defined a distance, called the nested distance, between dependent sequences of random variables. This nested distance allows to bound the gap between a MSP and its approximated problems and thus to provide good approximation of MSP.

Note that, in risk neutral stochastic optimization, we generally assume that the distributions of random variables are known by the decision maker. This assumption makes stochastic programs dependent on the choice of the distribution while modeling a practical problem. The impact of this choice is studied by the field of sensitivity analysis [Dup90, RS91] which consists in understanding how the value of a stochastic problem varies with small perturbation of the distribution. In practice, it is often not necessary to solve precisely, or exactly, a stochastic program since the modeling error can overshadow the optimization or approximation error. However, from a theoretical point of view, providing mathematical insights on the optimal solutions of a problem with known probability distributions opens the door to improvements of existing algorithms and ideas to develop new algorithms.

In this thesis, instead of the above approaches, we develop exact quantization leveraging polyhedral geometry tools.

2.2 Polyhedral geometry, combinatorics and complexity

Optimization has found fruitful bridges with the fields of geometry and combinatorics. We first recall some standard connections between linear problems and polyhedral geometry. We then present briefly "higher order" tools in polyhedral geometry, such as the secondary fan which will be a key tool in this thesis, and their link with parametric linear programming. These more fundamental notions are originally motivated from polynomial systems and elimination theory, a topic we will not consider in this thesis, but that is briefly evoked in Section 2.2.3. Finally, after introducing computational complexity theory, we recall complexity results in polyhedral geometry and stochastic programming.

2.2.1 Linear optimization and polyhedra

Linear optimization is the subfield of optimization where the objective function is linear and the constraint are defined as affine inequalities (see [MG07]). A polyhedron is a set of points which is equal to the intersection of a finite number of halfspaces. We give a more formal definition of polyhedron in Section 3.1 and refer to [Zie12, Grü13] for a complete introduction on polyhedral

geometry. Thus, the set of admissible solutions of a linear problem (LP) is a polyhedron. A key result in linear optimization is that the infimum of a linear function on a polyhedron is attained on at least one vertex, i.e. an extreme point of the polyhedron, whenever the polyhedron admits vertices and the infimum is finite. The simplex method is one of the most efficient and commonly used algorithms to solve LP. It consists in following a path on the vertices of the polyhedron of admissible solutions while decreasing the objective function.

$$
\begin{array}{lll}
\min\limits_{x\in\mathbb{R}^2} & c^\top x := x_1 + x_2 & \\
\text{s.t.} & x_1 \leqslant 1 & (1)\\
& x_1 - x_2 \leqslant 2 & (2)\\
& -x_1 - x_2 \leqslant 2 & (3)\\
& -x_1 + x_2 \leqslant 2 & (4)\\
& x_2 \leqslant 1 & (5)
\end{array}
$$

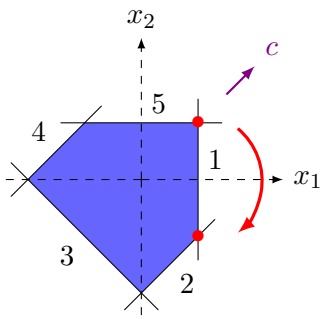


Figure 2.3: Illustration of a step of the simplex algorithm for a simple linear problem. We pivot from the basis $\{1, 5\}$ to the basis $\{1, 2\}$.

To implement the simplex method in a computer, we need more combinatorial notions. A basis is a set of indices, of the rows or columns of a matrix, which yields an invertible submatrix. The vertices of a polyhedron can be retrieved and computed thanks to the bases. Then from a combinatorial point of view, the simplex method goes from basis to basis thanks to a pivot rule, see Fig. 2.3. The links between polyhedral geometry and linear programming are widely studied, for example to improve the performance of the simplex method. Fundamental complexity problems for linear programming such as the 9th problem of Smale [Sma98] are thus heavily linked with polyhedral geometry fundamental conjectures such as the Hirsch conjecture [San12].

2.2.2 Higher order polyhedral notions

The parametric linear programming point of view consists in studying the behavior of linear problems when parameters, such as the cost vector, the constraint matrix and the constraint right-hand side vector, vary. A combinatorial approach of parametric linear programming was developed by Walkup and Wets [WW69]. Their basis decomposition theorem describes how the value of a linear program in standard form varies with respect to the cost and the right-hand side of the constraints. A more recent approach by Sturmfels and Thomas [ST97] linked this theorem with the theory of regular subdivisions. Regular subdivisions are combinatorial objects which describes the behavior of finite sets of points (resp. vectors) called configurations of points (resp. vectors). We give a formal definition of regular subdivisions in Definition 3.17 and refer to the monograph [DLRS10] for a complete introduction on this theory. Informally, the regular subdivision of a vector configuration $A = (a_i)_{i\in[n]}$ associated with a height vector $\omega \in \mathbb{R}^n$ is the collection of sets of indices $I_F = \{i \in [n] \,|\, a_i \in F\}$ when F describes the lower faces of the lifted cone $\mathrm{Cone}(a_i, \omega_i)_{i\in[n]}$ see Fig. 2.4. The structure of regular subdivisions is better understood thanks to a fundamental geometric notion called secondary fan[3] Σ-fan(A). More precisely, two height vectors give the same regular subdivision if and only if they belong in the same relative interior of a secondary cone (i.e. a cone of the secondary fan).

This fan happens to be the normal fan of a certain polytope called secondary polytope and denoted $\Sigma(A)$. The vertices of $\Sigma(A)$ can be retrieved with the regular triangulations of A seen as a vector configuration, thanks to the so called GKZ coordinates (for Gelfand, Kapranov

[3] A fan is a polyhedral complex whose cells are cones.

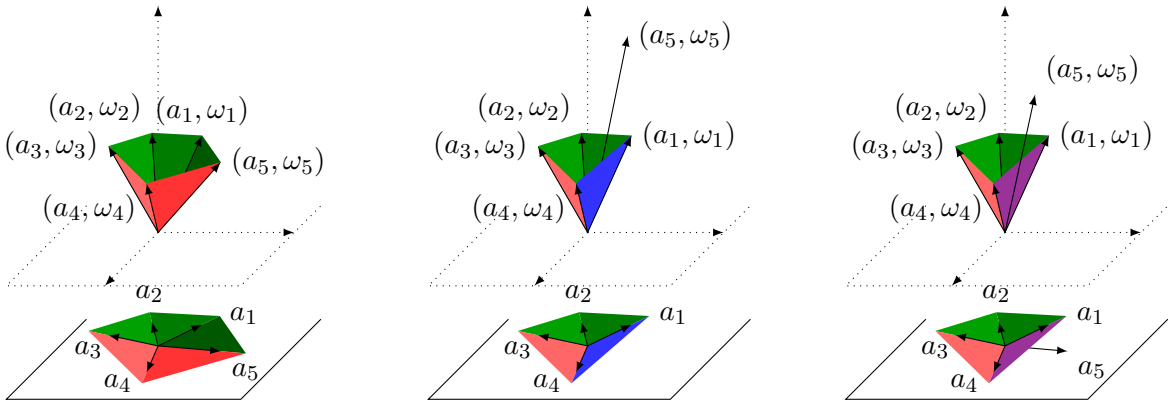


Figure 2.4: Three lifted vector configurations and the projection of their lower faces, see Section 3.3.1 for more details.

and Zelevinsky, see next section). In [BS92], Billera and Sturmfels defined another fundamental object called the fiber polytope, to enlighten the structure of a more general class of subdivisions. They also showed that the secondary polytope was a special case of fiber polytope.

2.2.3 Motivation from polynomial systems and elimination theory

These advanced and recent higher order notions, used throughout this thesis in the context of stochastic programming, are motivated, in another context, from the study of polynomial systems.

In elimination theory, the *resultant* $R(P, Q)$, or Sylvester determinant, of two univariate polynomials P and Q, is a scalar which equals 0 if and only if P and Q have a common root. The resultant of P and Q is defined as the determinant of a matrix which depends on the coefficients of P and Q. Thus, the function resultant $R_d : (P, Q) \mapsto R(P, Q)$, restricted to polynomials of degrees smaller than d, can be seen as a multivariate polynomial in the coefficients of P and Q. The *discriminant* $\Delta_d(P)$ of a univariate polynomial P is defined as the resultant $R(P, P')$ of P and its derivative P'. Like the resultant, the function discriminant Δ_d, restricted to polyhedron of degree smaller than d, can also be seen as a multivariate polynomial. These objects are fundamental in elimination theory to solve polynomials systems.

We now consider the multivariate setting. Let $A \subset \mathbb{N}^n$ be a finite set of non-negative integer vectors. Both notions of resultants and discriminants can be extended to multivariate polynomials with non-null coefficients in A, we then speak of A-resultant and A-discriminant, denoted R_A and Δ_A (see [GKZ94, Chapter 8 and 9]). Finally, to $P := \sum_{a \in A} c_a X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$, a multivariate polynomial with exponent in A, we can associate its *Newton polytope* $\mathrm{Nw}(P) := \mathrm{Conv}\{a \in A \,|\, c_a \neq 0\}$, see Fig. 2.5. The Newton polytope is a key notion to study the behavior of multivariate polynomials (see [EK08, MTY19]).

In [GKZ90] (see also [GKZ94, Chapter 13]), Gelfand, Kapranov and Zelevinski showed that the discriminant Δ_A had strong connection with the theory of regular subdivisions. More precisely, they proved that the Newton polytope $\mathrm{Nw}(\Delta_A)$ of the discriminant Δ_A was equal, up to a scalar factor, to the secondary polytope $\Sigma(A)$. Sturmfels subsequently showed in [Stu94] that the Newton polytope $\mathrm{Nw}(R_A)$ of the resultant was a summand of a special fiber polytope.

2.2.4 Computational complexity

Computation complexity theory classifies mathematical problems according to how hard they are to solve. A decision problem D can be proven at least as hard as a problem D' by showing that every instance of D' is equivalent to an instance of D. We say that D' can be reduced to D. On the contrary, to prove that a problem D can be solved in reasonable time, we can study

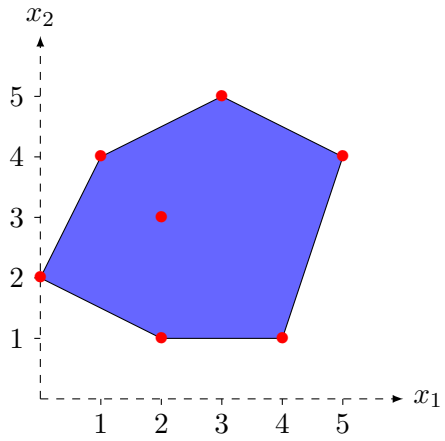


Figure 2.5: The Newton polytope $\mathrm{Nw}(P)$ of the polynomial $P := X_1^3X_2^5 + 3X_1^5X_2^4 - 7X_1^4X_2 + X_1^2X_2 - 8X_2^2 + X_1X_2^4 - 4X_1^2X_2^3$

the complexity of any algorithm that solves D. We refer to Arora and Barak monograph [AB09] for an introduction on complexity theory.

In [Kha80], Khachiyan showed that the ellipsoid method solves linear programs in polynomial time, thus proving that linear programming was in the complexity class P. We refer to the book [GLS12] of Grötschel, Lovász and Schrijver for an introduction on complexity theory for linear programs. However, adding some constraints, like having integer admissible solutions, or modifying the structure of the problem, with stochasticity, changes a lot the complexity class. Indeed, it was proven in [vzGS78], that integer linear programming (ILP) was NP-hard. When a problem has a difficult complexity status, a rather common approach, and often the best one can hope for, is to show that suitably restrictions with fixed parameters of the problem become polynomial. For example, Lenstra proved in [LJ83] that, when the dimension of the set of admissible points is fixed, integer linear programming becomes polynomial. Similarly, it was proven that counting the number of integer points in a polytope was a ♯P-complete problem. Barvinok then showed in [Bar94] (see also [Bar08]) that when the dimension is fixed, counting the number of integer points in a polytope becomes a polynomial problem. Perhaps, more surprisingly, the problem of computing the volume of a polytope is comparable to counting the number of integer points, i.e. ♯P-complete in general [DF88] and polynomial when the dimension is fixed.

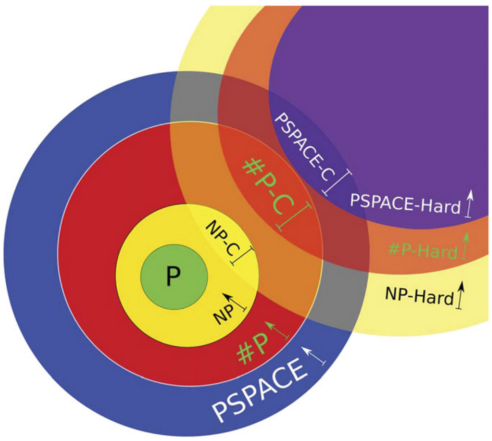


Figure 2.6: Different classes of complexity. The figure is borrowed from [IJCN15].

The complexity of stochastic programming has been extensively studied. In [DS06], Dyer and Stougie proved that a particular case of MSLP was PSPACE-hard, and that 2-stage stochastic programming with discrete distributions is ♯P-hard, by reducing to it the problem of graph reliability. Hanasusanto, Kuhn and Wiesemann [HKW16] then showed that solving, with a sufficiently high accuracy, the 2-stage linear programming (2SLP) with continuous distribution is also ♯P-hard, exploiting the ♯P-completeness of the computation of the volume of knapsack polytopes and order polytopes. Several papers such as [Sha06] studied the complexity of 2SLP and MSLP by analyzing the convergence of SAA methods. Shapiro and Nemirovski [SN05] showed in particular that the 2SLP problem can be solved by the SAA method, with high probability and up to precision ε, in a time which is pseudo-polynomial, *i.e.* polynomial in $1/\varepsilon$ and in the input size when the dimensions of decision spaces are fixed. This entails that 2SLP problems can be solved in a scalable way with a moderate accuracy. Finally, Lan [Lan20] and Zhang and Sun [ZS19] independently analysed the complexity of Stochastic Dual Dynamic Programming (SDDP). It follows from their results that finitely supported MSLP can be solved approximately in pseudo-polynomial time in the error approximation ε when all the dimensions are fixed. In other words the complexity of these SDDP methods is polynomially bounded in $1/\varepsilon$.

2.3 The exact quantization problem

In this thesis, we wonder how to deal with general (thus non-finitely supported) distribution in stochastic programming. The main question is then

How to solve *exactly* multistage stochastic problems with *general distributions* ?

In other words, do there exist analytical formulas and algorithms to compute exact solutions of stochastic programs with general distribution ?

We saw in Section 2.1.3 that we can approximate a MSP, either by SAA or by Jensen and Edmundson-Madansky inequalities, by replacing the continuous random variables by finitely supported random variable, see Table 2.1. These techniques can be seen as approximate quantization method, we now define the notion of exact quantization.

Definition 2.1 (Exact quantization)**.** *We say that a MSP (with stagewise independence) admits a* local exact quantization *at time t on x_{t-1} if there exists a finitely supported $(\check{\boldsymbol{\xi}}_t)_{t\in[T]}$ that yields the same expected cost-to-go functions* i.e. *such that*

$$V_t(x_{t-1}) = \mathbb{E}\big[\hat{V}_t(x_{t-1}, \boldsymbol{\xi}_t)\big] = \mathbb{E}\big[\hat{V}_t(x_{t-1}, \check{\boldsymbol{\xi}}_t)\big].$$

In particular, we have $V(x_{t-1}) = \sum_{\xi\in\mathrm{supp}(\check{\boldsymbol{\xi}}_t)} \mathbb{P}\big[\boldsymbol{\xi}_t = \xi\big]\hat{V}(x_{t-1}, \xi)$.

We call a quantization uniformly exact *if it is locally exact at all $x_{t-1} \in \mathbb{R}^{n_{t-1}}$, and all $t \in [T]$.*

We say that a quantization is partition based *if there exists a (measurable, finite) partition $\mathcal{P}_{t,x_{t-1}}$ of the uncertainty set Ξ_t such that, for $P \in \mathcal{P}_{t,x_{t-1}}$,*

$$\mathbb{P}\big[\check{\boldsymbol{\xi}}_t = \check{\xi}_{t,P}\big] = \check{p}_{t,P} \quad \text{with} \quad \check{p}_{t,P} := \mathbb{P}\big[\boldsymbol{\xi}_t \in P\big], \quad \check{\xi}_{t,P} := \mathbb{E}\big[\boldsymbol{\xi}_t \,|\, \boldsymbol{\xi}_t \in P\big]^4.$$

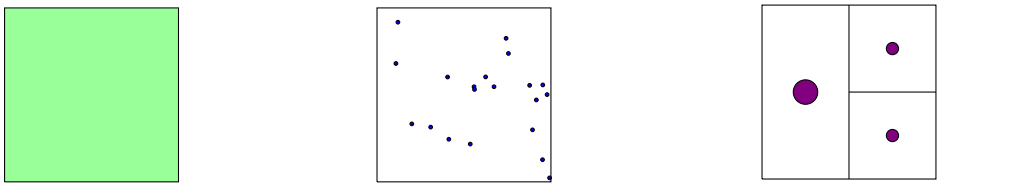
In particular, the partition based quantization reads

$$V_t(x_{t-1}) = \sum_{P\in\mathcal{P}_{t,x_{t-1}}} \check{p}_{t,P}\hat{V}(x_{t-1}, \check{\xi}_{t,P}).$$

If there exists a partition $\mathcal{P}_{t,x_{t-1}}$ such that the induced quantization is exact at time t on x_{t-1}, for all distributions of $(\boldsymbol{\xi}_\tau)_{\tau\in[T]}$, we call the exact quantization universal.

	SAA	Jensen/Edmundson-Madansky	Exact quantization
Setting	General	Convex	Linear
Approximations	Confidence interval	Upper and lower bounds	Exact values

Table 2.1: Comparison of advantages and drawbacks of approximation methods with exact quantization



$\boldsymbol{\xi}_t$ continu

$(\check{\xi}_{t,k})_{k=N}$ pour SAA avec $N = 20$

$(\check{\xi}_{t,P})_{P\in\mathcal{P}}$ pour une discrétisation induite par $\mathcal{P}$

Figure 2.7: Comparaisons des scénari choisis selon la discrétisation.

In this thesis, we only discuss partition based quantization. Before diving in theorems of exact quantization in the core of the thesis, we first explain in this section why we cannot expect to have general exact quantization theorem in the non-linear setting. We then focus on stochastic linear problems (SLP). By giving counter examples, we show that there is neither local and uniform partition based exact quantization when the recourse matrix is stochastic with non-finitely supported distribution, and that there is no uniform exact quantization when the constraints are stochastic with non-finitely supported distribution.

2.3.1 Non-linear case

If $\mathcal{P}_{t,x_{t-1}}$ is a partition of Ξ_t, the law of total expectation yields

$$V_t(x_{t-1}) = \sum_{P\in\mathcal{P}_{t,x_{t-1}}} \check{p}_{t,P}\mathbb{E}\big[\hat{V}_t(x_{t-1},\boldsymbol{\xi}_t)\,|\,\boldsymbol{\xi}_t\in P\big].$$

If we assume that the cost-to-go function $\hat{V}_t$ is convex in the uncertainty parameter ξ_t, by Jensen's inequaliy, we have that for all $P\in\mathcal{P}_{t,x_{t-1}}$, $\hat{V}_t(x_{t-1},\check{\xi}_{t,P}) = \hat{V}_t\Big(x_{t-1},\mathbb{E}\big[\boldsymbol{\xi}_t\,|\,\boldsymbol{\xi}_t\in P\big]\Big) \leqslant \mathbb{E}\Big[\hat{V}_t(x_{t-1},\boldsymbol{\xi}_t)\,|\,\boldsymbol{\xi}_t\in P\Big]$. In particular, we have the partition based exact quantization result if and only if we have the equality in Jensen's inequality, for all $P\in\mathcal{P}_{t,x_{t-1}}$ such that $\check{p}_{t,P}>0$. When we have strict convexity, Jensen's inequality is generally strict, we also say that the Jensen gap is positive (see e.g. [Wal14]). Then, a partition-based exact quantization result is highly improbable when $\hat{V}_t$ is strictly convex in ξ_t. We have similar conclusions if $\hat{V}_t$ is strictly concave in ξ_t, by inverting the inequality. Although we might show some exact quantization results in particular cases, this discussion shows that we cannot hope to have a general exact quantization theorem if $\hat{V}_t$ is not piecewise affine in ξ_t.

We thus focus on the linear setting.

2.3.2 Counter examples in 2SLP

Multistage stochastic linear problems (MSLP) are special cases of MSP where the loss functions are linear and the constraints are affine. For sake of simplicity, we first 2-stage linear problems

[4]When $\check{p}_{t,P} := \mathbb{P}\big[\boldsymbol{\xi}_t\in P\big]$ is equal to 0, $\mathbb{E}\big[\boldsymbol{\xi}_t\,|\,\boldsymbol{\xi}_t\in P\big]$ is not well-defined. Then, we take an arbitrary $\check{\xi}_{t,P}$ in P. This choice does not matter since $\check{\xi}_{t,P}$ will only appear in functions multiplied by $\check{p}_{t,P}=0$ in formulas.

(2SLP) which are special cases of MSLP with horizon 2. In 2SLP, we only have to study a single expected cost-to-go function $V := V_2$. We also denote the dimensions $n := n_1$, $m := n_2$, the state variables $x := x_1$, $y := x_2$ and $\Xi := \operatorname{supp}(\boldsymbol{\xi}) \subset \mathbb{R}^m \times \mathbb{R}^{\ell \times n} \times \mathbb{R}^{\ell \times m} \times \mathbb{R}^\ell$ the support of $\boldsymbol{\xi} = (\boldsymbol{q}, \boldsymbol{W}, \boldsymbol{T}, \boldsymbol{h}) := \boldsymbol{\xi}_2 = (\boldsymbol{c}_2, \boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{b}_2)$.

We thus consider the expected cost-to-go function

$$\hat{V} : \begin{cases} \mathbb{R}^n \times \Xi & \to \mathbb{R} \cup \{-\infty, +\infty\} \\ (x, \xi) & \mapsto \begin{cases} \min_{y \in \mathbb{R}^m} & q^\top y \\ \text{s.t.} & Tx + Wy \leqslant h \end{cases} \end{cases}$$

We define

$$V(x) := \mathbb{E}\big[\hat{V}(x, \boldsymbol{\xi})\big]$$

We now show that in the general case, there is neither local or uniform partition based exact quantization when the recourse matrix $\boldsymbol{W}$ is stochastic:

Example 2.2 (Stochastic $\boldsymbol{W}$). Let $m = \ell = 1$, $\boldsymbol{q} = (1)$, $\boldsymbol{T} = (0)$, $\boldsymbol{W} = (-\boldsymbol{u})$, $\boldsymbol{h} = (-1)$ where $\boldsymbol{u}$ denotes a uniform random variable on $[1, 2]$. We then have for all $x \in \mathbb{R}^n$:

$$\hat{V}(x, \xi) = \begin{array}{ll} \min\limits_{y \in \mathbb{R}} & y \\ \text{s.t.} & uy \geqslant 1 \end{array} = \frac{1}{u} \tag{2.3}$$

Thus, $\hat{V}$ is constant in x and strictly convex in ξ. We then have for all finite partition $\mathcal{P}$ of Ξ:

$$\sum_{P \in \mathcal{P}} \check{p}_P \hat{V}(x, \check{\xi}_P) < V(x) \tag{2.4}$$

with $\check{p}_P = \mathbb{P}[\boldsymbol{\xi} \in P]$, $\check{\xi}_P = \mathbb{E}[\boldsymbol{\xi} \,|\, \boldsymbol{\xi} \in P]$. Thus, for all $x \in \mathbb{R}^n$, there is no partition-based local, thus uniform, exact quantization result at x when $\boldsymbol{W}$ is non-finitely supported. We give a more detailed proof in Appendix A.

An obvious necessary condition for having uniform exact quantization at all times t is that the value functions V_t are polyhedral functions, meaning that they take value in $\mathbb{R} \cup \{+\infty\}$ and their epigraphs are (possibly empty) polyhedra. Indeed, for each $\xi \in \operatorname{supp}(\boldsymbol{\xi})$, $Q^\xi : (x, y) \to q^\top y + \mathbb{I}_{Tx+Wy \leqslant h}$ is polyhedral. Thus, $\hat{V}(\cdot, \xi) = \min_{y \in \mathbb{R}^m} Q^{c,\xi}(\cdot, y)$ is polyhedral as $\operatorname{epi} \hat{V}(\cdot, \xi)$ is a projection of $\operatorname{epi} Q^\xi$ (see Fig. 2.8 and [JKM08]). Finally, V, being a positive linear combination of polyhedral functions, is also polyhedral. Hence, the following examples show that if the constraints have non-discrete distributions, there is no hope to have a uniform exact quantization theorem.

Example 2.3 (Stochastic $\boldsymbol{T}$). Here, and in the next example, $\boldsymbol{u}$ denotes a uniform random variable on $[0, 1]$.

$$V(x) = \mathbb{E}\begin{bmatrix} \min\limits_{y \in \mathbb{R}^m} & y \\ \text{s.t.} & \boldsymbol{u}x \leqslant y \\ & 1 \leqslant y \end{bmatrix} = \mathbb{E}\big[\max(\boldsymbol{u}x, 1)\big] = \begin{cases} 1 & \text{if } x \leqslant 1 \\ \frac{x}{2} + \frac{1}{2x} & \text{if } x \geqslant 1 \end{cases}$$

Example 2.4 (Stochastic $\boldsymbol{h}$).

$$V(x) = \mathbb{E}\begin{bmatrix} \min\limits_{y \in \mathbb{R}^m} & y \\ \text{s.t.} & \boldsymbol{u} \leqslant y \\ & x \leqslant y \end{bmatrix} = \mathbb{E}\big[\max(x, \boldsymbol{u})\big] = \begin{cases} \frac{1}{2} & \text{if } x \leqslant 0 \\ \frac{x^2+1}{2} & \text{if } x \in [0, 1] \\ x & \text{if } x \geqslant 1 \ . \end{cases}$$

In contrast, we shall see, in a perhaps surprising way, that there is a uniform and universal exact quantization theorem without any restriction on the cost distribution. Moreover, there also exists a local quantization result when $\boldsymbol{T}$ and $\boldsymbol{h}$ are stochastic, see Table 2.2.

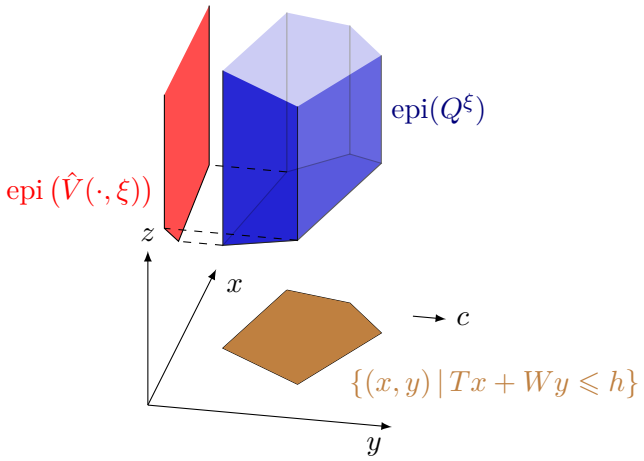


Figure 2.8: epi $\left(\hat{V}(\cdot,\xi)\right)$ is the projection of epi(Q^ξ)

2.4 Contributions

We now sum-up in a concise way the main contributions and results of this thesis.

In this thesis, we show how the tools from polyhedral geometry enlighten the structure of multistage stochastic linear problems.

In Chapter 3, we reformulate results from polyhedral geometry and combinatorics with useful notations for the stochastic programming community. Although the results in this chapter may look like common knowledge for the polyhedral community, we believe that they both introduce the topic to the stochastic optimizers and give formal proofs that can be hard to find explicitly in the literature.

2.4.1 Understanding MSLP through polyhedral geometry

First, we build a bridge between polyhedral geometry and stochastic programming.

Contribution 1. *We show that the expected cost-to-go functions are piecewise affine on explicit and universal polyhedral complexes, when the constraints are finitely supported and the costs have general distributions.*

For the 2-stage case, we indeed prove in Theorem 4.9 that the expected cost-to-go function is affine on every cell of the chamber complex, a universal polyhedral complex only depending on the constraints. This also allows us to derive combinatorial formulas for the expected cost-to-go functions and its subdifferential in Theorem 4.9. We then extend this result to MSLP in Theorem 4.14 by constructing by backward induction universal chamber complexes where the expected cost-to-go functions are piecewise affine, independently of the cost distributions.

Contribution 2. *We give an explicit characterization of expected cost-to-go functions in terms of support functions of nested weighted fiber polyhedra, when the constraints are deterministic and the costs have general distributions.*

We define a generalization of the fiber polytope of Billera and Sturmfels [BS92], called weighted fiber polyhedron, and show in Theorem 4.5 that the expected cost-to-go function in 2SLP is equal, up to an affine change of variable, to the support function of the weighted fiber polyhedron. This result is extended to MSLP in Theorem 4.17 by defining nested fiber polyhedra through dynamic programming.

2.4.2 Exact quantization results

In this thesis, we also give several exact quantization results that we sum up here.

Contribution 3. *We give local and universal exact quantization results for 2SLP with fixed recourse.*

Indeed, we show in Theorem 4.2 that any refinement of the opposite of the normal fan $\mathcal{N}(P_x)$ of the second stage admissible set P_x at x yields a *local and universal* exact quantization when the $\boldsymbol{c}$ is stochastic. By taking the dual problem, we also show in Theorem 5.3 that we have a *local and universal* exact quantization result when $\boldsymbol{T}$ and $\boldsymbol{h}$ are stochastic in the 2-stage case. This *local and universal* exact quantization result extends to the 2-stage case where $\boldsymbol{q}$, $\boldsymbol{T}$ and $\boldsymbol{h}$ are all stochastic in Theorem 5.15. We sum up the exact quantization results in 2SLP in Table 2.2.

	Local and universal	Uniform and universal
$\boldsymbol{W}$	×: Counter ex 2.2	×: Counter ex 2.2
$(\boldsymbol{T}, \boldsymbol{h})$	✓: $\mathcal{R}_x$ in Theorem 5.3	×: Counter ex 2.3 and 2.4
$\boldsymbol{q}$	✓: $-\mathcal{N}(P_x)$ in Theorem 4.2	✓: $\bigwedge_{\sigma\in\mathcal{C}(P,\pi)} -\mathcal{N}_\sigma$ in Theorem 4.3
$(\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})$	✓: $\mathcal{R}_x$ in Theorem 5.15	×: Counter ex 2.3 and 2.4

Table 2.2: Existence of partition based uniform and local (both universal) quantization for 2SLP with the associated partitions or counter examples

Contribution 4. *We give uniform and universal exact quantization results for MSLP with general cost distribution* $\boldsymbol{c}$ *and finitely supported constraints.*

When the constraints are deterministic, the local exact quantization in 2SLP is valid at all the first state x belonging to the same relative interior of a cell of the chamber complex. By taking the common refinement on all cells of the chamber complex, we derive in Theorem 4.3 a *uniform and universal* exact quantization for general cost $\boldsymbol{c}$. By propagating the chamber complexes through dynamic programming, we extend the *uniform and universal* exact quantization to the *multistage* case and to finitely supported constraints in Theorem 4.15.

2.4.3 New algorithms

The new perspectives of polyhedral geometry and exact quantization allow to define new algorithms, summed up in Table 2.3, to solve stochastic problems.

Contribution 5. *We present a generalized adaptive partition based method (GAPM) to solve 2SLP with stochastic* $\boldsymbol{T}$ *and* $\boldsymbol{h}$ *and give an extension to stochastic* $\boldsymbol{q}$.

Adaptive partition based methods (APM) were introduced by Song and Luedtke [SL15] to reduce the size of finitely supported 2SLP by gathering scenarios thanks to a partition. Ramirez-Pico and Moreno showed in [RPM21] that APM can be adapted to general distributions, calling it generalized adaptive partition based method (GAPM), if we are able to respect an analytical sufficient condition. We present an algorithm, see Algorithm 5.1, that allows to implement a GAPM methods for every distribution by applying an exact quantization result. We also give convergence and complexity results for APM methods. Thanks to an implementation of the algorithm in Julia, we compare GAPM with SAA methods by providing numerical results.

[5] We give ideas to develop a higher order simplex algorithm for 2SLP with general cost distribution in the conclusion of this manuscript.

	2SLP	MSLP
Local $(\boldsymbol{T}, \boldsymbol{h})$	GAPM in Chapter 7	GAPDP in Chapter 6
Local $\boldsymbol{q}$	Secondary Simplex[5]in Chapter 7	GAPDP in Chapter 6
Uniform $\boldsymbol{q}$	Extended GAPM in Chapter 5	Nested fiber and chamber complex in Chapter 4

Table 2.3: Algorithms and tools to tackle exact quantization in 2SLP and MSLP

Contribution 6. *We provide a framework called Trajectory Following Dynamic Programming (TFDP) which encompasses at least 14 algorithms (see Table 6.1), including SDDP and its variants. We then explain how to adapt all these algorithms, originally designed for finitely supported distribution, to the case of general distributions.*

In Chapter 6, we show how all these variants integrates the framework, see Algorithm 6.1, by discussing the choice of the upper and lower approximations called cuts, the choice of the trajectory depending on the uncertain parameter, *i.e.* the node selection, and the optimization problem we solve in the forward phase.

2.4.4 Complexity results

We present several results concerning the complexity of multistage stochastic linear programs.

Contribution 7. *We show that MSLP with general cost can be solved approximately in polynomial time in* $\log(1/\varepsilon)$*, when* T*,* $n_2, \cdots, n_T$ *are fixed for a large class of regular density functions.*

In particular, the first state dimension is not fixed. The proof relies on the theory of linear programming with oracles [GLS12] as well as on upper bound theorems of McMullen [McM70] and Stanley [Sta75] concerning the number of vertices and the size of a triangulation of a polyhedron. This complexity theorem, see Corollary 4.30, is widely applicable since the distribution cost can now be essentially arbitrary; we only assume that it is given implicitly through an appropriate oracle (see Definition 4.27) – this applies in particular to any distribution with a smooth density with respect to Lebesgue measure. Moreover, we obtain a similar polynomial complexity bounds in the exact (Turing) model of computation for appropriate classes of distributions, such as uniform on polytopes or exponential in Corollary 4.25. For comparison, to best of our knowledge, all the other complexity bounds were in pseudo-polynomial time, *i.e.* polynomial in $1/\varepsilon$.

Moreover, we give complexity bounds in the number of iterations for APM methods and for the algorithms following the TFDP framework. In multistage setting, assume, for every time step, relatively complete recourse, that the admissible sets have diameter smaller than D and dimensions at most d, and that the expected cost-to-go functions have a Lipschitz constants smaller than L. Let γ be the approximation error accumulated during an iteration.

Contribution 8. *The algorithms encompassed by TFDP framework find an approximate* ε*-solution in at most* $\left(\frac{2DL}{\varepsilon-\gamma}\right)^n (T-1)^{n+1}$ *iterations , when implemented with the problem child (resp. explorative) node selection, and in at most* $(T-1)\left(\frac{4DL(T-1)}{\varepsilon-\gamma}\right)^{n+2(T-1)}$ *iterations when implemented with a randomized node selection.*

In particular, this provides new complexity results, see Corollary 6.17 and Theorem 6.21, for at least 10 algorithms in the literature (see Table 6.1). For 2SLP, we show in Theorem 5.11 that APM methods, including GAPM, find an ε-solution in at most $(1+\frac{LD}{\varepsilon})^n$ iterations.

2.5 Structure of the thesis

The rest of the thesis is laid out as follows:

Chapter 3 gives an introduction on polyhedral geometry for stochastic optimizers. It recalls the definition of basic notions such as polyhedra and their representations, fans and chamber complexes. We give formal proof for the correspondences between the faces of a polyhedron, its normal cones and its sets of active constraints. We define the more combinatorial notions of regular subdivisions and secondary fan and show how they are related to the notion of active constraint sets. Finally, we present an adaptation of the basis decomposition theorem of Walkup and Wets [WW69] to our notations.

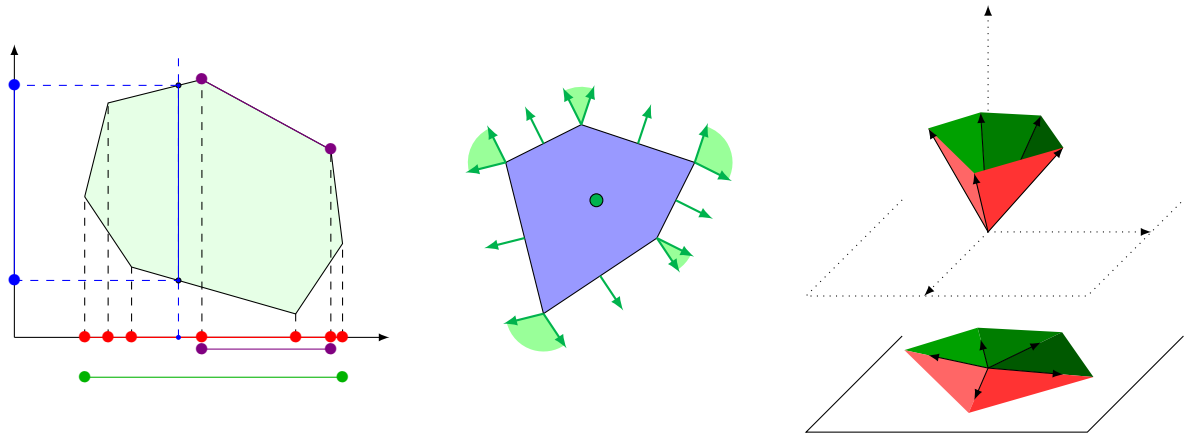
In Chapter 4, we study the polyhedral structure of MSLP with general cost distribution. We first prove that we have a universal and uniform exact quantization of 2SLP with general cost distribution which allows us to give a characterization of the expected cost-to-go function. We give a dual interpretation by defining a new object called weighted fiber polyhedron. This results are then extended to the multistage case. By constructing, through dynamic programming, universal chamber complexes, where the expected cost-to-go function is piecewise affine, we give a uniform and universal exact quantization results for MSLP with general cost distributions. We finally derive complexity results for MSLP showing that with fixed parameters, MSLP becomes polynomial for every regular distribution. Most of the results of Chapter 4 comes from the preprint [FGL21], which was awarded the best student paper price at ECSO-CMS 2022, with some new results detailed in the introduction of the chapter.

Chapter 5 deals with Adaptive Partition-based Methods (APM), which are numerical methods that solve, in particular, two-stage stochastic linear problems (2SLP). We say that a partition of the uncertainty space is *adapted* to the current first stage control $\check{x}$ if it induces a local exact quantization results at $\check{x}$. The core idea of APM is to iteratively constructs an *adapted* partition to all past tentative first stage controls. Relying on the normal fan of the dual admissible set, we present a local exact quantization when the technology matrix $\boldsymbol{T}$ and the right-hand side vector $\boldsymbol{h}$ have general distributions. This allows us to derive a necessary and sufficient condition for a partition to be adapted even for non-finite distribution, and provide a geometric oracle to obtain an adapted partition. Further, by showing the connection between APM and the L-shaped algorithm, we prove convergence and complexity bounds of the APM methods. We give some numerical results and compare GAPM with SAA methods. We finally extend the results to general distribution for the second stage cost. The majority of this chapter can be found in the paper [FL22b], but the extension to general second stage cost is written in [FL22a, Section 3].

In Chapter 6, we introduce a class of algorithms, called Trajectory Following Dynamic Programming (TFDP) algorithms, that iteratively refines approximations of expected cost-to-go functions of multistage stochastic problems with independent random variables. This framework encompasses most variants of the Stochastic Dual Dynamic Programming algorithm. Leveraging a Lipschitz assumption on the expected cost-to-go functions, we provide a new convergence and complexity proof that allows random variables with non-finitely supported distributions. In particular, this leads to new complexity results for numerous known algorithms. Further, we detail how TFDP algorithms can be implemented without the finite support assumption, either through approximations or exact computations. This chapter presents the results of the preprint [FL22a], except section 3, included in the previous chapter.

We conclude and give some perspectives for subsequent works in Chapter 7.

3 Polyhedral geometry toolbox for stochastic optimizers



Contents

This chapter presents the tools and results from polyhedral geometry that are used throughout this thesis. We reformulate properties from polyhedral geometry and combinatorics with useful notations for the stochastic programming community. While most of these results are considered as general knowledge, some proofs can be hard to find explicitly in the literature without falling in circular reasoning. With this in mind, this chapter might be seen as a Bourbaki-like polyhedral geometry toolbox for stochastic programmer.

We first define polyhedra and explain how they can be implemented through V-representation or H-representation. We then recall the notion of polyhedral complexes in order to focus on two particular cases, namely the normal fan and the chamber complex. These two particular polyhedral complexes happen to be key tools to describe the structure of stochastic linear programming. To derive explicit formulas and implement computations, we eventually present combinatorial notions such as the active constraints sets. We end this chapter by recalling the link with the general theory of regular subdivisions.

3.1 Polyhedra, cones and faces

Polyhedra are geometrical objects that are central to optimization. In this chapter, we give a brief introduction to polyhedral geometry. We refer to [Zie12, Grü13, Fuk16, JT13] for a detailed introduction to this field.

3.1.1 H-representation, V-representation

Polyhedra are convex[1] sets that can be represented, either as an intersection of a finite number of closed half spaces, we speak of H-representation, or as a set generated by vertices and rays called the V-representation.

More formally, a subset P of $\mathbb{R}^d$ is a *polyhedron* if there exists a finite collection $(H_i)_{i\in[p]}$ of p hyperplanes $H_i := \{x \in \mathbb{R}^d \,|\, a_i^\top x \leqslant b_i\}$, with a_i vector in $\mathbb{R}^d$ and b_i a scalar such that $P = \bigcap_{i=1}^p H_i$. In particular, the polyhedron can be written in the following *H-representation* as

$$P = \{x \in \mathbb{R}^d \mid a_i^\top x \leqslant b_i, \forall i \in [p]\} \tag{3.1}$$

We define the *convex hull* of vectors $v_1, \dots, v_s \in \mathbb{R}^d$ as

$$\mathrm{Conv}(v_1, \dots, v_s) := \Big\{ \sum_{j=1}^s \lambda_j v_j \mid \lambda_j \geqslant 0, \sum_{i=1}^s \lambda_j = 1 \Big\} \tag{3.2}$$

A subset Q of $\mathbb{R}^d$ is a *polytope* if there exist vectors $(v_i)_{i\in[s]}$ such that $Q = \mathrm{Conv}(v_1, \cdots, v_s)$. We say that $(v_1, \cdots, v_s)$ is a V-representation of Q.

For $d \in \mathbb{N}$, the *simplices* of $\mathbb{R}^d$ are the polytopes which can be written $\mathrm{Conv}(v_1, \cdots, v_n)$ with $n \leqslant d+1$. As stated in Theorem 3.4, the polytopes are exactly the bounded polyhedra.

The *conic hull* of vectors $r_1, \dots, r_t \in \mathbb{R}^d$ is

$$\mathrm{Cone}(r_1, \dots, r_t) := \Big\{ \sum_{j=1}^t \mu_j r_j \mid \mu_j \geqslant 0 \Big\} \tag{3.3}$$

A *cone* is a non-empty set of point C stable by multiplication by a positive scalar:

$$C \text{ is a cone} \iff \forall x \in C,\ \forall \lambda \in \mathbb{R}_+,\ \lambda x \in C \tag{3.4}$$

A polyhedral cone is a cone which is also a polyhedron. Since in this thesis, we are only interested in polyhedral cones, we use the name cone to refer to polyhedral cones. Note that

[1] There exist a large field of mathematics that deals with non-convex polyhedra. In this thesis, the terms polyhedron and polyhedra always refer to *convex* polyhedra.

(a) H-representation: $P = \{x \mid Ax \leqslant b\}$ (b) V-representation: $P = \text{Conv}(v_1, ..., v_n)$

Figure 3.2: 2 representations of the same polytope

the set $\text{Cone}(r_1, \cdots, r_t)$ is a cone. We state in Theorem 3.3 that every (polyhedral) cone C can be written in this way and we say that $(r_i)_{i\in[t]}$ is a V-representation of C.

The *simplicial cones* of $\mathbb{R}^d$ are the cones which can be written $\text{Cone}(r_1, \cdots, r_t)$ with $t \leqslant d$.

The Minkowski sum of X and Y two subsets of $\mathbb{R}^d$ is

$$X + Y := \{x + y \mid x \in X, y \in Y\} \tag{3.5}$$

We now present the Minkowski-Weyl Theorem which states that a subset of $\mathbb{R}^d$ is a polyhedron if and only if it is the Minkowski sum of a polytope and a cone.

Theorem 3.1 (Minkowski-Weyl's theorem for polyhedra see [Zie12] 1.2 or [Fuk16] Thm 3.9)**.** *For $P \subset \mathbb{R}^d$, the following statements are equivalent :*

1. *There exist $(a_i)_{i\in[q]} \in (\mathbb{R}^d)^q$ and $(b_i)_{i\in\mathbb{R}^q}$ such that $P := \{x \in \mathbb{R}^d \mid a_i^\top x \leqslant b_i, \forall i \in [q]\}$.*
2. *There exist finite families of vectors v_i and r_j in $\mathbb{R}^d$ such that $P = \text{Conv}(v_1, \ldots, v_s) + \text{Cone}(r_1, \ldots, r_t)$*

In particular, P is a polyhedron if and only if it satisfies one of this statement.

Thus, we have two ways to represent, for example in a computer, the polyhedron P. We say that $(a_i, b_i)_{i\in[q]}$ is a *H-representation* of P. Similarly, we say that $((v_i)_{i\in[s]}, (r_i)_{i\in[t]})$ is a *V-representation* of P.

Note that H-representation and V-representation are not unique. Indeed, we can add inequalities $a_i^\top x \leqslant b_i$ that are always valid without changing the polyhedron P. Similarly, adding points v_i already in the convex hull does not change the polyhedron P either. However, a polytope Q admits a unique minimal V-representation. We call the elements of these minimal representation the *vertices* of Q. On the other hand, a full dimensional polyhedron admits a unique minimal H-representation $(a_i, b_i)_{i\in[q]}$ up to positive scalar factors.

To ease the reading, we prefer to rely on matrix notation. If we define the matrix A as the matrix whose rows are the row vectors $a_i^\top$ and the vector b with coefficients b_i. The polyhedron P can be written as

$$P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} \tag{3.6}$$

Equivalently we can write the convex and conic hulls with a matrix representation. We define the matrices $V \in \mathbb{R}^{d\times s}$ and $R \in \mathbb{R}^{d\times t}$ as the matrices whose columns are respectively the vectors $(v_i)_{i\in[s]}$ and $(r_i)_{i\in[t]}$. We can equivalently define the convex hull of V as

$$\text{Conv}(V) := V\Delta_{s-1} \tag{3.7}$$

where Δ_d stands for the canonical simplex of dimension d as $\Delta_d := \{x \in \mathbb{R}^{d+1} \,|x_i \geqslant 0, \sum_{i=1}^{d+1} x_i = 1\}$.

Similarly, we define the conic hull of R as

$$\mathrm{Cone}(R) := \mathrm{Cone}(R) = \{R\mu \mid \mu \geqslant 0, \mu \in \mathbb{R}^t\} \tag{3.8}$$

We now restate the Minkowski-Weil theorem with this notation.

Theorem 3.2 (Minkowski-Weyl's Theorem for polyhedra with matrix notation)**.** *For $P \subset \mathbb{R}^d$, the following statements are equivalent :*

1. *There exist $A \in \mathbb{R}^{q\times d}$ and $b \in \mathbb{R}^q$ such that $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$.*
2. *There exist $V \in \mathbb{R}^{d\times s}$ and $R \in \mathbb{R}^{d\times t}$ such that $P = \{V\lambda + R\mu \mid \mu \geqslant 0, \mathbf{1}^\top\lambda = 1, \lambda \geqslant 0\} = V\Delta_{s-1} + R\mathbb{R}^t_+$*

In particular, P is a polyhedron if and only if it satisfies one of this statement.

Similarly, we say that (A, b) is a *H-representation* of P and that (V, R) is a *V-representation* of P. The idea of the proof is to add a lift variable and represent the polyhedron $P \subset \mathbb{R}^d$ as a cone $\widetilde{P} = \{(x, \lambda) \in \mathbb{R}^d \times \mathbb{R}_+ \mid x \in \lambda P\}$ in $\mathbb{R}^{d+1}$. With this homogenization trick, the Theorem 3.2 is a consequence of the following Minkowski-Weil for cones:

Theorem 3.3 (Minkowski-Weyl's Theorem for cones (see [Fuk16] Thm 3.10))**.** *For $P \subset \mathbb{R}^d$, the following statements are equivalent :*

1. *There exist $A \in \mathbb{R}^{q\times d}$ such that $P = \{x \in \mathbb{R}^d \mid Ax \leqslant 0\}$.*
2. *There exists $R \in \mathbb{R}^{d\times t}$ such that $P = \{R\mu \mid \mu \geqslant 0\} = R\mathbb{R}^t_+$.*

In particular, P is a cone if and only if it satisfies one of these conditions.

The idea of the proof is to obtain the H-representation from the V-representation by Fourier-Motzkin elimination. The backward implication is proved by a duality argument.

We eventually adapt the Minkowski-Weil theorem for polytopes which states that the polytopes are exactly the bounded polyhedra.

Theorem 3.4 (Minkowski-Weyl's Theorem for polytopes)**.** *For $P \subset \mathbb{R}^d$ non-empty, the following statements are equivalent :*

1. *There exist $A \in \mathbb{R}^{q\times d}$ and $b \in \mathbb{R}^q$ such that $P = \{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ and P is bounded.*
2. *There exists $V \in \mathbb{R}^{d\times s}$ such that $P = \{V\lambda \,|\, \mathbf{1}^\top\lambda = 1, \lambda \geqslant 0\} = V\Delta_{s-1}$*

In particular, P is a polytope if and only if P satisfies one of these statements or P is empty.

Proof. By Theorem 3.1, we only have to prove that for $P := \mathrm{Conv}(v_1, \cdots, v_n) + \mathrm{Cone}(r_1, \cdots, r_t)$, we have that

P is bounded if and only if $t = 0$.

($\Rightarrow$) If $t \geqslant 1$, for all $\mu \in \mathbb{R}_+$, $\mu r_1 \in P$, thus P is not bounded.

($\Leftarrow$) If $t = 0$, let $M = \max(\|v_1\|, \cdots, \|v_n\|)$. Then, for all $x \in P = \mathrm{Conv}(v_1, \cdots, v_n)$, $\|x\| \leqslant M$. Thus, P is bounded. □

3.1.2 Basic properties of polyhedra

Polyhedra are stable by intersection, Minkowski sum, linear transformation and projection.

Proposition 3.5. *For P and Q be two polyhedra, A a matrix and π a linear projection. We have the following properties.*

- *$P \cap Q$ is a polyhedron.*

- *$P + Q$ is a polyhedron.*

- *AP is a polyhedron.*

- *$\pi(P)$ is a polyhedron.*

If $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a function, we say that f is *polyhedral function* (see [Roc15, Chapter 19]) if its epigraph $\mathrm{epi}(f)$ is a polyhedron of $\mathbb{R}^{d+1}$.

Proposition 3.6. *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a polyhedral function. Then there exists $(\alpha_i)_{i\in[n]} \in (\mathbb{R}^d)^n$, $(\beta_i)_{i\in[n]} \in \mathbb{R}^n$, a matrix A and a vector b such that*

$$f : x \mapsto \mathbb{I}_{Ax\leqslant b} + \max_{i\in[n]} \{\alpha_i^\top x + \beta_i\} \tag{3.9}$$

In particular, $\mathrm{dom}(f) = \{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ is a polyhedron.

Proof. If $\mathrm{epi}(f)$ is empty, the result is trivial. We now assume that $\mathrm{epi}(f)$ is not empty. Since $\mathrm{epi}(f)$ is a polyhedron, there exist a matrix M and vectors v and h such that $\mathrm{epi}(f) = \{(x, z) \,|\, Mx + vz \leqslant h\}$.

By contradiction, assume that there exists an index i such that $v_i > 0$. Let $(x, z) \in \mathrm{epi}(f)$, i.e. $f(x) \leqslant z$. Then, for all $z \geqslant z'$, $f(x) \leqslant z \leqslant z'$ and thus $(x, z') \in \mathrm{epi}(f)$ and $M_i x + v_i z' \leqslant h_i$. If we take $z' := \frac{h_i - M_i x + 1}{v_i}$, then $z \leqslant z'$ and $M_i x + v_i z' \leqslant h_i$. By simplifying, we get the contradiction $1 \leqslant 0$. Then, for all i, $v_i \leqslant 0$. Let $I^0 = \{i | v_i = 0\}$ and $I^- = \{i | v_i < 0\}$. We then have (x, z) is in $\mathrm{epi}(z)$ if and only if $M_{I^0} x \leqslant h_{I^0}$ and $z \geqslant \frac{h_i}{v_i} - \frac{1}{v_i} M_i x$ for all $i \in I^-$. We define $A := M_{I^0}$ the submatrix of M with row indices in I^0 and $b := h_{I^0}$ the subvector of h. We also define, up to changing the indices, $\alpha_i := -\frac{1}{v_i} M_i^\top$ and $\beta_i := \frac{h_i}{v_i}$. With this notation, we now have $\mathrm{epi}(f) := \{(x, z) \,|\, Ax \leqslant b,\ \alpha_i^\top x + \beta_i \leqslant z,\ \forall i\}$ which gives Eq. (3.9). □

We now define basic notions on polyhedra.

The *polar* of a convex set $C \subset \mathbb{R}^d$ is the convex cone

$$C^\circ := \{\alpha \in \mathbb{R}^d \mid \forall x \in C,\ \alpha^\top x \leqslant 0\}. \tag{3.10}$$

The *lineality space* of C is the following vector space

$$\mathrm{Lin}(C) := \{r \in \mathbb{R}^d \mid \forall x \in C, \forall \lambda \in \mathbb{R}, x + \lambda r \in C\}. \tag{3.11}$$

The *recession cone* of a convex set $C \subset \mathbb{R}^m$ is the set

$$\mathrm{rc}(C) := \{r \in \mathbb{R}^d \mid \forall x \in C, \forall \mu \in \mathbb{R}^+, x + \mu r \in C\} \tag{3.12}$$

Note that, if P is a polyhedron such that $P = Q + C$ with Q a polytope and C a cone, as in Theorem 3.1, we then have $\mathrm{rc}(P) = C$.

We say that a cone is *pointed* if its lineality space is reduced to $\{0\}$. A pointed cone C has a minimal V-representation up to positive scalar factors. An element of a minimal V-representation of C is called a *ray* of C.

Proposition 3.7. *The recession cone and the lineality space of a non-empty polyhedron $P = \{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ with $A \in \mathbb{R}^{q\times d}$ and $b \in \mathbb{R}^q$ are respectively*

$$\mathrm{rc}(P) = \{x \in \mathbb{R}^d \,|\, Ax \leqslant 0\} = A^{-1}(\mathbb{R}^q_-) \tag{3.13a}$$

$$\mathrm{Lin}(P) = \{x \in \mathbb{R}^d \,|\, Ax = 0\} = \mathrm{Ker}(A) \tag{3.13b}$$

$$\big(\mathrm{rc}(P)\big)^\circ = \mathrm{Cone}(A^\top) = A^\top \mathbb{R}^q_+. \tag{3.13c}$$

Proof. Let $r \in \mathrm{rc}(P)$, since P is not empty, there exists $x \in P$. We have that $A(x+\lambda r) \leqslant b$, for all $\lambda \in \mathbb{R}_+$. In particular, $Ar \leqslant (b - Ax)/\lambda$. By letting λ goes to infinity, we have $Ar \leqslant 0$. Now, if we take r such that $Ar \leqslant 0$. Then, for $x \in P$ and $\lambda \in \mathbb{R}_+$, we have that $A(x + \lambda r) \leqslant Ax + \lambda \times 0 \leqslant b$, thus $x + \lambda r$ belongs to P. It follows that $r \in \mathrm{rc}(P.)$ The proof for the lineality space is similar. □

3.1.3 Faces

A *face* F of a polyhedron $P \subset \mathbb{R}^d$ is a subset of $\mathbb{R}^d$ such that there exists $c \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ with

$$P \subset \{x \in \mathbb{R}^d \,|\, c^\top x \leqslant \beta\} \tag{3.14a}$$
$$F = P \cap \{x \in \mathbb{R}^d \,|\, c^\top x = \beta\} \tag{3.14b}$$

We denote by $\mathcal{F}(P)$ the set of faces of P.

Some faces are reduced to singleton we call them *vertices* and we denote by $\mathrm{Vert}(P)$ the set of vertices of P.

Note that, for a polyhedron P, we denote

$$P^\psi := \arg\min_{x \in P} \psi^\top x \tag{3.15}$$

By taking $c = -\psi$ and $\beta = \max_{x \in P} \psi^\top x$ we see that P^ψ is a face of P. More generally, for any set $E \subset \mathbb{R}^d$, we write

$$P^E := \cap_{\psi \in E} P^\psi. \tag{3.16}$$

In many contexts, it is useful to consider polyhedra in $\mathbb{R}^d \times \mathbb{R}$, e.g. for epigraphs of polyhedral functions. In these cases, the last coordinate plays a particular role that can be interpreted as a "height" or a lift variable. We say that a face F of the polyhedron $P \in \mathbb{R}^d \times \mathbb{R}$ is a *lower face* if it is "visible from below" *i.e.* , if there exists a vector $\varphi \in \mathbb{R}^d$ such that $F = P^{(\varphi,1)}$.

Proposition 3.8. *If P is a non-empty polyhedron, we have that*

3.8.(a) *$\emptyset$ and P are faces of P.*

3.8.(b) *Any face $F \in \mathcal{F}(P)$ is a polyhedron.*

3.8.(c) *If F and F' are faces of P, then $F \cap F'$ is also a face of P.*

3.8.(d) *A set F is a non-empty face of P if and only if there exists $\psi \in -\big(\mathrm{rc}(P)\big)^\circ$ such that $F = P^\psi$.*

3.8.(e) *$\mathcal{F}(P)$ is a finite collection of polyhedra.*

3.8.(f) *The relative interior of P is given by*

$$\mathrm{ri}(P) = P \backslash \bigcup_{F \in \mathcal{F}(P) \backslash \{P\}} F \tag{3.17}$$

Proof. 3.8.(a). By taking $c = 0$ and $\beta = 1$, we have that $P \subset \{x \in \mathbb{R}^d \,|\, 0^\top x \leqslant 1\}$ and $\emptyset = P \cap \{x \in \mathbb{R}^d \,|\, 0^\top x = 1\}$. Similarly, by taking $c = 0$ and $\beta = 0$, we get $P \subset \{x \in \mathbb{R}^d \,|\, 0^\top x \leqslant 0\}$ and $P = P \cap \{x \in \mathbb{R}^d \,|\, 0^\top x = 0\}$.

3.8.(b). Is direct from the definition of polyhedron as intersection of half-spaces.

3.8.(c). Let c, c', β, β' such that $F = P \cap \{x \in \mathbb{R}^d \,|\, c^\top x = \beta\}$, $F' = P \cap \{x \in \mathbb{R}^d \,|\, c'^\top x = \beta'\}$ and for all $x \in P$, $c^\top x \leqslant \beta$ and $c'^\top x \leqslant \beta'$. We then have that for all $x \in P$, $(c + c')^\top x \leqslant \beta + \beta'$ and $F \cap F' = P \cap \{x \in \mathbb{R}^d \,|\, (c + c')^\top x = \beta + \beta'\}$.

3.8.(d). Since $x \mapsto c^\top x$ is continuous and P is non-empty and closed, we have that P^ψ is empty if and only if $\inf_{x \in P} \psi^\top x = -\infty$. Moreover, if $F := P \cap \{x \in \mathbb{R}^d \,|\, c^\top x = \beta\}$, with $P \subset \{x \in \mathbb{R}^d \,|\, c^\top x \leqslant \beta\}$, is a non-empty face of P, then $\beta = \max_{x \in P} c^\top x$ and $F = P^{-c}$. We first prove that $\inf_{x \in P} \psi^\top x = -\infty$ is empty if and only if $\psi \in -\big(\mathrm{rc}(P)\big)^\circ$:

($\Rightarrow$) Let $\psi \notin -\mathrm{rc}(P)^\circ$. By definition of the polar cone, there exists $r \in \mathrm{rc}(P)$ such that $-\psi^\top r > 0$. By definition of the recession cone, there exists $x \in P$, such that, for any $\mu \in \mathbb{R}^+$, $x + \mu r \in P$. Thus, we have $\lim_{\mu \to +\infty} \psi^\top (x + \mu r) = -\infty$ and then $\inf_{x \in P} \psi^\top x = -\infty$.

($\Leftarrow$) Let $\psi \in -\mathrm{rc}(P)^\circ$. Then for all $r \in \mathrm{rc}(P)$, $-c^\top r \leqslant 0$, and $\inf_{r \in \mathrm{rc}(P)} c^\top r = 0$. By Minkowski Weyl theorem's 3.1, there exists a polytope Q such that $P = Q + \mathrm{rc}(P)$. Thus, $\inf_{x \in P} \psi^\top x = \min_{x \in Q, r \in \mathrm{rc}(P)} \psi^\top (x + r) = \min_{x \in Q} \psi^\top x$ is finite as Q is bounded.

We now prove 3.8.(d). Let $F = P \cap \{x \in \mathbb{R}^d \,|\, c^\top x = \beta\}$ be a non-empty face of F with $P \subset \{x \in \mathbb{R}^d \,|\, c^\top x \leqslant \beta\}$. Let $b = \sup_{x \in P} c^\top x$. We cannot have $\beta < b$ because $P \subset \{x \in \mathbb{R}^d \,|\, c^\top x \leqslant \beta\}$. We cannot have $\beta > b$ otherwise F is empty. Thus, $b = \beta$ and $F = P^{-c}$. Reciprocally, we already saw that P^ψ is a face of P and that it is non-empty if $\psi \in -\big(\mathrm{rc}(P)\big)^\circ$.

We prove 3.8.(e). and 3.8.(f). in Section 3.2.1 after defining active constraints sets.

□

3.1.4 Polyhedral complex, fan and chamber complex

Polyhedral complexes are collections of polyhedra satisfying some combinatorial and geometrical properties. In particular the relative interiors of the elements of a polyhedral complex (without the empty set) form a partition of their union. We refer to [DLRS10] for a complete introduction to polyhedral complexes and triangulations.

A finite collection of polyhedra $\mathcal{C}$ is a *polyhedral complex* if it satisfies

1. For every polyhedron $P \in \mathcal{C}$ and every non-empty[2] face F of P, $F \in \mathcal{C}$.
2. For any polyhedra P and Q in $\mathcal{C}$, $P \cap Q$ is a (possibly empty) face of P and Q.

We denote by $\mathrm{supp}\,\mathcal{C} := \bigcup_{P \in \mathcal{C}} P$ the *support* of a polyhedral complex. Further, if all the elements of $\mathcal{C}$ are polytopes (resp. cones, simplices, simplicial cones), we say that $\mathcal{C}$ is a *polytopal complex* (resp. a *fan*, a *simplicial complex*, a *simplicial fan*).

For example, the set of faces $\mathcal{F}(P)$ of a polyhedron P is a polyhedral complex.

Proposition 3.9. *For any polyhedral complex $\mathcal{C}$, the relative interiors of its elements (without the empty set) form a partition of its support:* $\mathrm{supp}(\mathcal{C}) = \bigsqcup_{P \in \mathcal{C}} \mathrm{ri}(P)$.

We recall that the notation $\sqcup$ refers to a disjoint union.

Let $\mathcal{C}$ and $\mathcal{R}$ be two polyhedral complexes, we say that $\mathcal{R}$ is a *refinement* of $\mathcal{C}$, denoted $\mathcal{R} \preccurlyeq \mathcal{C}$, if for every cell $R \in \mathcal{R}$ there exists a cell $C \in \mathcal{C}$ containing R: $R \subset C$.

Note that $\preccurlyeq$ defines a partial order on the space of polyhedral complexes, and the *meet* associated with this order is given by the *common refinement* of two polyhedral complexes $\mathcal{C}$ and $\mathcal{C}'$ defined as the polyhedral complex of the intersections of cells of $\mathcal{C}$ and $\mathcal{C}'$:

$$\mathcal{C} \wedge \mathcal{C}' := \{R \cap R' \,|\, R \in \mathcal{C}, R' \in \mathcal{C}'\} \tag{3.18}$$

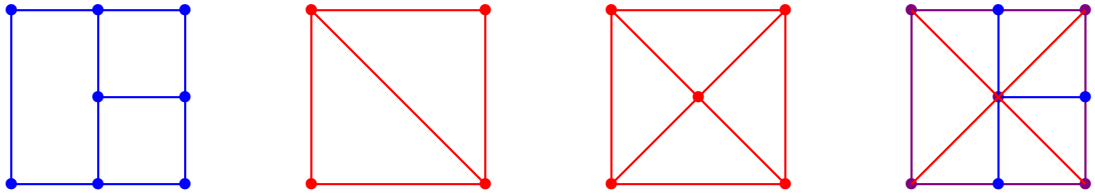
A *triangulation $\mathcal{T}$ of a polytope Q* is a refinement of $\mathcal{F}(Q)$ such that the cells of dimension 0 of $\mathcal{T}$ are the vertices of Q and $\mathcal{T}$ is a simplicial complex. A *triangulation $\mathcal{T}$ of a cone K* is a refinement of $\mathcal{F}(K)$ such that the cells of dimension 1 of $\mathcal{T}$ are the rays of K and $\mathcal{T}$ is a simplicial fan.

We recall that a *simplex* of dimension d is the convex hull of $d+1$ affinely independent points and that a *simplicial cone* of dimension d is the conical hull of d linearly independent vectors. We speak of triangulations because simplices in dimension 2 and 3 are respectively triangles and tetrahedra. Thus, simplices are generalization of triangles in higher dimensions.

Finally, to study the projection of the faces of P, we present classical polyhedral tools known as fibers and chambers complex[3] (see [BS92, RZ96, Ram96]).

[2] For some authors, a polyhedral complex must contain the empty set. We do not make this requirement.

[3] The terms *validity domain*, instead of chamber, and *parametrized polyhedron*, instead of fiber, are also used in the literature [CL98, LW97].



(a) $\mathcal{P}$ (b) $\mathcal{T}$ (c) $\mathcal{U}$ (d) $\mathcal{P} \wedge \mathcal{U}$.

Figure 3.3: A polyhedral complex $\mathcal{P}$, a triangulation $\mathcal{T}$ of the square, another triangulation $\mathcal{U}$ which refines $\mathcal{T}$. The common refinement $\mathcal{P} \wedge \mathcal{U}$.

Definition 3.10 (Chamber complex and fiber)**.** *Let $P \subset \mathbb{R}^d$ be a polyhedron and π a linear projection defined on $\mathbb{R}^d$. For $x \in \pi(P)$ we define the* chamber of x for P along π *as*

$$\sigma_{P,\pi}(x) := \bigcap_{F \in \mathcal{F}(P) \text{ s.t. } x \in \pi(F)} \pi(F). \tag{3.19}$$

The chamber complex $\mathcal{C}(P,\pi)$ *of P along π is defined as the (finite) collection of chambers,* i.e.

$$\mathcal{C}(P,\pi) := \{\sigma_{P,\pi}(x) \mid x \in \pi(P)\}. \tag{3.20}$$

We define the fiber P_x *of P at x along π as the projection of $P \cap \pi^{-1}(\{x\})$ on the space* $\operatorname{Ker}(\pi)$.

Further $\mathcal{C}(P,\pi)$ is a polyhedral complex such that $\operatorname{supp}\mathcal{C}(P,\pi) = \pi(P)$. *In particular,* $\{\operatorname{ri}(\sigma) \mid \sigma \in \mathcal{C}(P,\pi)\}$ *is a partition of $\pi(P)$.*

More generally, the chamber complex of a polyhedral complex $\mathcal{P}$ is

$$\mathcal{C}(\mathcal{P},\pi) := \{\sigma_{\mathcal{P},\pi}(x) \mid x \in \pi(\operatorname{supp}(\mathcal{P}))\}. \tag{3.21}$$

with $\sigma_{\mathcal{P},\pi}(x) := \bigcap_{F \in \mathcal{P} \text{ s.t. } x \in \pi(F)} \pi(F)$.

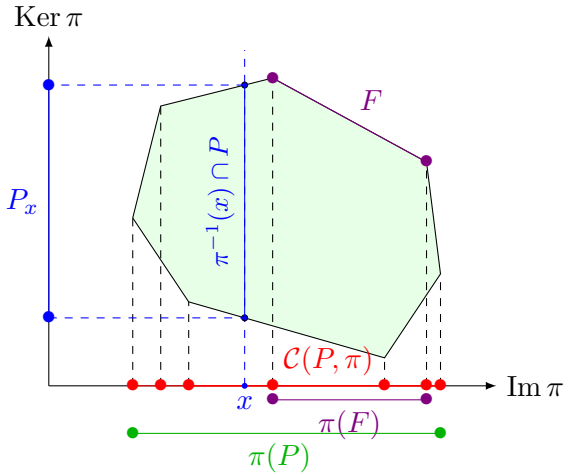


Figure 3.4: A polytope P in light green, its chamber complex in red on the x-axis and a fiber P_x in blue on the y-axis, for the orthogonal projection π on the horizontal axis.

3.2 Active constraints sets and normal fan

We now introduce the notions of active constraints and normal fan with basic properties. These notions are in one-to-one correspondences and the collection of active constraint sets will provide H-representations of the faces of a polyhedron and V-representations of its normal cones.

3.2.1 Active constraints sets

We introduce in this section the collection of active constraints which we use to obtain explicit formulas and make computations in practice. This notion is algebraic and depends on the matrix A and vector b used to define the polyhedron, i.e. its *H-representation*.

For any matrix $A \in \mathbb{R}^{q\times p}$ and a subset $I \subset [q]$, we denote by A_I the submatrix composed of the *rows* of indices in I of A

$$A_I := A_{I,\cdot} = (A_{i,j})_{i\in I, j\in[p]} \tag{3.22}$$

For $i \in [q]$, we also denote $A_i := A_{\{i\}}$ the ith row of A. To avoid confusion, we use the parenthesis rule $A_I^\top := (A_I)^\top$.

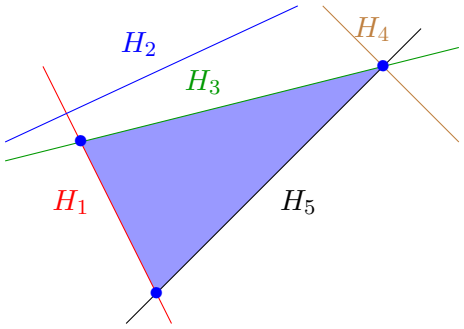


Figure 3.5: An example of a polyhedron $P = \{x \mid Ax \leqslant b\}$ with an H-representation (A, b). Each H_i corresponds to the hyperplane $\{x \mid A_i x = b_i\}$ and the label H_i is located in the halfspace $\{x \mid A_i x \geqslant b_i\}$. We have $\mathcal{I}(A, b) = \{\emptyset, \{1\}, \{1, 3\}, \{3\}, \{3, 4, 5\}, \{5\}, \{1, 5\}\}$. Constraint 2 is never active ($2 \notin \operatorname{supp} \mathcal{I}(A, b)$), and constraint 4 is redundant with constraints 3 and 5.

We now give basic properties of active constraints sets. In particular, the collection of active constraints sets $\mathcal{I}(A, b)$ provides H-representations of the non-empty faces of P.

Definition 3.11 (Active constraints set)**.** *For a polyhedron $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$, we denote by $I_{A,b}(x)$ the set of* active constraints *of P in $x \in \mathbb{R}^d$, with the H-representation $(A, b) \in \mathbb{R}^{q\times d} \times \mathbb{R}^q$:*

$$I_{A,b}(x) := \{i \in [q] \mid A_i x = b_i\} \tag{3.23}$$

More generally, for a set $E \subset P$, we write $I_{A,b}(E) := \bigcap_{x\in E} I_{A,b}(x)$.

We denote by $\mathcal{I}(A, b)$, the collection of active constraints sets of P with the H-representation (A, b):

$$\mathcal{I}(A, b) := \{I_{A,b}(x) \mid Ax \leqslant b\} \tag{3.24}$$

We now give elementary properties for the collection of active constraints sets.

Proposition 3.12 (Basic properties of active constraint sets)**.** *Let $P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$, for $I \in [q]$, we define the face $P(I) := \{x \in P \mid A_I x = b_I\}$. We then have the following propositions.*

3.12.(a) *If $I \in \mathcal{I}(A, b)$, then the relative interior of $P(I)$ is*

$$\operatorname{ri}(P(I)) = \{x \in \mathbb{R}^n \mid A_I x = b_I \text{ and } A_j x < b_j, \forall j \in [q]\backslash I\} \tag{3.25}$$

3.12.(b) *The mapping*

$$\begin{cases} \mathcal{I}(A, b) & \to \mathcal{F}(P)\backslash\{\emptyset\} \\ I & \mapsto P(I) \end{cases} \tag{3.26}$$

is a one-to-one correspondence whose inverse is $F \mapsto I_{A,b}(F)$.

3.12.(c) *For any I and J in $\mathcal{I}(A, b)$, we have $I \cap J \in \mathcal{I}(A, b)$.*

3.12.(d) *For $F \in \mathcal{F}(P)\backslash\{\emptyset\}$, we have $I_{A,b}(F) = I_{A,b}(x)$ for every $x \in \mathrm{ri}(F)$.*

3.12.(e) *For $F, F' \in \mathcal{F}(P)\backslash\{\emptyset\}$ and $I, I' \in \mathcal{I}(A,b)$*

$$I \subset I' \iff P(I) \supset P(I') \tag{3.27a}$$
$$F \subset F' \iff I_{A,b}(F) \supset I_{A,b}(F') \tag{3.27b}$$

3.12.(f) *For $I \in \mathcal{I}(A,b)$ and $F \in \mathcal{F}(P)\backslash\{\emptyset\}$,*

$$\dim\left(P(I)\right) = n - \mathrm{rg}(A_I) \tag{3.28a}$$
$$\dim(F) = n - \mathrm{rg}(A_{I_{A,b}(F)}) \tag{3.28b}$$

Proof. 3.12.(a) We first show that $\mathrm{Aff}\left(P(I)\right)$ is equal to $H_I := \{x \in \mathbb{R}^d \,|\, A_I x = b_I\}$. Indeed, $P(I)$ is included in H_I. Moreover, there exists x such that $A_I x = b_I$ and $A_j x < b_j$ for all $j \in [q]\backslash I$. Then, for $\varepsilon > 0$ small enough, $B(x,\varepsilon) \cap H_I \subset P$. In particular, every affine subset which contains P also contains H_I, which implies $\mathrm{Aff}\left(P(I)\right) = H_I$. Finally, a point x is in $\mathrm{ri}\left(P(I)\right)$ if and only if there exists $\varepsilon > 0$ such that $B(x,\varepsilon) \cap H_I \subset P(I)$ which is equivalent to $A_I x = b_I$ and $A_j x < b_j$ for all $j \in [q]\backslash I$.

3.12.(b) Let $F \in \mathcal{F}(P)\backslash\{\emptyset\}$. We define $I := I_{A,b}(F)$. We first prove that $F = P(I)$. If $x \in F$, we have that $I = I_{A,b}(F) \subset I_{A,b}(x)$. In particular, $A_I x = b_I$ and $x \in P(I)$. Reciprocally, we now assume that $x \notin F$. By 3.8.(d), there exists $\psi \in \mathbb{R}^d$ such that $F = P^{-\psi}$. We now show that $\psi \in \mathrm{Cone}(A_I)$. Let $\varphi \in \mathbb{R}^d$ such that $\varphi \notin \mathrm{Cone}(A_I)$. By Farkas lemma, there exists $z \in \mathbb{R}^d$, such that $A_I z \geqslant 0$ and $\varphi^\top z < 0$. Since F is non-empty there exists $y \in F \subset P(I)$. Let $\varepsilon > 0$, we have $A_I(y - \varepsilon z) = b_I - \varepsilon A_I z \leqslant b_I$. Since for all $j \notin I$, $A_j y < b_j$, for ε small enough, $A_j(y - \varepsilon z) < b_j$ for all $j \notin I$ and thus $y - \varepsilon z$ belongs to P. As $\varphi^\top z < 0$, we get $-\varphi^\top(y - \varepsilon z) < -\varphi^\top y$ and then $F \neq P^{-\varphi}$. In particular, since $F = P^{-\psi}$, we have that $\psi \in \mathrm{Cone}(A_I)$, i.e. there exists $\mu \in \mathbb{R}^I_+$ such that $\psi = A_I \mu$. Since $x \notin F = P^{-\psi}$, there exists $y \in F = P(I)$ such that $-\psi^\top y < -\psi^\top x$ which rewrites $\mu^\top A_I y > \mu A_I x$. As $A_I y = b_I$, we have $\mu^\top(b_I - A_I x) > 0$ and then there exists $i \in I$ such that $A_i x < b_i$. Thus, $x \notin P(I)$ and we can conclude $F = P(I) = P(I_{A,b}(F))$.

Finally, by 3.12.(a) and since F is not empty, there exists $x \in \mathrm{ri}(F)$ and we have $I_{A,b}(F) = I_{A,b}(x) \in \mathcal{I}(A,b)$. The mapping $I \mapsto P(I)$ is then well-defined and surjective. We now prove that it is injective. Let $I, J \in \mathcal{I}(A,b)$ with $I \neq J$. By symmetry, we can assume without loss of generality that there exists $i \in I\backslash J$. Since $J \in \mathcal{I}(A,b)$, there exists $x \in P$ such that $A_J x = b_J$ and $A_k x < b_k$ for all $k \in [q]\backslash J$. In particular, $x \in P(J)$ and $A_i x < b_i$ and then $x \notin P(I)$. Thus, $P(I) \neq P(J)$ which proves the mapping is injective.

We already proved that $F = P(I_{A,b}(F))$. Then, by bijectivity, $F \mapsto I_{A,b}(F)$ is the inverse of $I \mapsto P(I)$.

3.12.(c) Since $I, J \in \mathcal{I}(A,b)$, there exist $x, y \in P$, such that $A_I x = b_I$, $A_k x < b_k$ for all $k \in [q]\backslash I$ and $A_J y = b_J$, $A_k x < b_k$ for all $k \in [q]\backslash J$. Then, for $i \in I \cap J$, $A_i(x+y)/2 = b_i$ and for $k \in [q]\backslash(I \cap J)$, $A_k(x+y)/2 < b_k$. Thus, $I \cap J = I_{A,b}(x + y/2)$ as $(x+y)/2$ is in P, we conclude $I \cap J \in \mathcal{I}(A,b)$.

3.12.(d) By 3.12.(b), there exists $I \in \mathcal{I}(A,b)$ such that $F = P(I)$. Then, for all $x \in F$, $A_I x = b_I$ and then $I_{A,b}(x) \subset I$. We thus have that $I_{A,b}(F) = \cap_{x \in F} I_{A,b}(x) \subset I$. Moreover, by 3.12.(a), for all $x \in \mathrm{ri}(F)$, $I_{A,b}(x) = I$. We then have for $x \in \mathrm{ri}(F)$, $I = I_{A,b}(x) \subset \cap_{x \in F} I_{A,b}(x) = I_{A,b}(F) \subset I$, Thus, we have $I_{A,b}(x) = I_{A,b}(F)$ for all $x \in \mathrm{ri}(F)$.

3.12.(e) The implications ($\Rightarrow$) are easy. The backward implications ($\Leftarrow$) are obtained because $I \mapsto P(I)$ and $F \mapsto I_{A,b}(F)$ are the inverse of each other.

3.12.(f) We saw in the proof of 3.12.(a) that $\operatorname{Aff} P(I) = H_I$. Thus, $\dim P(I) = \dim H_I = \dim \operatorname{Ker}(A_I) = d - \operatorname{rg}(A_I)$, by the rank theorem. The other equality is obtained by the one-to-one correspondence in 3.12.(b). □

We are now able to prove 3.8.(e) and 3.8.(f) which stated that $\mathcal{F}(P)$ is finite and that

$$\operatorname{ri}(P) = P \setminus \bigcup_{F \in \mathcal{F}(P) \setminus \{P\}} F \tag{3.29}$$

Proof of 3.8.(e) *and* 3.8.(f). The point 3.8.(e) is a consequence of 3.12.(b): $\sharp\mathcal{F}(P) = \sharp\mathcal{F}(P)\setminus\{\emptyset\}+1 = \sharp\mathcal{I}(A,b)+1 \leqslant \sharp 2^{[q]}+1 = 2^q+1 < +\infty$.

Let $I = I_{A,b}(P)$ and $x \in P$. We first notice that $\exists j \in [q]\setminus I$, $A_j x = b_j$ if and only if $x \in \bigcup_{F\in\mathcal{F}(P)\setminus\{P\}} F$:

$$\begin{aligned}
\exists j \in [q]\setminus I,\ A_j x = b_j &\iff I_{A,b}(x) \neq I && \text{(3.30a)}\\
&\iff P(I_{A,b}(x)) \neq P && \text{(3.30b)}\\
&\iff \exists F \in \mathcal{F}(P)\setminus\{P\},\ x \in F && \text{(3.30c)}\\
&\iff x \in \bigcup_{F\in\mathcal{F}(P)\setminus\{P\}} F && \text{(3.30d)}
\end{aligned}$$

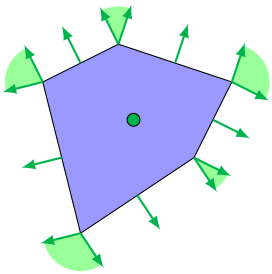
By point 3.12.(a), we have $\operatorname{ri}(P) = \{x \in P \,|\, \forall j \in [q]\setminus I,\ A_j x < b_j,\} = P\setminus\{x \in P \,|\, \exists j \in [q]\setminus I,\ A_j x = b_j\}$ from which we get the result. □

3.2.2 Normal fan

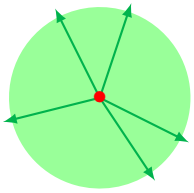
The normal fan is the collection of the normal cones of all faces of a polyhedron. See [LR08] for a review of normal fan properties.

The *normal cone* of a convex set $C \subset \mathbb{R}^m$ at the point x is the set $N_C(x) := \{\alpha \in \mathbb{R}^m \mid \forall y \in C,\ \alpha^\top(y-x) \leqslant 0\}$. More generally, for a set $E \subset C$, we define $N_C(E) := \bigcap_{x\in E} N_C(x)$.

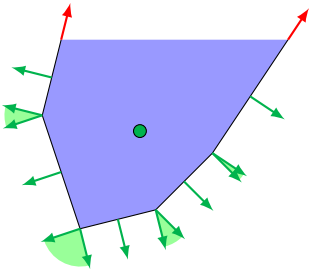
We also remark that by definition P^ψ in Section 3.1.3, $x \in P^\psi$ if and only if $-\psi \in N_P(x)$. Indeed, both are equivalent to $\psi^\top x \leqslant \psi^\top y$ for all $y \in P$.



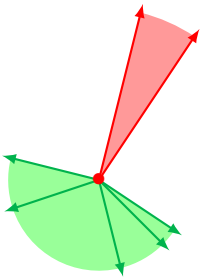
(a) A polytope P and its normal fan $\mathcal{N}(P)$



(b) The recession cone of $\operatorname{rc} P = \{0\}$ in red and its normal fan $\mathcal{N}(P)$ in green. $\operatorname{supp}(\mathcal{N}) = \mathbb{R}^2 = \{0\}^\circ$



(c) A polyhedron P and its normal fan.



(d) The recession cone of P in red and its normal fan $\mathcal{N}(P)$ in green.

Figure 3.6: Examples of polyhedra and their normal fans and recession cones.

Definition 3.13 (Normal Fan)**.** *The* normal fan[4] *of a convex set C is the collection of polyhedral cones*

$$\mathcal{N}(C) := \{N_C(x) \mid x \in C\} \tag{3.31}$$

We say that two convex sets C and C' are normally equivalent *if they have the same normal fan : $\mathcal{N}(C) = \mathcal{N}(C')$.*

We now state basic properties of normal fans. In particular, we show that the collection of active constraints sets $\mathcal{I}(A, b)$ provides V-representations of the normal cones in $\mathcal{N}(P)$.

Proposition 3.14 (Basic properties of normal fans)**.** *Let $P := \{x \in \mathbb{R}^n \mid Ax \leqslant b\}$, we have*

3.14.(a) *The support of the normal fan $\mathcal{N}(P)$ can be expressed geometrically as the polar of the recession cone of P,* i.e.

$$\operatorname{supp} \mathcal{N}(P) = \operatorname{rc}(P)^\circ \tag{3.32}$$

3.14.(b) *For all $x \in P$, we have $N_P(x) = \operatorname{Cone}(A_I)$ with $I := I_{A,b}(x)$.*

3.14.(c) *The mapping*

$$\begin{cases} \mathcal{I}(A, b) & \to \mathcal{N}(P) \\ I & \mapsto \operatorname{Cone}(A_I) \end{cases} \tag{3.33}$$

is a one-to-one correspondence whose inverse is $N \mapsto I_N := \{i \in \operatorname{supp} \mathcal{I}(A, b) \mid A_i^\top \in N\}$.

3.14.(d) *$\mathcal{N}(P)$ is a finite collection of normal cones.*

3.14.(e) *For $I, J \in \mathcal{I}(A, b)$ and $N, N' \in \mathcal{N}(P)$*

$$I \subset J \iff \operatorname{Cone}(A_I) \subset \operatorname{Cone}(A_J) \tag{3.34a}$$
$$N \subset N' \iff I_N \subset I_{N'} \tag{3.34b}$$

3.14.(f) *For $I, J \in \mathcal{I}(A, b)$ and $N, N' \in \mathcal{N}(P)$*

$$\operatorname{Cone}(A_{I \cap J}) = \operatorname{Cone}(A_I) \cap \operatorname{Cone}(A_J) \tag{3.35a}$$
$$I_{N \cap N'} = I_N \cap I_{N'} \tag{3.35b}$$

3.14.(g) *The normal fan $\mathcal{N}(P)$ is a fan, i.e. a polyhedral complex whose elements are cones, which by definition means:*

-For all $N, N' \in \mathcal{N}(P)$, $N \cap N'$ is a face of both N and N'.

-For all $N \in \mathcal{N}(P)$, all non-empty faces of N belong to $\mathcal{N}(P)$.

3.14.(h) *For $I \in \mathcal{I}(A, b)$ and $N \in \mathcal{N}(P)$,*

$$\dim\big(\operatorname{Cone}(A_I)\big) = \operatorname{rg}(A_I) \tag{3.36a}$$
$$\dim(N) = \operatorname{rg}(A_{I_N}) \tag{3.36b}$$

[4]Sometimes called *outer* normal cones and fan, as opposed to *inner* cones obtained either by inverting the inequality in the definition of the normal cone or by taking the opposite cones respect to the origin.

3.14.(i) *The mapping*

$$\begin{cases} \mathcal{F}(P)\backslash\{\emptyset\} & \to \mathcal{N}(P) \\ F & \mapsto N_P(F) \end{cases} \tag{3.37}$$

is a one-to-one correspondence whose inverse is $N \mapsto P^{-N}$.

Moreover, these mapping are composition of the later mappings as described in Fig. 3.7.

3.14.(j) *For* $F, F' \in \mathcal{F}(P)\backslash\{\emptyset\}$ *and* $N, N' \in \mathcal{N}(P)$

$$N \subset N' \iff P^{-N} \supset P^{-N'} \tag{3.38a}$$
$$F \subset F' \iff N_P(F) \supset N_P(F') \tag{3.38b}$$

3.14.(k) *For* $F \in \mathcal{F}(P)\backslash\{\emptyset\}$ *and* $N \in \mathcal{N}(P)$*, we have*

$$N_P(F) = N_P(x), \quad \text{for every } x \in \mathrm{ri}(F) \tag{3.39a}$$
$$P^{-N} = P^{-\psi}, \quad \text{for every } \psi \in \mathrm{ri}(N). \tag{3.39b}$$

We sum up the propositions Propositions 3.12 and 3.14 in figure 3.7, that shows that, if P is a polyhedron, its normal fan $\mathcal{N}(P)$, its set of non-empty faces $\mathcal{F}(P) \setminus \{\emptyset\}$ and its collection of sets of active constraints $\mathcal{I}(A, b)$ are in one-to-one correspondence. Furthermore, the orders are preserved or inverted by the correspondences.

The collection of active constraints sets $\mathcal{I}(A, b)$ then provides H-representations of the faces of P and V-representations of the normal cones in $\mathcal{N}(P)$. Indeed, each normal cone $N \in \mathcal{N}(P)$ can be written $\mathrm{Cone}(A_I)$ then $A_I^\top$ is a V-representation of N. Similarly, for a non-empty face $F \in \mathcal{F}(P)\backslash\{\emptyset\}$, there exists $I \in \mathcal{I}(A, b)$ with $F = \{x \in \mathbb{R}^d \,|\, Ax \leqslant b,\, A_I x = b_I\}$. Then, $\big((A, -A_I), (b, -b_I)\big)$ is a H-representation of F.

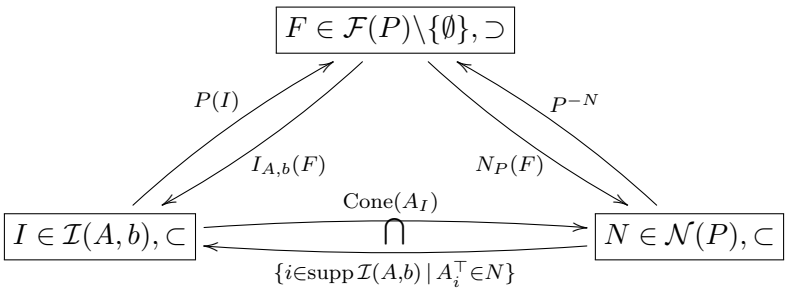
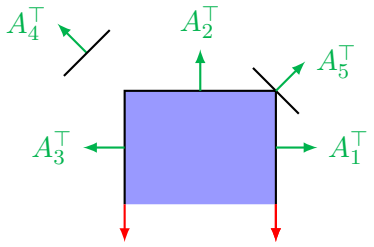


Figure 3.7: Monotonous one-to-one correspondences between normal fan, collection of active constraints sets and set of faces of a non-empty polyhedron $P = \{x \mid Ax \leqslant b\}$. The sign $\cap$ means that the mappings are morphisms for the intersection.

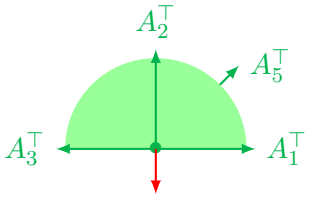
We illustrate these correspondences on an example in Table 3.1.

Proof. 3.14.(a) If $\psi \in \mathbb{R}^d$, we have

$$\psi \in \mathrm{supp}\,\mathcal{N}(P) \iff \exists N \in \mathcal{N}(P),\ \psi \in N \tag{3.40a}$$
$$\iff \exists x \in P,\ \psi \in N_P(x) \tag{3.40b}$$
$$\iff \exists x \in P,\ x \in P^{-\psi} \tag{3.40c}$$
$$\iff P^{-\psi} \neq \emptyset \tag{3.40d}$$
$$\iff \psi \in \mathrm{rc}(P)^\circ \qquad \text{by 3.8.(d)} \tag{3.40e}$$



(a) The polyhedron $P = [-1, 0] \times \mathbb{R}_-$



(b) The normal fan $\mathcal{N}(P)$ of P.

Figure 3.8: Illustration of correspondences of Table 3.1.

F	I	N	$\dim(N)$	$\dim(F)$
$\{(-1,0)\}$	$\{2,3\}$	$\mathbb{R}_- \times \mathbb{R}_+$	2	0
$\{(0,0)\}$	$\{1,2,5\}$	$\mathbb{R}_+ \times \mathbb{R}_+$	2	0
$\{0\} \times \mathbb{R}_-$	$\{1\}$	$\mathbb{R}_+ \times \{0\}$	1	1
$[-1,0] \times \{0\}$	$\{2\}$	$\{0\} \times \mathbb{R}_+$	1	1
$\{-1\} \times \mathbb{R}_-$	$\{3\}$	$\mathbb{R}_- \times \{0\}$	1	1
$[-1,0] \times \mathbb{R}_-$	$\emptyset$	$\{(0,0)\}$	0	2

Table 3.1: Correspondences illustrated in Fig. 3.8 between the normal fan $\mathcal{N}(P)$, the collection of active constraints sets $\mathcal{I}(A,b)$ and the collection of non-empty faces $\mathcal{F}(P)\backslash\{\emptyset\}$ for $P = [-1,0] \times \mathbb{R}_-$ given by the H-representation (A,b) with

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ -1 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}.$$

Indeed, we already saw that $x \in P^{-\psi}$ is equivalent to $\psi \in N_P(x)$.

3.14.(b) ($\supset$) Let $\psi = A_I^\top \mu$ with $\mu \in \mathbb{R}_+^I$. For $y \in P$, we have that $\psi^\top(y-x) = \mu^\top A_I(y-x) = \mu^\top(A_I y - b_I) \leqslant 0$, since $A_I x = b_I$. Then, $\psi \in N_P(x)$.

($\subset$) By contraposition, assume that there is no $\mu \in \mathbb{R}_+^I$ such that $A_I^\top \mu = \psi$. By Farkas lemma, there exists $z \in \mathbb{R}^d$ such that $A_I z \geqslant 0$ and $\psi^\top z < 0$. We define $y := x - \varepsilon z$ for $\varepsilon > 0$. Since for all $j \in [q]\backslash I$, $A_j x < b_j$, we can take $\varepsilon > 0$ small enough such that $A_j y < b_j$ for all $j \in [q]\backslash I$. Moreover, $A_I y = b_I - \varepsilon A_I z \leqslant b_I$. Then, for $\varepsilon > 0$ small enough, $y \in P$. Finally, $\psi^\top(y-x) = -\varepsilon\psi^\top z > 0$, and thus $\psi \notin N_P(x)$.

3.14.(c) We start by proving that for $j \in \operatorname{supp}\mathcal{I}(A,b)$ and $I \in \mathcal{I}(A,b)$, we have

$$A_j^\top \in \operatorname{Cone}(A_I) \iff j \in I \tag{3.41}$$

Indeed, the implication ($\Leftarrow$) is trivial. We now assume $j \notin I$. Since $I \in \mathcal{I}(A,b)$, there exists x in P such that $A_I x = b_I$ and $A_j x < b_j$. Moreover, since $j \in \operatorname{supp}\mathcal{I}(A,b)$, there exists $y \in P$ such that $A_j y = b_j$. We set $z = x - y$, then we have that $A_j z = A_j x - b_j < 0$ and $A_I z = b_I - A_I x \geqslant 0$. Then, by Farkas lemma, there is no $\mu \in \mathbb{R}_+^I$ such that $A_j^\top = A_I^\top \mu$. Thus, $A_j^\top \notin \operatorname{Cone}(A_I)$ leading to Eq. (3.41).

The mapping $I \mapsto \operatorname{Cone}(A_I)$ from $\mathcal{I}(A,b)$ to $\mathcal{N}(P)$ is well-defined and surjective by 3.14.(b). Let I and J be two active constraints sets in $\mathcal{I}(A,b)$. If $J \subset I$, it is trivial that $\operatorname{Cone}(A_J) \subset \operatorname{Cone}(A_I)$. Reciprocally if $\operatorname{Cone}(A_J) \subset \operatorname{Cone}(A_I)$, then, for all $j \in J$, $A_j^\top \in \operatorname{Cone}(A_I)$ and by Eq. (3.41) we have $j \in I$ for all $j \in J$, leading to $J \subset I$. Thus, $\operatorname{Cone}(A_J) \subset \operatorname{Cone}(A_I)$ is

equivalent to $J \subset I$. In particular, we have that $\mathrm{Cone}(A_I) = \mathrm{Cone}(A_J)$ if and only if $I = J$. Proving that $I \mapsto \mathrm{Cone}(A_I)$ is injective.

It remains to prove that $N \mapsto \{i \in \operatorname{supp}\mathcal{I}(A,b) \,|\, A_i^\top \in N\}$ is the inverse function. Let $I \in \mathcal{I}(A,b)$, we now prove that $I = \{i \in \operatorname{supp}\mathcal{I}(A,b) \,|\, A_i^\top \in \mathrm{Cone}(A_I)\}$. The inclusion $\subset$ is easy since $A_i^\top \in \mathrm{Cone}(A_I)$ for all $i \in I$. For the other inclusion, let $j \in [q]\backslash I$. If $j \notin \operatorname{supp}\mathcal{I}(A,b)$, we have that j does not belong to $\{i \in \operatorname{supp}\mathcal{I}(A,b) \,|\, A_i^\top \in \mathrm{Cone}(A_I)\}$. If $j \in \operatorname{supp}\mathcal{I}(A,b)$, by Eq. (3.41), $A_j^\top \notin \mathrm{Cone}(A_I)$ and then $j \notin \{i \in \operatorname{supp}\mathcal{I}(A,b) \,|\, A_i^\top \in \mathrm{Cone}(A_I)\}$.

3.14.(d) $\sharp\mathcal{N}(P) = \sharp\mathcal{I}(A,b) \leqslant 2^q < +\infty$.

3.14.(e) We already proved $I \subset J \iff \mathrm{Cone}(A_I) \subset \mathrm{Cone}(A_J)$ in the proof of 3.14.(c). The other equivalence is obtained thanks to the one-to-one correspondence of 3.14.(c).

3.14.(f) ($\subset$) Let $\psi \in \mathrm{Cone}(A_{I\cap J})$, there exists $\mu \in \mathbb{R}_+^{I\cap J}$ such that $\psi = A_{I\cap J}^\top \mu$ then $\psi = A_{I\cap J}^\top \mu + A_{I\backslash J}^\top 0 \in \mathrm{Cone}(A_I)$. By symmetry, $\psi \in \mathrm{Cone}(A_J)$. Thus, $\mathrm{Cone}(A_{I\cap J}) \subset \mathrm{Cone}(A_I) \cap \mathrm{Cone}(A_J)$.

($\supset$) Let $\psi \notin \mathrm{Cone}(A_{I\cap J})$. If $\psi \notin \mathrm{Cone}(A_I)$, we have $\psi \notin \mathrm{Cone}(A_I) \cap \mathrm{Cone}(A_J)$. We now assume $\psi \in \mathrm{Cone}(A_I)$. There exist $\mu \in \mathbb{R}_+^{I\cap J}$ and $\lambda \in \mathbb{R}_+^{I\backslash J}$ such that $\psi = A_{I\cap J}^\top \mu + A_{I\backslash J}^\top \lambda$. Since $I \in \mathcal{I}(A,b)$, there exists x such that $A_I x = b_I$ and for $j \notin I$, $A_j x < b_j$. Similarly, since $J \in \mathcal{I}(A,b)$, there y such that $A_J y = b_J$ and for $i \notin J$, $A_i x < b_i$. We define $z := y - x$.

$$\begin{aligned}
\psi^\top z &= \psi^\top y - \psi^\top x && \text{(3.42a)}\\
&= \mu^\top A_{I\cap J} y + \lambda^\top A_{I\backslash J} y - \mu^\top A_{I\cap J} x - \lambda^\top A_{I\backslash J} x && \text{(3.42b)}\\
&= \mu^\top b_{I\cap J} + \lambda^\top A_{I\backslash J} y - \mu^\top b_{I\cap J} - \lambda^\top b_{I\backslash J} && \text{(3.42c)}\\
&= \lambda^\top (A_{I\backslash J} y - b_{I\backslash J}) < 0 && \text{(3.42d)}
\end{aligned}$$

Indeed, the last term is negative because for all $i \in I\backslash J, A_i y < b_i$ and because, since $\psi \in \mathrm{Cone}(A_I)\backslash \mathrm{Cone}(A_{I\cap J})$, there exists at least one index of λ which is positive. Moreover, $A_J z = A_J y - A_J x = b_J - A_J x \geqslant 0$. To sum up, we have the existence of z such that $\psi^\top z < 0$ and $A_J z \geqslant 0$. By Farkas lemma, there is no $\nu \in \mathbb{R}_+^J$ such that $\psi = A_J^\top \nu$, i.e. $\psi \notin \mathrm{Cone}(A_J)$. In particular, $\psi \notin \mathrm{Cone}(A_I) \cap \mathrm{Cone}(A_J)$ which concludes the proof.

The other equality is a direct consequence of the correspondence 3.14.(c).

3.14.(g) - Let $N, N' \in \mathcal{N}(P)$, there exists $x, x' \in P$ such that $N = N_P(x)$ and $N' = N_P(x')$. Let $I = I_{A,b}(x)$ and $x' = I_{A,b}(x')$. In the proof of 3.12.(c), we saw that $I_{A,b}(\frac{x+x'}{2}) = I_{A,b}(x) \cap I_{A,b}(x') = I \cap I'$. Thus, by Item 3.14.(f) and 3.14.(b), we get $N \cap N' = \mathrm{Cone}(A_{I\cap I'}) = N_P(\frac{x+x'}{2})$. We have that $N \cap \{\psi \in \mathbb{R}^d \,|\, (x'-x)^\top \psi \leqslant 0\}$ since $N = N_P(x)$ and $x' \in P$. We now prove that $N \cap N' = N_P(\frac{x+x'}{2})$ is the face $N \cap \{\psi \in \mathbb{R}^d \,|\, (x'-x)^\top \psi = 0\}$ of N. Let $\psi \in N$ such that $(x'-x)^\top \psi = 0\}$, we have $\psi^\top (y - \frac{x+x'}{2}) = \psi^\top (y-x) + \frac{1}{2}\psi^\top (x'-x) = \psi^\top (y-x) \leqslant 0$ and then $\psi \in N_P(\frac{x+x'}{2})$. Reciprocally, assume $\psi \in N_P(\frac{x+x'}{2})$. Then, by taking $y = x \in P$ we have $\frac{1}{2}\psi^\top (x-x') = \psi^\top (y - \frac{x+x'}{2}) \leqslant 0$ and by taking $y = x' \in P$, we get $\frac{1}{2}\psi^\top (x'-x) = \psi^\top (y - \frac{x+x'}{2}) \leqslant 0$ and thus $(x'-x)^\top \psi = 0$. Finally, $N_P(\frac{x+x'}{2}) = \mathrm{Cone}(A_{I\cap I'}) \subset \mathrm{Cone}(A_I) = N$. By symmetry, we also have that $N \cap N'$ is a face of N'.

-Let N' a non-empty face of N. By definition of a face, there exists $z \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that $N' = N \cap \{\psi \in \mathbb{R}^d \,|\, z^\top \psi = \beta\}$ and $P \subset \{\psi \in \mathbb{R}^d \,|\, z^\top \psi \leqslant \beta\}$. Since $0 \in P$, we have $0 \leqslant \beta$. Moreover, for all $\psi \in N$ and all $\lambda \in \mathbb{R}_+^\star$, $\lambda \psi^\top z \leqslant \beta$, in particular by letting λ goes to infinity, we have $\psi^\top z \leqslant 0$ for all $\psi \in N$. If $\beta > 0$, then we would have $N' = \{\psi \in \mathbb{R}^d \,|\, \psi^\top z = \beta\} \cap N = \emptyset$. Since we assumed that N' is not empty we have $\beta = 0$.

Let $x \in P$ such that $N = N_P(x)$ and $I \in \mathcal{I}(A,b)$ such that $I = I_{A,b}(x)$. By Item 3.14.(b), $N = \mathrm{Cone}(A_I)$. Since, $N \subset \{\psi \in \mathbb{R}^d \,|\, \psi^\top z \leqslant 0\}$ and as $A_i^\top$ belongs to $N = \mathrm{Cone}(A_I)$ for all $i \in I$, we have $A_I z \leqslant 0$. Then, there exists $\varepsilon > 0$ such that $A(x + 2\varepsilon z) \leqslant b$, i.e. $x + 2\varepsilon z \in P$. By convexity, we also have $x + \varepsilon z \in P$. We now prove that $N_P(x+\varepsilon z) = N' = N \cap \{\psi \in \mathbb{R} \,|\, z^\top \psi = 0\}$.

($\supset$) Let $\psi \in N' = N \cap \{\psi \in \mathbb{R} \,|\, z^\top \psi = 0\}$. Then, for every $y \in P$, we have $\psi^\top (y - (x + \varepsilon z)) = \psi^\top (y - x) \leqslant 0$ and thus $\psi \in N_P(x + \varepsilon z)$.

($\subset$) Let $\psi \in N_P(x+\varepsilon z)$. By taking $y = x \in P$ we have $-\varepsilon^\top z = \psi^\top(y-(x+\varepsilon)) \leqslant 0$ and by taking $y = x + 2\varepsilon \in P$, we have $\varepsilon\psi^\top z = \psi^\top(y-(x+\varepsilon)) \leqslant 0$, thus $\psi^\top z = 0$ and $N_P(x+\varepsilon z) \subset \{\psi \in \mathbb{R} \,|\, z^\top\psi = 0\}$. We eventually prove that $N_P(x+\varepsilon z) \subset N = N_P(x)$. Let $j \notin I = I_{A,b}(x)$, we have $A_j x < b_j$ and, since $x + 2\varepsilon z \in P$, $A_j(x+2\varepsilon z) \leqslant b_j$ leading to $A_j(x+\varepsilon z) < b_j$. In particular, $j \notin I_{A,b}(x+\varepsilon z)$ and thus $I_{A,b}(x+\varepsilon z) \subset I_{A,b}(x)$. By 3.14.(b) and 3.14.(e), we conclude $N_P(x+\varepsilon z) \subset N = N_P(x)$.

3.14.(h) We have that $\mathrm{Cone}(A_I) \subset A_I\mathbb{R}^I$. Moreover, let $a = \sum_{i\in I} A_i^\top$ there exists $\varepsilon > 0$ such that $B(a,\varepsilon) \cap A_I\mathbb{R}^I \subset \mathrm{Cone}(A_I)$. Thus, $\dim\big(\mathrm{Cone}(A_I)\big) = \dim(A_I^\top\mathbb{R}^I) = \mathrm{rg}(A_I^\top) = \mathrm{rg}(A_I)$. The other equality is obtained by the correspondence in 3.14.(c).

3.14.(i) It is enough to prove that we have the composition of functions summed up in Fig. 3.7. We first prove that $N_P(F) = \mathrm{Cone}(A_{I_{A,b}(F)})$, for $F \in \mathcal{F}(P)\backslash\{\emptyset\}$. We denote by $\mathcal{I}_F$ the set $\{I_{A,b}(x) \,|\, x \in F\}$. By applying successively 3.14.(b) and 3.14.(f), as $\mathcal{I}_F \subset \mathcal{I}(A,b)$ is finite, we have

$$
\begin{aligned}
N_P(F) &= \cap_{x\in F} N_P(x) && \text{(3.43a)}\\
&= \cap_{x\in F} \mathrm{Cone}(A_{I_{A,b}(x)}) && \text{(3.43b)}\\
&= \cap_{I\in\mathcal{I}_F} \mathrm{Cone}(A_I) && \text{(3.43c)}\\
&= \mathrm{Cone}(A_{\cap_{I\in\mathcal{I}_F} I}) && \text{(3.43d)}\\
&= \mathrm{Cone}(A_{I_{A,b}(F)}) && \text{(3.43e)}
\end{aligned}
$$

We now prove that $P^{-N} = P(I_N)$ with $I_N := \{i \in \mathrm{supp}\,\mathcal{I}(A,b) \,|\, A_i^\top \in N\}$. Let $x \in P$.

$$
\begin{aligned}
x \in P(I_N) &\iff \forall i \in I_N,\ A_i x = b_i && \text{by definition of } P(I_N) && \text{(3.44a)}\\
&\iff I_N \subset I_{A,b}(x) && \text{by definition of } I_{A,b}(x) && \text{(3.44b)}\\
&\iff \mathrm{Cone}(A_{I_N}) \subset \mathrm{Cone}(A_{I_{A,b}(x)}) && \text{by 3.14.(e)} && \text{(3.44c)}\\
&\iff N \subset N_P(x) && \text{by 3.14.(c) } \mathrm{Cone}(A_{I_N}) = N \text{ and by 3.14.(b)} && \text{(3.44d)}\\
&\iff \forall\psi \in N, \forall y \in P, \psi^\top(y-x) \leqslant 0 && && \text{(3.44e)}\\
&\iff x \in P^{-N} && && \text{(3.44f)}
\end{aligned}
$$

3.14.(j) It is a consequence of the composition in Fig. 3.7 and the monotonicity results 3.12.(e) and 3.14.(e).

3.14.(k) We just proved that $N_P(F) = \mathrm{Cone}(A_{I_{A,b}(F)})$. If $x \in \mathrm{ri}(F)$,by 3.12.(d), $I_{A,b}(F) = I_{A,b}(x)$ then $N_P(F) = \mathrm{Cone}(A_{I_{A,b}(x)})$. Finally, by 3.14.(b), we get $N_P(F) = N_P(x)$.

We now prove that for $\psi \in \mathrm{ri}(N)$, $P^{-\psi} = P^{-N}$. By definition of as P^{-N} as an intersection, $P^{-N} \subset P^{-\psi}$. By 3.14.(j), $N_P(P^{-\psi}) \subset N_P(P^{-N}) = N$. By Item 3.14.(g), $N_P(P^{-\psi}) = N_P(P^{-\psi}) \cap N$ is a face of N. Moreover, for all $x \in P^-\psi$ and $y \in P$ $\psi^\top(y-x) \leqslant 0$ thus $\psi \in N_P(P^{-\psi})$. In particular, $N_P(P^{-\psi}) \cap \mathrm{ri}(N) \neq \emptyset$. By 3.8.(f), $\mathrm{ri}(N) = P\backslash\bigcup_{N'\in\mathcal{F}(N)\backslash\{N\}} N'$ and thus we have $N_P(P^{-\psi}) = N$. By the correspondence Item 3.14.(i), we conclude $P^{-\psi} = P^{-N}$.

□

3.2.3 Vertices, maximal constraint sets and full-dimensional cones

We now focus on the collection of maximal elements of the normal fan $\mathcal{N}(P)$ and active constraints set $\mathcal{I}(A,b)$ and on the collection of minimal elements of the non-empty faces $\mathcal{F}(P)\backslash\{\emptyset\}$.

For a collection $\mathcal{C}$, we denote by $\overline{\mathcal{C}}$ the subcollection of its maximal elements according to the inclusion and by $\underline{\mathcal{C}}$ the subcollection of its minimal element:

$$\overline{\mathcal{C}} := \{E \in \mathcal{C} \,|\, \forall F \in \mathcal{C},\ E \subset F \Rightarrow F = E\} \tag{3.45a}$$
$$\underline{\mathcal{C}} := \{E \in \mathcal{C} \,|\, \forall F \in \mathcal{C},\ F \subset E \Rightarrow F = E\} \tag{3.45b}$$

Proposition 3.15. *Let $A \in \mathbb{R}^{q\times p}, b \in \mathbb{R}^q$, $P := \{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ and $r = \mathrm{rg}(A)$. If P is not empty, we have*

$$\overline{\mathcal{I}(A,b)} = \{I \in \mathcal{I}(A,b) \,|\, \mathrm{rg}(A_I) = r\} \tag{3.46a}$$
$$\overline{\mathcal{N}(P)} = \{N \in \mathcal{N}(P) \,|\, \dim(N) = r\} \tag{3.46b}$$
$$\underline{\mathcal{F}(P)\backslash\{\emptyset\}} = \{F \in \mathcal{F}(P) \,|\, \dim(F) = d - r\} \tag{3.46c}$$

Proof. We only prove that the minimal faces are the faces of dimension $d-r$. The other equalities follow from Proposition 3.14 and Proposition 3.12. For all face $F \in \mathcal{F}(P)\backslash\{\emptyset\}$, by Eq. (3.13b), we have that $\mathrm{Lin}(F) = \mathrm{Ker}(A) = \mathrm{Lin}(P)$ and then $\dim(F) \geqslant d - r$.

($\supset$) Let $F \in \mathcal{F}(P)$ such that $\dim(F) = d - r$. Let G a non-empty face of P included in F. There exist $I, J \in \mathcal{I}(A,b)$ such that $F = P(I)$ and $G = P(J)$. If $G \subsetneq P$, we have $P \cap H_J = P(J) \subsetneq P(I) = P \cap H_I$. Then, $H_J \subsetneq H_I$ and then $\dim(G) = \dim(H_J) < \dim(H_I) = \dim(F) = d - r$. We have a contradiction with the fact that all non-empty faces have dimension at least $d - r$. Thus, $G = F$ and $F \in \underline{\mathcal{F}(P)\backslash\{\emptyset\}}$.

($\subset$). Let $F \in \underline{\mathcal{F}(P)\backslash\{\emptyset\}}$. Then, $\mathcal{F}(F) = \{F, \emptyset\}$ and by 3.8.(f), $\mathrm{ri}(F) = F$. There exists $I \in \mathcal{I}(A,b)$, such that $F = P(I)$. We now prove that $\mathrm{Ker}(A_I) = \mathrm{Ker}(A)$. It is trivial that $\mathrm{Ker}(A) \subset \mathrm{Ker}(A_I)$. Let $v \in \mathrm{Ker}(A_I)$ and assume, by contradiction, that there exists $j \in [q]\backslash I$ such that $A_j \neq 0$. Without of generality, we choose j such that for all $j' \in [q]\backslash I$, $|A_j| \leqslant |A_{j'}|$. Then, by setting $\lambda = \frac{b_j - A_j x}{A_j v}$ we have that $A_I(x+\lambda v) = b_I$ and for all $j' \in [q]\backslash I$, $A_{j'}(x+\lambda v) \leqslant b_{j'}$. Moreover, $A_j(x + \lambda v) = b_j$ and thus $x + \lambda v \in P(I)\backslash \mathrm{ri}\left(P(I)\right) = F \backslash \mathrm{ri}(F) = \emptyset$. We then have a contradiction and $\mathrm{Ker}(A_I) = \mathrm{Ker}(A)$. We conclude by remarking that $\dim(F) = \dim\left(P(I)\right) = \dim \mathrm{Ker}(A_I) = \dim \mathrm{Ker}(A) = d - r$. □

In particular, when $r = d$, we define the following assumption.

Assumption 3.1. Let $P := \{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ be a non-empty polyhedron such that one of the following equivalent statements holds:

- P admits at least one vertex, i.e. $\mathrm{Vert}(P) \neq \emptyset$
- A is a full rank matrix: $\mathrm{rg}(A) = d$
- The lineality space of P is trivial: $\mathrm{Lin}(P) = \{0\}$.
- All maximal normal cones are full-dimensional cones.
- From every maximal active constraint set $I \subset \mathcal{I}(A,b)$, we can extract a basis $B_I \subset I$.

Under Assumption 3.1, the correspondences of Fig. 3.7 can be simplified as presented in Fig. 3.9. In particular, we replace the set of non-empty faces $\mathcal{F}(P)\backslash\{\emptyset\}$ by the set of vertices $\mathrm{Vert}(P)$. For $I \in \overline{\mathcal{I}(A,b)}$, the minimal non-empty faces $P(I)$ are then reduced to singleton $\{x(I)\}$. Since, $x(I)$ satisfies $A_I x(I) = b_I$ and since we can extract a basis B_I from I, $x(I)$ can only be equal to $A_{B_I}^{-1} b_{B_I}$. We say that $x(I)$ is a basic point. We thus replace the mapping $I \mapsto P(I)$ by $I \mapsto A_{B_I}^{-1} b_{B_I}$.

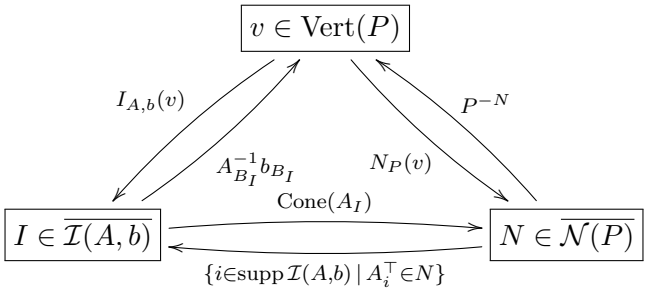


Figure 3.9: One-to-one correspondences between the collections of vertices of $P := \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$, maximal active constraint sets and full-dimensional normal cones of P, in the case where A is a full rank matrix.

Lemma 3.16. *Let $\mathcal{C}$ a collection of sets and $\mathcal{D}$ a polyhedral complex such that*

$$\overline{\mathcal{C}} \subset \overline{\mathcal{D}} \tag{3.47a}$$
$$\mathrm{supp}(\mathcal{C}) = \mathrm{supp}(\mathcal{D}) \tag{3.47b}$$

Then,

$$\mathcal{C} = \mathcal{D} \tag{3.48}$$

Moreover, if $\mathcal{C}$ is also polyhedral complex, then $\mathcal{C}$ and $\mathcal{D}$ are equal up-to the empty set.

$$\mathcal{C} \backslash \{\emptyset\} = \mathcal{D} \backslash \{\emptyset\} \tag{3.49a}$$
$$\mathcal{C} \cup \{\emptyset\} = \mathcal{D} \cup \{\emptyset\} \tag{3.49b}$$

Proof. If $\mathcal{C}$ or $\mathcal{D}$ is reduced to $\{\emptyset\}$ or $\emptyset$, the result is trivial. We assume $\mathcal{C}$ and $\mathcal{D}$ are different from $\{\emptyset\}$ and $\emptyset$ We first prove that $\overline{\mathcal{C}} = \overline{\mathcal{D}}$.

Let $D \in \mathcal{D}$. Since $D \subset \mathrm{supp}(\mathcal{C}) = \mathrm{supp}(\mathcal{D}) = \cup_{C \in \overline{\mathcal{C}}} C$, then $D = \cup_{C \in \overline{\mathcal{C}}} D \cap C$. In particular, there exists $C \in \overline{\mathcal{C}}$ such that $\dim(C \cap D) = \dim(D)$. Since $\overline{\mathcal{C}} \subset \overline{\mathcal{D}}$, we have $C \in \overline{\mathcal{D}}$, then D and C are two elements of $\overline{\mathcal{D}}$. However, since $\mathcal{D}$ is a polyhedral complex $C \cap D$ is a face of D. By Item 3.12.(f), the only face of D of dimension $\dim(D)$ is D itself. Then, $C \cap D = D$ and $D \subset C$. Finally, as $D \in D$ is maximal, we have $D = C \in \mathcal{C}$ and thus $\overline{\mathcal{C}} = \overline{\mathcal{D}}$.

In the case where $\mathcal{C}$ is a polyhedral complex, we just notice that a polyhedral complex, up to the empty set, is fully determined by its collection of maximal elements: $\mathcal{C} \cup \{\emptyset\} = \{F \mid \exists C \in \overline{C}, F \in \mathcal{F}(C)\}$.

□

3.3 Regular subdivisions and secondary fan

In this section, we explain the link between the *regular subdivisions* presented in the monograph of De Loera, Rambau and Santos [DLRS10] and the collection of *active constraint sets*. We first recall the definition of a regular subdivision $\mathcal{S}(V, \omega)$ of a *vector configuration* V induced by a *height vector* ω. We recall and prove formally that the collection of active constraints sets $\mathcal{I}(A, b)$ is exactly the regular subdivision $\mathcal{S}(A^\top, b)$ of the vector configuration of the rows of A induced by the height vector b.

We then introduce the *secondary fan* Σ-fan(V) thanks to the equivalence classes of height vectors b giving the same regular subdivision $\mathcal{S}(V, \omega)$. The secondary fan is a fundamental geometric object, which has strong links with elimination theory. Indeed, Gelfand, Kapranov and Zelevinsky [GKZ94] proved that the secondary fan was the normal fan of the secondary polytope which is equal to the Newton polytope of a discriminant.

3.3.1 Regular subdivisions

Let $(v_i)_{i\in[q]}$ be a family of vectors in $\mathbb{R}^d$. We define the matrix $V \in \mathbb{R}^{d\times q}$ whose columns are the vector v_i and refer for simplicity to the family $(v_i)_{i\in[q]}$ as V. As V is a finite set of vectors, each corresponding to n index in $[q]$, where we allow the set to have repeated vectors, V is often called a *vector configuration*. A *subdivision* is a collection of sets of indices, here in $[q]$, satisfying some properties we do not detail here (see [DLRS10, Definition 2.3.1]). Here, we are only interested in particular subdivisions, called regular subdivisions and presented hereafter.

Let $\omega \in \mathbb{R}^d$ be a vector called the *height vector*. For each $i \in [q]$, we draw the vectors $(v_i, \omega_i) \in \mathbb{R}^d \times \mathbb{R}$ and refer to this collection as the *lifted configuration*. Recall that the lower faces are the faces of a polyhedron $P \in \mathbb{R}^d \times \mathbb{R}$ that are visible from below, *i.e.* the faces F which can be written $F = P^{(\varphi,1)}$ for $\varphi \in \mathbb{R}^d$. We now consider all the lower faces of the lifted cone $LC_{V,\omega}$, defined as the conic hull of the vectors (v_i, ω_i) in $\mathbb{R}^d \times \mathbb{R}$. Each face can be represented by the set of indices i such that (v_i, ω_i) belongs to the face. The *regular subdivision* induced by ω is the collection of sets of indices representing the lower faces of $LC_{V,\omega}$. More formally, we have the following definition.

Definition 3.17 (Regular subdivision)**.** *Let $(v_i)_{i\in[q]}$ be a vector configuration represented by a matrix $V \in \mathbb{R}^{d\times q}$ and a vector $\omega \in \mathbb{R}^q$. We denote by $LC_{V,\omega}$ the lifted cone.*

$$LC_{V,\omega} := \mathrm{Cone}\left(\begin{pmatrix} v_i \\ \omega_i \end{pmatrix}_{i\in[q]}\right) \tag{3.50}$$

When $\omega \in V^\top\mathbb{R}^d + \mathbb{R}^q_+$, we define the regular subdivision *associated to the configuration of vector V induced by the height vector ω:*

$$\mathcal{S}(V,\omega) := \{I_F \,|\, F \in \mathcal{F}_{\mathrm{low}}(LC_{V,\omega})\} \tag{3.51}$$

where

$$I_F := \{i \in [q] \,|\, (v_i, \omega_i) \in F\}. \tag{3.52}$$

We illustrate these notions in Fig. 3.10.

The Lemma 2.5.11 in [DLRS10] states that the regular subdivision of a vector configuration is well-defined if the function $v_i \mapsto \omega_i$ differs from a non-negative function only by a linear function: this gives an interpretation why ω should be in $V^\top\mathbb{R}^d + \mathbb{R}^q_+$. The next proposition explains why this condition on ω is natural.

Lemma 3.18. *Let $A \in \mathbb{R}^{q\times d}$ and $b \in \mathbb{R}^q$. Then, the following statements are equivalent*

1. *The polyhedron $\{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ is not empty.*
2. *The set of lower faces is not trivial: $\mathcal{F}_{\mathrm{low}}(LC_{A^\top,b}) \neq \{\emptyset\}$.*
3. *$\mathcal{S}(A^\top, b)$ is defined meaning that $b \in A\mathbb{R}^d + \mathbb{R}^q_+$.*

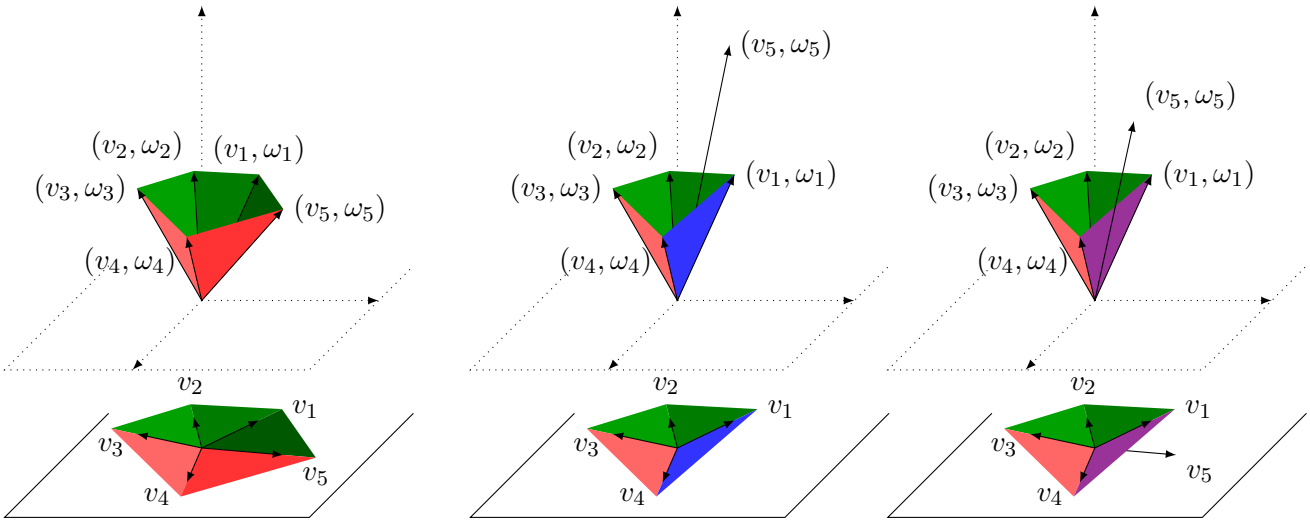
Proof. We first show that x does not verify $Ax \leqslant b$ if and only if the lower face $LC^{(-x,1)}_{A^\top,b}$ is empty:

$$\exists i \in [q],\ A_i x > b_i \iff \exists i \in [q],\ 0 > -x^\top A_i^\top + b_i \tag{3.53a}$$

$$\iff \exists (\alpha,\beta) \in LC_{A^\top,b},\ 0 > -x^\top\alpha + \beta \quad \text{as } LC_{A^\top,b} \text{ is the conic hull of } \begin{pmatrix} A_i^\top \\ b_i \end{pmatrix}_{i\in[q]} \tag{3.53b}$$

$$\iff \inf_{(\alpha,\beta)\in LC_{A^\top,b}} -x^\top\alpha + \beta = -\infty \qquad \text{because } LC_{A^\top,b} \text{ is a cone} \tag{3.53c}$$

$$\iff LC^{(-x,1)}_{A^\top,b} = \emptyset \tag{3.53d}$$



(a) For small ω_5, the lifted vector (v_5, ω_5) is salient and belongs to three lower faces: $\mathcal{S}(V, \omega) = \mathcal{S}_{co} \cup \big\{\{5\}, \{4, 5\}, \{1, 5\}\big\}$

(b) For large ω_5, the lifted vector (v_5, ω_5) is pointed inward, and belongs to no lower face: $\mathcal{S}(V, \omega) = \mathcal{S}_{co} \cup \big\{\{1, 4\}\big\}$

(c) In the limit case, the lifted vector (v_5, ω_5) belongs to one lower face: $\mathcal{S}(V, \omega) = \mathcal{S}_{co} \cup \big\{\{1, 4, 5\}\big\}$

Figure 3.10: Three lifted vector configurations, their projections and the regular subdivisions $\mathcal{I}(V, \omega)$ induced for different values of ω_5. We define $\mathcal{S}_{co} := \Big\{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{2, 3\}, \{3, 4\}\Big\}$

In particular, we have that $\{x \in \mathbb{R}^d \,|\, Ax \leqslant b\}$ is empty if and only if $\mathcal{F}_{\text{low}}(LC_{A^\top, b})$ is equal to $\{\emptyset\}$:

$$\{x \in \mathbb{R}^d \,|\, Ax \leqslant b\} = \emptyset \iff \forall x \in \mathbb{R}^d,\ \exists i \in [q],\ A_i x > b_i \tag{3.54a}$$

$$\iff \forall x \in \mathbb{R}^d, LC^{(-x,1)}_{A^\top, b} = \emptyset \tag{3.54b}$$

$$\iff \mathcal{F}_{\text{low}}(LC_{A^\top, b}) = \{\emptyset\} \tag{3.54c}$$

We now prove the equivalence 1. $\iff$ 3. If there exists $x \in \mathbb{R}^d$ such that $Ax \leqslant b$, then $b = Ax + (b - Ax) \in A\mathbb{R}^d + \mathbb{R}^q_+$. Reciprocally, if $b \in A\mathbb{R}^d + \mathbb{R}^q_+$, let $x \in \mathbb{R}^d$ and $\eta \in \mathbb{R}^q_+$ such that $b = Ax + \eta$, then $b - Ax = \eta \geqslant 0$ and thus $Ax \leqslant b$. □

We now state that the collection of active constraints sets $\mathcal{I}(A, b)$ is exactly the regular subdivision $\mathcal{S}(A^\top, b)$ of the vector configuration of the rows of A induced by the height vector b.

Proposition 3.19. *Let $A \in \mathbb{R}^{q \times d}$ and $b \in \mathbb{R}^q$. If one of the equivalent statements of Lemma 3.18 is true, we have*

$$\mathcal{I}(A, b) = \mathcal{S}(A^\top, b) \tag{3.55}$$

Proof. Let x such that $Ax \leqslant b$, by Eq. (3.53c), we know that $\inf_{(\alpha,\beta) \in LC_{A^\top, b}} -x^\top \alpha + \beta$ is different from $-\infty$. Since, $LC_{A^\top, b}$ is a cone, we have $-x^\top \alpha + \beta \geqslant 0$, for all $(\alpha, \beta) \in LC_{A^\top, b}$, otherwise $\inf_{(\alpha,\beta) \in LC_{A^\top, b}} -x^\top \alpha + \beta$ is equal to $-\infty$. Moreover, $LC_{A^\top, b}$ contains 0, thus $\min_{(\alpha,\beta) \in LC_{A^\top, b}} -x^\top \alpha +$

$\beta = 0$. We now show that $I_{A,b}(x) = I_{LC^{(-x,1)}_{A^\top,b}}$:

$$\begin{aligned}
i \in I_{A,b}(x) &\iff A_i x = b_i && (3.56a)\\
&\iff 0 = -x^\top A_i^\top + b_i && (3.56b)\\
&\iff \min_{(\alpha,\beta)\in LC_{A^\top,b}} -x^\top\alpha + \beta = -x^\top A_i^\top + b_i && (3.56c)\\
&\iff (A_i^\top, b_i) \in LC^{(-x,1)}_{A^\top,b} && (3.56d)\\
&\iff i \in I_{LC^{(-x,1)}_{A^\top,b}} && (3.56e)
\end{aligned}$$

We conclude the proof by using the definitions of regular subdivisions, lower faces and active constraints sets:

$$\begin{aligned}
\mathcal{S}(A^\top, b) &= \{I_F \,|\, F \in \mathcal{F}_{\text{low}}(LC_{A^\top,b})\} && (3.57a)\\
&= \{I_{LC^{(-x,1)}_{A^\top,b}} \,|\, x \in \mathbb{R}^d, I_{LC^{(-x,1)}_{A^\top,b}} \neq \emptyset\} && (3.57b)\\
&= \{I_{A,b}(x) \,|\, x \in \mathbb{R}^d, Ax \leqslant b\} && (3.57c)\\
&= \mathcal{I}(A, b) && (3.57d)
\end{aligned}$$

□

3.3.2 Secondary fan

We now introduce a fundamental notion in polyhedral geometry called the secondary fan. We refer to [DLRS10], in particular its section 5.2, for a complete introduction on secondary fans.

Definition 3.20. *Let $V \in \mathbb{R}^{d\times q}$ be a vector configuration. We define the secondary cone $\Sigma\text{-}C(V,\mathcal{I})$, and respectively its relatively open secondary cone $\Sigma\text{-}C^\circ(V,\mathcal{I})$[5], of a regular subdivision $\mathcal{I}$ of V as the set of height vectors ω such that $\mathcal{I}$ refines, resp equals, $\mathcal{S}(V,\omega)$.*

$$\Sigma\text{-}C(V,\mathcal{I}) := \{\omega \in \mathbb{R}^q \,|\, \mathcal{I} \preccurlyeq \mathcal{S}(V,\omega)\} \quad (3.58a)$$

$$\Sigma\text{-}C^\circ(V,\mathcal{I}) := \{\omega \in \mathbb{R}^q \,|\, \mathcal{I} = \mathcal{S}(V,\omega)\} \quad (3.58b)$$

The secondary fan is the set of all secondary cones of regular subdivisions.

$$\Sigma\text{-fan}(V) := \{\Sigma\text{-}C(V,\mathcal{I}) \,|\, \exists \omega_{\mathcal{I}} \in \text{Cone}(V^\top) + \mathbb{R}^q_+,\ \mathcal{I} = \mathcal{S}(V,\omega_{\mathcal{I}})\} \quad (3.59a)$$

$$\Sigma\text{-fan}^\circ(V) := \{\Sigma\text{-}C^\circ(V,\mathcal{I}) \,|\, \exists \omega_{\mathcal{I}} \in \text{Cone}(V^\top) + \mathbb{R}^q_+,\ \mathcal{I} = \mathcal{S}(V,\omega_{\mathcal{I}})\} \quad (3.59b)$$

We remark that $\Sigma\text{-fan}^\circ(V)$ is the set of equivalence classes of the relation $\sim_V$ defined as $\omega \sim_V \omega'$ if $\mathcal{S}(V,\omega) = \mathcal{S}(V,\omega')$. In particular, $\Sigma\text{-fan}^\circ(V)$ is a partition of $V^\top\mathbb{R} + \mathbb{R}^q_+$.

We now give a H-representations of secondary cones when the vector configuration is full dimensional.

Proposition 3.21. *Let $A \in \mathbb{R}^{d\times q}$ such that $\text{rg}(A) = d$ and $\mathcal{I}$ a regular subdivision of $A^\top$. For all $I \in \overline{\mathcal{I}}$, we can extract a basis $B_I \subset I$. Then, for all such choices of bases B_I for every $I \in \overline{\mathcal{I}}$, we have that*

$$\Sigma\text{-}C(A^\top,\mathcal{I}) = \left\{ b \in \text{Cone}(A) + \mathbb{R}^q_+ \;\middle|\; \forall I \in \overline{\mathcal{I}},\ \begin{array}{ll} A_i A_{B_I}^{-1} b_{B_I} = b_i, & \forall i \in I \\ A_i A_{B_I}^{-1} b_{B_I} \leqslant b_i, & \forall j \in [q]\backslash I \end{array} \right\} \quad (3.60a)$$

$$\Sigma\text{-}C^\circ(A^\top,\mathcal{I}) = \left\{ b \in \text{Cone}(A) + \mathbb{R}^q_+ \;\middle|\; \forall I \in \overline{\mathcal{I}},\ \begin{array}{ll} A_i A_{B_I}^{-1} b_{B_I} = b_i, & \forall i \in I \\ A_i A_{B_I}^{-1} b_{B_I} < b_i, & \forall j \in [q]\backslash I \end{array} \right\}. \quad (3.60b)$$

[5] We adapt here the notation in [DLRS10]. The sign $\circ$ in exponent refers to the relative interior and should not be confused with the polar of a set.

In particular, $\Sigma\text{-}C^\circ(A^\top,\mathcal{I}) = \operatorname{ri}\left(\Sigma\text{-}C(A^\top,\mathcal{I})\right)$.

Proof. Let $I \in \overline{\mathcal{I}}$ and $b \in \operatorname{Cone}(A) + \mathbb{R}^q_+$, we have that

$$\begin{aligned}
\exists I' \in \mathcal{I}(A,b), I \subset I' &\iff \exists x \in P,\ I \subset I_{A,b}(x) && \text{(3.61a)}\\
&\iff \exists x \in P,\ A_I x = b_I \text{ and } A_j x \leqslant b_j,\ \forall j \in [q]\backslash I && \text{(3.61b)}\\
&\iff A_I A_{B_I}^{-1} b_{B_I} = b_I \text{ and } A_j A_{B_I}^{-1} b_{B_I} \leqslant b_j,\ \forall j \in [q]\backslash I && \text{(3.61c)}
\end{aligned}$$

Indeed, since we can extract a basis $B_I \subset I$. The only x satisfying $A_I x = b_I$ is $A_{B_I}^{-1} b_{B_I}$. Similarly, we have

$$\begin{aligned}
\exists I' \in \mathcal{I}(A,b), I = I' &\iff \exists x \in P,\ I = I_{A,b}(x) && \text{(3.62a)}\\
&\iff \exists x \in P,\ A_I x = b_I \text{ and } A_j x < b_j,\ \forall j \in [q]\backslash I && \text{(3.62b)}\\
&\iff A_I A_{B_I}^{-1} b_{B_I} = b_I \text{ and } A_j A_{B_I}^{-1} b_{B_I} < b_j,\ \forall j \in [q]\backslash I && \text{(3.62c)}
\end{aligned}$$

Finally, since a regular subdivision is completely determined by its sets of maximal elements, we have

$$\begin{aligned}
\mathcal{I} \preccurlyeq \mathcal{I}(A,b) &\iff \forall I \in \overline{\mathcal{I}},\ \exists I' \in \mathcal{I}(A,b),\ I \subset I' && \text{(3.63a)}\\
\mathcal{I} = \mathcal{I}(A,b) &\iff \forall I \in \overline{\mathcal{I}},\ \exists I' \in \mathcal{I}(A,b),\ I = I' && \text{(3.63b)}
\end{aligned}$$

which concludes the proof. □

3.3.3 Regular triangulations

Regular triangulations are a particularly interesting classes of regular subdivisions are triangulations which give their name to the book of De Loera, Rambau and Santos [DLRS10] a monograph on the theory of regular subdivisions.

Definition 3.22 (Regular triangulation)**.** *Let $(v_i)_{i\in[q]}$ be a vector configuration represented by a matrix $V \in \mathbb{R}^{d\times q}$. We say that the regular subdivision $\mathcal{S}$ of the point configuration V is a regular* triangulation *if, for all $I \in \mathcal{S}$, the $(v_i)_{i\in I}$ is an independent family of vectors.*

We remark that if $\mathcal{I}$ is a regular triangulation of $A^\top$, then all maximal collections of indices $I \in \overline{\mathcal{I}}$ are basis. Indeed, $\operatorname{Cone}(A_I)$ is full dimensional and A_I is a full rank matrix thus, A_I is invertible.

We now state that the regular triangulations are the regular subdivisions with a generic height vector:

Proposition 3.23. *Let $(v_i)_{i\in[q]}$ be a vector configuration represented by a matrix $V \in \mathbb{R}^{d\times q}$ and a vector $\omega \in \mathbb{R}^q$. Then, the regular subdivision $\mathcal{S}(V,\omega)$ is a regular triangulation if and only if there exists a maximal relatively open cone S of Σ-fan$^\circ(V)$ such that $\omega \in S$.*

3.4 Basis decomposition theorems

In this section, we study the value of the standard linear problem:

$$\operatorname{val}(\mathrm{LP}^{=}_{A,b,c}) = \begin{cases} \min_{x\in\mathbb{R}^d} & c^\top x \\ \text{s.t.} & Ax = b \\ & x \geqslant 0 \end{cases} \qquad (\mathrm{LP}^{=}_{A,b,c})$$

In the standard form, $A \in \mathbb{R}^{q\times d}$ is a rectangular matrix whose width is larger than its heigth $q \leqslant d$. For $J \subset [d]$, we denote by $A_{\cdot,J} = (A_{i\in[q],j\in J})$ the submatrix obtained by taking the columns of indices in J.

Before proving Theorem 3.27, we recall some usual definitions and results in linear programming's theory that can be found in any standard linear programming book, e.g. [MG07].

Definition 3.24 (Basic point, reduced cost)**.** *We say that $B \subset [d]$ is a* basis *of A if the submatrix $A_{\cdot,B} = (a_i)_{i\in B}$, where a_i is the i-th column of A, is invertible.*

We define the basic point *associated to the basis B as the vector in $\mathbb{R}^d$, with coordinates $x_B := A_{\cdot,B}^{-1} b$ for $i \in B$ and 0 for $i \notin B$.*

A basis B is said to be admissible *(resp.* optimal*), if its associated basic point is an admissible (resp. optimal), solution of $(\mathrm{LP}^{=}_{A,b,c})$.*

Finally, the reduced cost *associated to a basis B is the vector*

$$(c_j - a_j^\top A_{\cdot,B}^{-1\top} c_B)_{j\in[d]} \tag{3.64}$$

Reduced cost is a key ingredient of the simplex method. In particular, it is well known that an admissible basis is optimal if and only if its reduced cost is non-negative. More formally, we have:

Lemma 3.25. *Let $B \subset [d]$ a basis. If $x_B := A_{\cdot,B}^{-1} b \geqslant 0$ and for all $j \in [d]$, $c_j - a_j^\top A_{\cdot,B}^{-1\top} c_B \geqslant 0$. Then, B is optimal and in particular,*

$$\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) = c_B^\top A_{\cdot,B}^{-1} b \tag{3.65}$$

Finally, we recall a classical generalization of Carathéodory's theorem for conic hull.

Lemma 3.26 (Carathéodory)**.** *Let $A \in \mathbb{R}^{q\times d}$ be a matrix and a subset of indices $J \subset [d]$ such that $span((a_i)_{i\in J}) := A_{\cdot,J}\mathbb{R}^J = \mathbb{R}^q$ where a_i is the i-th column of A.*

Consider a vector b in the conic hull of $(a_i)_{i\in J}$, i.e. *$b \in \mathrm{Cone}(A_{\cdot,J})$. Then, there exists a basis $B \subset J$ such that h is in the conic hull of $(a_i)_{i\in B}$, i.e. $b \in \mathrm{Cone}(A_{\cdot,B})$.*

Proof. Let $I \subset J$ be such that $b \in \mathrm{Cone}(A_{\cdot,I})$ and $(a_i)_{i\in I}$ is spanning $\mathbb{R}^q$. There exist $(\mu_i)_{i\in I} \in \mathbb{R}_+^I$ nonnegative coefficients such that $b = \sum_{i=1}^m \mu_i a_i$. Assume that I is not a basis, then $(a_i)_{i\in I}$ is not linearly independent, that is there exists a collection $(\lambda_i)_{i\in I} \in \mathbb{R}^I$ such that $\sum_{i\in I} \lambda_i a_i = 0$ with at least one λ_i different of zero, that can be assumed w.l.o.g positive.

Define $j := \arg\min_{i\in I|\lambda_i>0} \frac{\mu_i}{\lambda_i}$. Then, we have $a_j = -\sum_{i\in I\setminus\{j\}} \frac{\lambda_i}{\lambda_j} a_i$ and thus $b = \sum_{i\in I\setminus\{j\}} (\mu_i - \mu_j \frac{\lambda_i}{\lambda_j}) a_i$. In particular, $(a_i)_{i\in I\setminus\{j\}}$ is spanning $\mathbb{R}^l$. We now show that each coefficient in this sum is nonnegative, that is $b \in \mathrm{Cone}(A_{\cdot,I\setminus\{j\}})$. Note that $\lambda_j > 0$ and for all $i \in I$, $\mu_i \geqslant 0$. Thus, if $\lambda_i \leqslant 0$ we have $\mu_i - \mu_j \frac{\lambda_i}{\lambda_j} \geqslant 0$. Otherwise, $\lambda_i > 0$, and by definition of j, $\frac{\mu_i}{\lambda_i} \geqslant \frac{\mu_j}{\lambda_j}$ and thus $\mu_i - \mu_j \frac{\lambda_i}{\lambda_j} \geqslant 0$. Which shows that $b = \sum_{i\in I\setminus\{j\}} (\mu_i - \mu_j \frac{\lambda_i}{\lambda_j}) a_i \in \mathrm{Cone}(A_{\cdot,I\setminus\{j\}})$.

By induction, we drop indices until we get a basis B. □

We can now state a reformulation the basis decomposition theorem, initially proved by Walkup and Wets in [WW69], and then adapted to modern vocabulary by Sturmfels and Thomas in [ST97].

Theorem 3.27 (Basis decomposition theorem for linear problem in standard form)**.** *Let $A \in \mathbb{R}^{q\times d}$ be matrix of rank q, a right-hand side vector $b \in \mathbb{R}^q$ and $c \in \mathbb{R}^d$, then*

1. *$(\mathrm{LP}^{=}_{A,b,c})$ is primal admissible if and only if $b \in \mathrm{Cone}(A)$.*
2. *$(\mathrm{LP}^{=}_{A,b,c})$ is dual admissible if and only if $c \in A^\top\mathbb{R}^q + \mathbb{R}^d_+$.*
3. *If $\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) \in \mathbb{R}$, then $\mathcal{I}(A^\top, c)$ is well-defined and there exist $I \in \mathcal{I}(A^\top, c)$ with $b \in \mathrm{Cone}(A_{\cdot,I})$ and $B \subset I$ a basis of A extracted from I.*
4. *If $\mathcal{I}(A^\top, c)$ is well-defined, for all $I \in \mathcal{I}(A^\top, c)$ such that $b \in \mathrm{Cone}(A_{\cdot,I})$ and all $B \subset I$ basis extracted from I, we have that B is an optimal basis:*

$$\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) = c_B^\top A_{\cdot,B}^{-1} b. \tag{3.66}$$

Remark 3.28. *In [ST97], Sturmfels and Thomas present the basis decomposition theorem of Walkup and Wets with the modern definitions of regular subdivision. Their second point reads as* $(\mathrm{LP}^{=}_{A,b,c})$ *is bounded for all* $b \in \mathrm{Cone}(A)$ *and all* $c \in \mathbb{R}^d$ *if and only if* $\mathrm{Ker}(A) \cap \mathbb{R}^d_- = \{0\}$. *We see that with our version of the basis decomposition theorem, we have* $(\mathrm{LP}^{=}_{A,b,c})$ *is bounded for all* $b \in \mathrm{Cone}(A)$ *and all* $c \in \mathbb{R}^d$ *if and only if* $A^\top \mathbb{R}^q + \mathbb{R}^d_+$ *is equal to* $\mathbb{R}^d$. *We recall that by [JW92, Lemma 2.2], for 2 closed convex cones* K *and* L, *we have* $(K \cap L)^\circ = K^\circ + L^\circ$. *In particular,* $A^\top \mathbb{R}^q + \mathbb{R}^d_+ = \big(\mathrm{Ker}(A) \cap \mathbb{R}^d_-\big)^\circ$. *Thus,* $A^\top \mathbb{R}^q + \mathbb{R}^d_+ = \mathbb{R}^d$ *is equivalent to* $\mathrm{Ker}(A) \cap \mathbb{R}^d_- = \{0\}$ *and then to* $\mathrm{Ker}(A) \cap \mathbb{R}^d_+ = \{0\}$. *Like Walkup and Wets in [WW69], Thomas and Sturmfels worked with a fixed cost* c. *For simplicity, they assume in the rest of their paper that* $\mathrm{Ker}(A) \cap \mathbb{R}^d_+$ *is reduced to* $\{0\}$. *As we want to deal with varying* b *and* c, *we chose to extend the basis decomposition theorem to the case where* $\mathrm{Ker}(A) \cap \mathbb{R}^d_+$ *is not necessarily reduced to* $\{0\}$.

Proof. 1. $\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) > -\infty$ is equivalent to the existence of a primal admissible point $x \geqslant 0$ such that $Ax = b$ which is by definition $b \in \mathrm{Cone}(A)$.

2. By the equivalence 1. $\iff$ 3. in Proposition 3.19. $c \in A^\top \mathbb{R}^q + \mathbb{R}^d_+$ is equivalent to $\emptyset \neq \{\lambda \in \mathbb{R}^q \mid A^\top \lambda \leqslant c\}$ which is the dual admissible set of $(\mathrm{LP}^{=}_{A,b,c})$.

3. If $\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) \in \mathbb{R}$, by 1. and 2. $c \in \mathrm{Cone}(A^\top) + \mathbb{R}^d_+$ and $b \in \mathrm{Cone}(A)$. Let $D_{A,c} := \{\lambda \in \mathbb{R}^q \mid A^\top \lambda \leqslant c\}$. The support $\mathrm{supp}\,\mathcal{N}(D_{A,c})$ of the normal fan is equal to $\mathrm{Cone}(A)$ by 3.14.(a) and Proposition 3.7. Then, there exists a full dimensional cone $N \in \mathcal{N}(D_{A,c})$ such that $b \in N$. By 3.14.(c), there exists $I \in \mathcal{I}(A^\top, c)$ such that $N = \mathrm{Cone}(A_{\cdot,I})$. Since the cone $\mathrm{Cone}(A_{\cdot,I})$ is full dimensional, the matrix $A_{\cdot,I}$ has full rank and thus we can extract a basis B from I.

4. Let $J \in \mathcal{I}(A^\top, c)$, $b \in \mathrm{Cone}(A_{\cdot,J})$ and $B \subset I$ a basis. Since B is an extracted basis, we know that the matrix $A_{\cdot,I}$ has full rank and thus $(a_i)_{i \in J}$ is spanning $\mathbb{R}^q$. Then by Caratheodory's Lemma 3.26, there exists a basis $B_b \subset I$, such that $b \in \mathrm{Cone}(A_{\cdot,B_b})$. In particular, we have that $x_{B_b} = A^{-1}_{\cdot,B_b} b \geqslant 0$, thus B_b is an admissible basis. Moreover, by definition of $\mathcal{I}(A^\top, b)$ in (3.24), there exists $\lambda(I)$ such that

$$\forall i \in I, \quad a_i^\top \lambda(I) = c_i \tag{3.67a}$$

$$\forall j \in [d] \backslash I, \quad a_j^\top \lambda(I) < c_j. \tag{3.67b}$$

As $B_b \subset I$, by (3.67a), we have that, for all $i \in B_b$, $a_i^\top \lambda(I) = c_i$ which in turn implies $\lambda(I) = A^{-1\top}_{\cdot,B_b} c_{B_b}$. Thus, for all $j \in [m]$, we can compute the reduced cost coordinate $c_j - a_j^\top A^{-1\top}_{\cdot,B_b} q_{B_b} = c_j - a_j^\top \lambda(I) \geqslant 0$, by (3.67a) and (3.67b). By Lemma 3.25, B_b is an optimal basis, leading to

$$\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) = c^\top_{B_b} A^{-1}_{\cdot,B_b} b.$$

We now prove that the following formula does not depend on the choice of the extracted basis B_b. As $B \subset I$, for all $i \in B$, $a_i^\top \lambda(I) = c_i$, we also have $A^{-1\top}_{\cdot,B} c_B = \lambda(I) = A^{-1\top}_{\cdot,B_b} c_{B_b}$. Thus, $\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) = c^\top_{B_b} A^{-1}_{\cdot,B_b} b = c^\top_B A^{-1}_{\cdot,B} b$.

□

We can also have the same basis decomposition theorem for the linear problem in inequality form

$$\mathrm{val}(\mathrm{LP}^{=}_{A,b,c}) = \begin{cases} \min_{x \in \mathbb{R}^d} & c^\top x \\ \text{s.t.} & Ax \leqslant b \end{cases} \tag{$\mathrm{LP}^{\leqslant}_{A,b,c}$}$$

Theorem 3.29 (Basis decomposition for linear problem in inequality form)**.** *Let* $A \in \mathbb{R}^{q \times d}$ *matrix of rank* d, *a right-hand side vector* $b \in \mathbb{R}^q$ *and* $c \in \mathbb{R}^d$, *then*

1. $(\mathrm{LP}^{\leqslant}_{A,b,c})$ *is primal admissible if and only if* $b \in A\mathbb{R}^q + \mathbb{R}^d_+$.
2. $(\mathrm{LP}^{\leqslant}_{A,b,c})$ *is dual admissible if and only if* $c \in \mathrm{Cone}(-A^\top)$.

3. *If* $\mathrm{val}(\mathrm{LP}^{\leqslant}_{A,b,c}) \in \mathbb{R}$, *then* $\mathcal{I}(A,b)$ *is well-defined and there exist* $I \in \mathcal{I}(A,b)$ *with* $c \in \mathrm{Cone}(-A_I)$ *and* $B \subset I$ *a basis of* $A^\top$ *extracted from* I.

4. *If* $\mathcal{I}(A,b)$ *is well-defined, for all such* $I \in \mathcal{I}(A,b)$ *such that* $c \in \mathrm{Cone}(-A_I)$ *and* $B \subset I$ *a basis extracted from* I, *we have that* B *is an optimal basis:*

$$\mathrm{val}(\mathrm{LP}^{\leqslant}_{A,b,c}) = c^\top A_B^{\top -1} b_B \tag{3.68}$$

Proof. The dual of $(\mathrm{LP}^{\leqslant}_{A,b,c})$ is

$$\max_{\mu \in \mathbb{R}^q} -b^\top \mu \tag{3.69a}$$
$$A^\top \mu = -c \tag{3.69b}$$
$$\mu \geqslant 0 \tag{3.69c}$$

By remarking that $\max_\mu -b^\top \mu = -\min_\mu b^\top \mu$, this problem is equivalent to $(LP^{=}_{A^\top,b,-c})$ with $\mathrm{val}(\mathrm{LP}^{\leqslant}_{A,b,c}) = -\mathrm{val}(\mathrm{LP}^{=}_{A^\top,b,-c})$ when we have strong duality. We conclude by applying Theorem 3.27 to $(LP^{=}_{A^\top,b,-c})$. □

3.5 Coupling polyhedron, fibers and chamber complex

In this section, we are interested in the case where the polyhedron depends on two variables. For $T \in \mathbb{R}^{\ell \times n}$, $W \in \mathbb{R}^{\ell \times m}$ and $h \in \mathbb{R}^\ell$, we denote the *coupling polyhedron* by

$$P := \{(x,y) \in \mathbb{R}^{n+m} \mid Tx + Wy \leqslant h\}. \tag{3.70}$$

Let π be the projection of $\mathbb{R}^n \times \mathbb{R}^m$ onto $\mathbb{R}^n$ such that $\pi(x,y) = x$. For any set E of $\mathbb{R}^{n+m}$, we recall that the projection of E is

$$\pi(P) = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m,\ (x,y) \in E\}. \tag{3.71}$$

Recall that, for any $x \in \mathbb{R}^n$, the polyhedron P_x the *fiber* P_x of P at x along π as the projection of $P \cap \pi^{-1}(\{x\})$ on the space $\mathrm{Ker}(\pi)$. Alternatively, we have

$$P_x := \{y \in \mathbb{R}^m \mid Tx + Wy \leqslant h\} \tag{3.72}$$

Note that P_x is not empty if and only if $x \in \pi(P)$.

We recall that for any subset of indices $I \in \mathcal{I}\Big((T,W),h\Big)$, we denote by $P(I)$ the face

$$P(I) := \{(x,y) \in P \mid T_I x + W_I y = h_I\}. \tag{3.73}$$

Similarly, for $x \in \pi(P)$ and $I \in \mathcal{I}(W, h - Tx)$, we define

$$P_x(I) := \{y \in P_x \mid W_I y = h_I - T_I x\} \tag{3.74}$$

Finally, we define $\mathcal{G}_x$ the collection of *lifted faces* above x, i.e. the faces G of the coupling polyhedron such that x belongs to the relative interior of the projection $\pi\big(\mathrm{ri}(G)\big) = \mathrm{ri}\,\big(\pi(G)\big)$:

$$\mathcal{G}_x := \{G \in \mathcal{F}(P) \,|\, x \in \pi\big(\mathrm{ri}(G)\big)\} \tag{3.75}$$

3.5.1 Normal equivalence on the chamber complex

We now show that all fibers P_x are normally equivalent for x in the relative interior of a cell of the chamber complex $\mathcal{C}(P,\pi)$ defined in Definition 3.10.

Proposition 3.30 (Normal equivalence above a chamber)**.** *Let $\sigma \in \mathcal{C}(P,\pi)$ be a chamber, and x and x' two points in its relative interior, then P_x and $P_{x'}$ are* normally equivalent*, i.e. they have the same normal fan $\mathcal{N}(P_x) = \mathcal{N}(P_{x'})$. Similarly, $\mathcal{I}(W, h - Tx)$ and $\mathcal{I}(W, h - Tx')$ are equal.*

Thus, we define the collection of active constraints set $\mathcal{I}_\sigma$ and the normal fan $\mathcal{N}_\sigma$ above[6] *$\sigma \in \mathcal{C}(P,\pi)$ by:*

$$\mathcal{N}_\sigma := \mathcal{N}(P_x) \qquad \text{for an arbitrary } x \in \mathrm{ri}(\sigma) \tag{3.76a}$$

$$\mathcal{I}_\sigma := \mathcal{I}(W, h - Tx) \qquad \text{for an arbitrary } x \in \mathrm{ri}(\sigma) \tag{3.76b}$$

Figure 3.11: Two normally equivalent polytopes P and P' and their normal fan $\mathcal{N}(P) = \mathcal{N}(P')$.

To prove the normal equivalence, we need to extend the correspondences of Fig. 3.7 to a subcollection of faces of P.

Proposition 3.31 (Basic properties of collection of lifted faces)**.** *Let $P := \{(x,y) \in \mathbb{R}^{n+m} \mid Tx + Wy \leqslant h\}$ and $x \in \pi(P)$.*

3.31.(a) *The mapping*

$$\begin{cases} \mathcal{G}_x & \to \mathcal{I}(W, h - Tx) \\ G & \mapsto I_{(T,W),h}(G) \end{cases} \tag{3.77}$$

is a one-to-one correspondence whose inverse is $I \mapsto P(I)$. Moreover, these mappings are morphisms with respect to the intersection $\cap$.

3.31.(b) *The mapping*

$$\begin{cases} \mathcal{G}_x & \to \mathcal{F}(P)\backslash\{\emptyset\} \\ G & \mapsto G_x := \{y \in \mathbb{R}^m \mid (x,y) \in G\} \end{cases} \tag{3.78}$$

is a one-to-one correspondence whose inverse is the function which, for an input $F \in \mathcal{F}(P)\backslash\{\emptyset\}$, returns the minimal face $G \in \mathcal{F}(P)$ such that $\{x\} \times F \subset G$.

Moreover, these mappings are composition of the later mappings as described in Fig. 3.12 where for all $I \in \mathcal{I}((T,W),h)$, we recall the notation $P(I) := \{(x,y) \in P \mid T_I x + W_I y = h_I\}$. and similarly, for $x \in \pi(P)$ and $I \in \mathcal{I}(W, h - Tx)$, we denote $P_x(I) := \{y \in P_x \mid W_I y = h_I - T_I x\}$.

[6] The normal fan $\mathcal{N}_\sigma \subset 2^{\mathbb{R}^m}$ above σ should not be confused with $\mathcal{N}(\sigma) \subset 2^{\mathbb{R}^n}$ the normal fan of σ which we never consider.

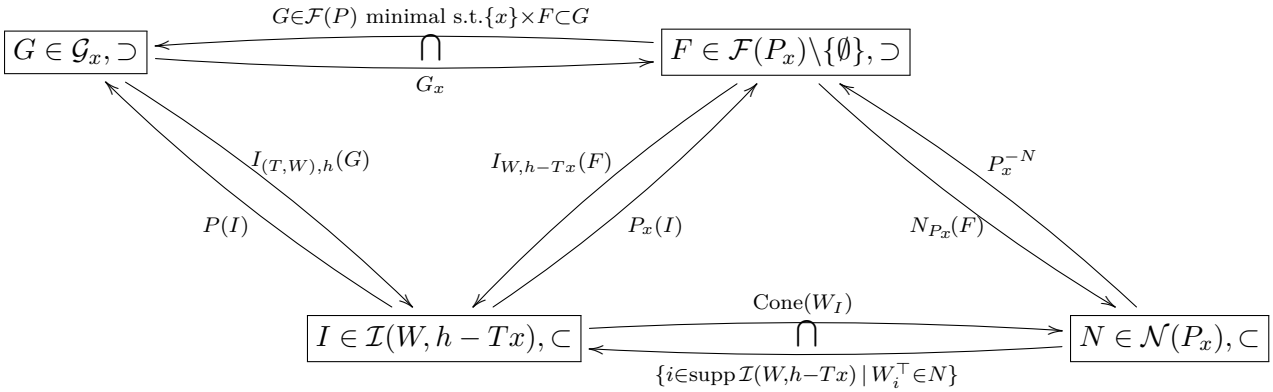


Figure 3.12: Monotonous one-to-one correspondences between normal fan, collection of active constraints sets and set of faces of a fiber P_x such that $x \in \mathrm{ri}(\sigma)$. For example, the downward arrow on the right reads $F_1 \subset F_2$ is equivalent to $N_{P_x}(F_1) \supset N_{P_x}(F_2)$.

Proof. 3.31.(a) Let $G \in \mathcal{G}_x$, we have that $x \in \pi\big(\mathrm{ri}(G)\big)$ and there exists $y \in \mathbb{R}^m$ such that $(x, y) \in \mathrm{ri}(G)$, then $I_{(T,W),h}(G) = I_{(T,W),h}(x, y) = I_{W,h-Tx}(y) \in \mathcal{I}(W, h - Tx)$. Then, the mapping is well-defined.

Moreover, on the one hand, we have by definition $\mathcal{G}_x \subset \mathcal{F}(P)\backslash\{\emptyset\}$. On the other hand, for $I \in \mathcal{I}(W, h - Tx)$, there exists $y \in \mathbb{R}^m$ such that $I_{W,h-Tx}(y) = I$. We then have $I = I_{(T,W),h}(x, y) \in \mathcal{I}\big((T, W), h\big)$ and thus $\mathcal{I}(W, h - Tx) \subset \mathcal{I}\big((T, W), h\big)$. Thus, the mappings in 3.31.(a) are restrictions of the mappings in 3.12.(b) with $A = (T, W)$ and $b = h$ which proves that they are inverse one-to-one correspondences. The morphism property is direct by noticing that y is in G_x and G'_x if and only if (x, y) is in G and G' which is equivalent to y being in $(G \cap G')_x$.

3.31.(b) We prove that these mappings are indeed the composition in Fig. 3.12. Let $G \in \mathcal{G}_x$, since $x \in \pi\big(\mathrm{ri}(G)\big)$, there exists $y \in \mathbb{R}^m$ such that $(x, y) \in \mathrm{ri}(G)$. In particular, $y \in G_x$. Since, by 3.8.(f), for all $G' \in \mathcal{F}(G)\backslash\{G\}$, $(x, y) \notin G'$ and then $y \notin G'_x$, we deduce that $y \in \mathrm{ri}(G_x)$. We then have $I_{(T,W),h}(G) = I_{(T,W),h}(x, y) = I_{W,h-Tx}(y) = I_{W,h-Tx}(G_x)$, which is exactly the composition of mappings in Fig. 3.12.

We now prove the other composition. Let $I \in \mathcal{I}(W, h - Tx)$. We show that $P(I)$ is the minimal face $G \in \mathcal{F}(P)$ such that $\{x\} \times P_x(I) \subset G$. First, for all $y \in P_x(I)$, $T_I x + W_I y = h_I$, thus, $\{x\} \times P_x(I) \subset P(I)$. Secondly, let $G \in \mathcal{F}(P)$ such that $\{x\} \times P_x(I) \subset G$ and $y \in \mathrm{ri}\,\big(P_x(I)\big)$. We have $I_{(T,W),h}(G) \subset \cap_{(x',y') \in G} I_{(T,W),h}(x', y') \subset I_{(T,W),h}(x', y') = I$. In particular, $I_{(T,W),h}(G) \subset I$ and by 3.12.(b) and 3.12.(e), we have $G \supset P(I)$. Thus, $P(I)$ is the minimal face $G \in \mathcal{F}(P)$ such that $\{x\} \times P_x(I) \subset G$. We have proved the composition which concludes the proof.

□

With these one-to-one correspondences, we are now able to prove the normal equivalence property.

Proof of Proposition 3.30. We define

$$\mathcal{G}_\sigma := \{G \in \mathcal{F}(P) \mid \mathrm{ri}\,(\sigma) \subset \mathrm{ri}\,\big(\pi(G)\big)\} = \mathcal{G}_x \qquad \text{for an arbitrary } x \in \mathrm{ri}(\sigma) \tag{3.79}$$

Let $x \in \mathrm{ri}(\sigma)$. For $G \in \mathcal{F}(P)$, $\mathrm{ri}(\sigma) \subset \mathrm{ri}\,\big(\pi(G)\big)$ implies $x \in \mathrm{ri}\,\big(\pi(G)\big)$. Then, $\mathcal{G}_\sigma \subset \mathcal{G}_x$. However, $\sigma = \cap_{G \in \mathcal{F}(P) \mid x \in \pi(G)} \pi(G)$. Then, if $G \in \mathcal{G}_x$, we have $x \in \mathrm{ri}\,\big(\pi(G)\big) \subset \pi(G)$ leading to $\sigma \subset \pi(G)$ and then $\mathrm{ri}(\sigma) \subset \mathrm{ri}\,\big(\pi(G)\big)$. We thus have $\mathcal{G}_x = \mathcal{G}_\sigma$.

To sum up, for all points x in the relative interior of a chamber $\mathrm{ri}(\sigma)$ have the same set of lifted faces $\mathcal{G}_x = \mathcal{G}_\sigma$.

By the correspondences of figure 3.12 and since $G \mapsto I_{(T,W),h}(G)$ does not depend on x. For all $x \in \mathrm{ri}(\sigma)$, $\mathcal{I}(W, h - Tx)$ is constant. Similarly, as $I \mapsto \mathrm{Cone}(W_I)$, $\mathcal{N}(P_x)$ does not depend on x, $\mathcal{N}(P_x)$ only depends on $\sigma \in \mathcal{C}(P, \pi)$ such that $x \in \mathrm{ri}(\sigma)$.

We rewrite figure 3.12 with these new notations in figure 3.13. □

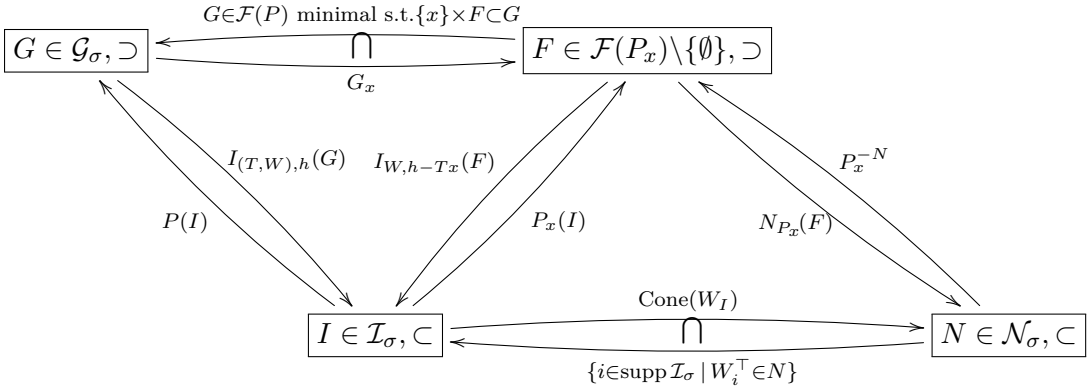


Figure 3.13: Monotonous one-to-one correspondences between normal fan, collection of active constraints sets and set of faces of a fiber P_x such that $x \in \mathrm{ri}(\sigma)$. For example, the downward arrow on the right reads $F_1 \subset F_2$ is equivalent to $N_{P_x}(F_1) \supset N_{P_x}(F_2)$.

We end with a monotonicity property concerning the cells of the chamber complex and its associated normal fans and collections of active constraint sets:

Proposition 3.32. *Let σ and τ two cells of the chamber complex $\mathcal{C}(P, \pi)$ such that*

$$\tau \subset \sigma. \tag{3.80}$$

Then, we have

$$\mathcal{N}_\sigma \preccurlyeq \mathcal{N}_\tau \tag{3.81a}$$
$$\mathcal{I}_\sigma \preccurlyeq \mathcal{I}_\tau \tag{3.81b}$$

Proof. We first prove that for every $G \in \mathcal{G}_\sigma$, there exists $G' \in \mathcal{G}_\tau$ such that $G \supset G'$. Let $G \in \mathcal{G}_\sigma$, by definition we have $\mathrm{ri}(\sigma) \subset \mathrm{ri}\,(\pi(G))$ and by continuity $\sigma \subset \pi(G)$. Let $x \in \mathrm{ri}(\tau)$, we have that $x \in \tau \subset \sigma \subset \pi(G)$. Then, there exists y such that $(x, y) \in G$. By Item 3.8.(f), there exists $G' \in \mathcal{F}(G)$ such that $(x, y) \in \mathrm{ri}(G')$. In particular, $x \in \pi\big(\mathrm{ri}(G')\big) = \mathrm{ri}\,\big(\pi(G')\big)$ which gives $G' \in \mathcal{G}_x = \mathcal{G}_\tau$. We thus have that there exists $G' \in \mathcal{G}_\tau$ such that $G \supset G'$.

Then, by the one-to-one monotonous correspondences of Fig. 3.13, we have that for all $I \in \mathcal{I}_\sigma$ there exists $I' \in \mathcal{I}_\tau$ such that $I \subset I'$ which is exactly $\mathcal{I}_\sigma \preccurlyeq \mathcal{I}_\tau$ and it is the same for $\mathcal{N}_\sigma \preccurlyeq \mathcal{N}_\tau$. □

3.5.2 H-representation of chamber and link with secondary fan

We now provide H-representations of chambers which allow us to interpret the chamber complex as a section of the secondary fan. For sake of simplicity, we assume in this section that W is a full rank matrix, i.e. $\mathrm{rg}(W) = m$. Under this assumption, see Assumption 3.1, every non-empty fiber $P_x = \{y \in \mathbb{R}^m \mid Tx + Wy \leqslant h\}$ have at least one vertex.

Proposition 3.33 (H-representation of the chamber)**.** *Assume that W is a full rank matrix, i.e. $\mathrm{rg}(W) = m$. For B a basis of $W^\top$, and $i \in [p]$, we denote by v_i^B the row vector*

$$v_i^B := T_i - W_i W_B^{-1} T_B \tag{3.82}$$

and by u_i^B the scalar

$$u_i^B := h_i - W_i W_B^{-1} h_B. \tag{3.83}$$

v_i^B and u_i^B can be understood as vectors and right-hand side of a reduced constraint.

Then, for every $I \in \mathcal{I}\big((T,W),h\big)$ from which we can extract a basis (i.e. $\operatorname{rg}(W_I^\top) = m$) and B_I such basis, we have

$$x \in \operatorname{ri}\big(\pi(P^I)\big) \iff \forall i \in I\backslash B_I,\ v_i^{B_I} x = u_i^{B_I} \text{ and } \forall j \in [\ell]\backslash I,\ v_j^{B_I} x < u_j^{B_I} \tag{3.84a}$$

$$\iff I \in \mathcal{I}(W, h - Tx) \tag{3.84b}$$

Moreover, for every $\sigma \in \mathcal{C}(P,\pi)$ we have the following H-representations:

$$\sigma = \left\{ x \in \mathbb{R}^n \;\middle|\; \forall I \in \overline{\mathcal{I}_\sigma},\ \begin{array}{ll} v_i^{B_I} x = u_i^{B_I}, & \forall i \in I\backslash B_I \\ v_j^{B_I} x \leqslant u_j^{B_I}, & \forall j \in [p]\backslash I \end{array} \right\} \tag{3.85a}$$

$$\operatorname{ri}(\sigma) = \left\{ x \in \mathbb{R}^n \;\middle|\; \forall I \in \overline{\mathcal{I}_\sigma},\ \begin{array}{ll} v_i^{B_I} x = u_i^{B_I}, & \forall i \in I\backslash B_I \\ v_j^{B_I} x < u_j^{B_I}, & \forall j \in [p]\backslash I \end{array} \right\} \tag{3.85b}$$

where for all $I \in \overline{\mathcal{I}_\sigma}$, B_I is an extracted basis of I.

Furthermore, the representations do not depend on the choice of these extracted bases B_I.

The set $\mathcal{I}_\sigma$ is thus a very useful tool for computation, as it provides a H-representation of the chamber σ and a V-representation of the normal fan $\mathcal{N}_\sigma$. For any $x \in \operatorname{ri}(P_x)$, it allows us to retrieve the H-representation of the faces of P_x and in particular its vertices. Finally, $\mathcal{I}_\sigma$ is a regular subdivision and then presents a rich and well-studied combinatorial structure.

Proof. (3.84a) We have $x \in \pi\big(\operatorname{ri} P(I)\big)$ if and only if there exists $y \in \mathbb{R}^m$ such that $(x,y) \in \operatorname{ri}\big(P(I)\big)$ which is equivalent to $T_I x + W_I y = h_I$ and for all $j \in [q]\backslash I$, $T_j x + W_j y < h_j$. We remark that the only $y \in \mathbb{R}^m$ which can satisfy these constraints is $W_{B_I}^{-1}(h_{B_I} - T_{B_I}x)$. Thus, x belongs to $\pi\big(\operatorname{ri} P(I)\big)$ if and only if $T_I x + W_I W_{B_I}^{-1}(h_{B_I} - T_{B_I}x) = h_{B_I}$ and for all $j \in [q]\backslash I$, $T_j x + W_j W_{B_I}^{-1}(h_{B_I} - T_{B_I}x) < h_j$. Since, for all $i \in B_i$, $v_i^{B_i} = 0$ and $u_i^{B_i} = 0$, the equation $v_i^{B_I} x = u_i^{B_I}$ is equivalent to the trivial equation $0 = 0$, we can only keep the indices in $I\backslash B_I$ for the equalities. The definitions of $v_i^{B_I}$ and $u_i^{B_I}$ then yield the equivalence.

(3.84b)
$$x \in \pi\big(\operatorname{ri}(P_I)\big) \iff \exists y \in \mathbb{R}^m,\ T_I x + W_I y = h_I \text{ and } T_{[\ell]\backslash I} x + W_{[\ell]\backslash I} y \ll h_{[\ell]\backslash I} \tag{3.86a}$$

$$\iff \exists y \in \mathbb{R}^m, I_{W,h-Tx}(y) = I \text{ and } Wy \leqslant h - Tx \tag{3.86b}$$

$$\iff I \in \mathcal{I}(W, h - Tx) \tag{3.86c}$$

In particular, $\pi(P^I)$ is contained in the vector subspace $E := \{x \in \mathbb{R}^n \mid v_i^{B_I} x = u_i^{B_I},\ \forall i \in I\backslash B\}$. Let $x \in \operatorname{ri}\big(\pi(P^I)\big) \neq \emptyset$, then since the inequality are strict, there exists $\varepsilon > 0$, such that $B(x,\varepsilon) \cap E \subset \operatorname{ri}\big(\pi(P^I)\big)$. Thus, $\dim(\pi(P^I)) = \dim(E)$.

Let $\sigma \in \mathcal{C}(P,\pi)$. Then, the collections $\underline{\mathcal{G}_\sigma}$, $\overline{\mathcal{I}_\sigma}$, $\overline{\mathcal{N}_\sigma}$ and $\operatorname{Vert}(P_x)$, for $x \in \operatorname{ri}(\sigma)$, are in one-to-one correspondences. By definition of $\underline{\mathcal{G}_\sigma}$, we have that $\operatorname{ri}(\sigma) = \cap_{G \in \underline{\mathcal{G}_\sigma}} \pi\big(\operatorname{ri}(G)\big)$, which is equal to $\cap_{I \in \overline{\mathcal{I}_\sigma}} \pi\big(\operatorname{ri} P(I)\big)$. Let $I \in \overline{\mathcal{I}_\sigma}$ be a maximal set of constraints. By the discussion in Section 3.2.3 and since $\operatorname{rg}(W) = m$, there exists a basis B_I of $W^\top$ such that $B_I \subset I$. By (3.84a), we have $x \in \pi\big(\operatorname{ri} P(I)\big)$ if and only if $\forall i \in I\backslash B_I$, $v_i^{B_I} x = u_i^{B_I}$ and $\forall j \in [\ell]\backslash I$, $v_j^{B_I} x < u_j^{B_I}$. Finally, since $\operatorname{ri}(\sigma) = \cap_{I \in \overline{\mathcal{I}_\sigma}} \pi\big(\operatorname{ri} P(I)\big)$, we then have

$$\operatorname{ri}(\sigma) = \left\{ x \in \mathbb{R}^n \;\middle|\; \forall I \in \overline{\mathcal{I}_\sigma},\ \begin{array}{ll} T_i x + W_i W_{B_I}^{-1}(h_{B_I} - T_{B_I}x) = h_i, & \forall i \in I\backslash B_I \\ T_j x + W_j W_{B_I}^{-1}(h_{B_I} - T_{B_I}x) < h_j, & \forall j \in [p]\backslash I \end{array} \right\} \tag{3.87}$$

We conclude using the definition of $v_i^{B_I}$ and $u_i^{B_I}$ and by taking the closure to get the representation of σ. □

We now show that the chamber complex can be interpreted as an affine section of the secondary fan.

Theorem 3.34. *Consider the affine function $a : x \mapsto h - Tx$. Then, the chamber complex $\mathcal{C}(P,\pi)$ is the inverse image, under a, of the secondary fan of $W^\top$:*

$$\mathcal{C}(P,\pi) = a^{-1}\Big(\Sigma\text{-fan}(W^\top)\Big) := \Big\{a^{-1}(S) \,\Big|\, S \in \Sigma\text{-fan}(W^\top)\Big\}\Big\backslash\Big\{\emptyset\Big\} \tag{3.88}$$

Proof. Assume for sake of simplicity that $\mathrm{rg}(W) = m$. The proof can be adapted by taking a quotient space if this assumption does not hold.

($\subset$) Let $\sigma \in \mathcal{C}(P,\pi)$, by (3.85), setting $b = h - Tx$, we have

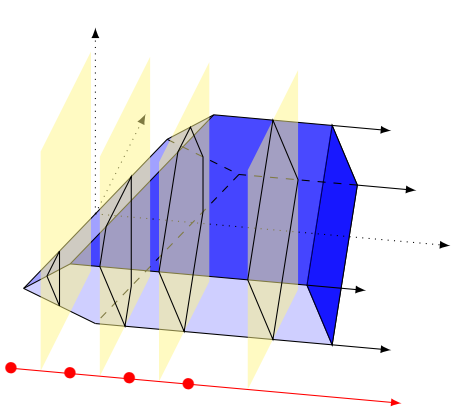
$$\mathrm{ri}(\sigma) = \left\{ x \in \mathbb{R}^n \;\middle|\; \exists b \in \mathbb{R}^p,\ b = h - Tx \text{ and } \forall I \in \overline{\mathcal{I}_\sigma},\ \begin{array}{ll} W_i W_{B_I}^{-1} b_{B_I} = b_i, & \forall i \in I \backslash B_I \\ W_j W_{B_I}^{-1} b_{B_I} < b_j, & \forall j \in [p] \backslash I \end{array} \right\}$$

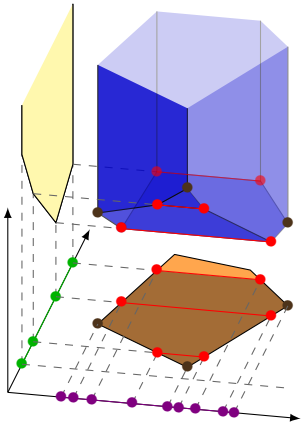
By Proposition 3.21, we then have $\mathrm{ri}(\sigma) = a^{-1}\big(\Sigma\text{-}C^\circ(W^\top, \mathcal{I}_\sigma)\big)$. By taking the closure and as a is continuous, we obtain $\sigma = a^{-1}\big(\Sigma\text{-}C(W^\top, \mathcal{I}_\sigma)\big)$. We then deduce that $\mathcal{C}(P,\pi) \subset a^{-1}\Big(\Sigma\text{-fan}(W^\top)\Big)$.

($\supset$) Reciprocally, we consider $S \in \Sigma\text{-fan}(W^\top)$ such that $a^{-1}(S)$ is not empty and we now prove that $a^{-1}(S) \in \mathcal{C}(P,\pi)$. As $S \in \Sigma\text{-fan}(W^\top)$, there exists a regular subdivision $\mathcal{I}$ such that $S = \Sigma\text{-}C(W^\top, \mathcal{I})$. If $a^{-1}(S)$ is not empty, then $\mathrm{ri}\,\big(a^{-1}(S)\big) = a^{-1}\big(\mathrm{ri}(S)\big) = a^{-1}\big(\Sigma\text{-}C^\circ(W^\top, \mathcal{I})\big)$ is not empty and there exists x such that $h - Tx \in \Sigma\text{-}C^\circ(W^\top, \mathcal{I})$ i.e. $\mathcal{I} = \mathcal{S}(W^\top, h - Tx) = \mathcal{I}(W, h - Tx)$. In particular, by Lemma 3.18, $h - Tx$ belongs to $\mathrm{Cone}(W) + \mathbb{R}^\ell_+$ and P_x is not empty, which is equivalent to $x \in \pi(P)$. Thus, there exists $\sigma \in \mathcal{C}(P,\pi)$ such that $x \in \mathrm{ri}(\sigma)$. We then have $\mathcal{I} = \mathcal{I}(W, h - Tx) = \mathcal{I}_\sigma$ and then by the previous discussion $a^{-1}(S) = a^{-1}\big(\Sigma\text{-}C(W^\top, \mathcal{I})\big) = a^{-1}\big(\Sigma\text{-}C(W^\top, \mathcal{I}_\sigma)\big) = \sigma \in \mathcal{C}(P,\pi)$ which concludes the proof.

□

4 Exact quantization of multistage stochastic linear problems





Contents

Most of this chapter is adapted from the preprint [FGL21], winner of the best paper student prize of ECSO-CMS 2022 conference in Venice.

⋆ refers to a supplementary section which was not in [FGL21].

4.1 Introduction

In this chapter, we show that every MSLP with general cost distribution is equivalent to an MSLP with finite distribution. In particular, we provide a uniform and universal exact quantization for MSLP with general cost. This leads to explicit representations of their value functions and to new complexity results. Most of this chapter is adapted from the preprint [FGL21], winner of the best paper student prize of ECSO-CMS 2022 conference in Venice. We added proofs for the normal equivalence property, for the characterization of the chamber complex as a section of the secondary fan and for the reduction of the Minkowksi integral to a Minkowski sum for the weighted fiber polyhedron. We also introduced a generalization of weighted fiber polyhedra to the multistage case by defining nested fiber polyhedra.

4.1.1 Multistage stochastic linear programming

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Given a sequence of independent random variables $\boldsymbol{c}_t \in L_1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_t})$ and $\boldsymbol{\xi}_t = (\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{b}_t)$, indexed by $t \in [T] := \{1, \dots, T\}$, we consider the MSLP given by

$$\begin{array}{lll} \min\limits_{(\boldsymbol{x}_t)_{t\in[T]}} & c_1^\top x_1 + \mathbb{E}\big[\sum\limits_{t=2}^{T} \boldsymbol{c}_t^\top \boldsymbol{x}_t\big] & \\ \text{s.t.} & A_1 x_1 \leqslant b_1 & \\ & \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{B}_t \boldsymbol{x}_{t-1} \leqslant \boldsymbol{b}_t \quad \text{a.s.} & \forall t \in \{2, \cdots, T\} \\ & \boldsymbol{x}_t \in L_\infty(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_t}) & \forall t \in \{2, \cdots, T\} \\ & \boldsymbol{x}_t \preccurlyeq \mathcal{F}_t & \forall t \in \{2, \cdots, T\} \end{array} \tag{4.1}$$

where $\boldsymbol{x}_1 \equiv x_1$, $\boldsymbol{A}_1 \equiv A_1$ and $\boldsymbol{b}_1 \equiv b_1$ are deterministic and $\mathcal{F}_t$ is the σ-algebra generated by $(\boldsymbol{c}_2, \boldsymbol{\xi}_2, \cdots, \boldsymbol{c}_t, \boldsymbol{\xi}_t)$. The last constraint, known as non-anticipativity, means that $\boldsymbol{x}_t$ is measurable with respect to $\mathcal{F}_t$.

Most results for MSLP with continuous distributions rely on discretizing the distributions. The Sample Average Approximation (SAA) method (see e.g. [SDR14, Chap. 5]) samples the costs and constraints. It relies on probabilistic results based on a uniform law of large number to give statistical guarantees. Obtaining a good approximation requires a large number of scenarios. In order to alleviate the computations, we can use scenario reduction techniques (see [DGKR03, HR03]). Latin Hypercube Sampling and variance reduction methods are also used to produce scenarios. Finally, one generates heuristically "good" scenarios, representing the underlying distribution (see [KW07]). Alternatively, we can leverage the structure of the problem to produce finite scenario trees (see [Kuh06, MAB14, MP18]) that yields bounds for the value of the true optimization problem. In each of these approaches, one solves an approximate version of the stochastic program, with or without statistical guarantee.

4.1.2 The exact quantization problem

Here, we aim at solving exactly the original problem, by finding an equivalent formulation with discrete distributions. This notion of equivalent formulation is best understood through the dynamic programming approach of MSLP. We define the *cost-to-go* function V_t inductively as follows. We set $V_{T+1} \equiv 0$ and for all $t \in \{2, \cdots, T\}$:

$$\begin{aligned} V_t(x_{t-1}) &:= \mathbb{E}\big[\hat{V}_t(x_{t-1}, \boldsymbol{c}_t, \boldsymbol{\xi}_t)\big] \\ \hat{V}_t(x_{t-1}, c_t, \xi_t) &:= \min_{x_t \in \mathbb{R}^{n_t}} \quad c_t^\top x_t + V_{t+1}(x_t) \\ &\qquad \text{s.t.} \quad A_t x_t + B_t x_{t-1} \leqslant b_t \end{aligned} \tag{4.2}$$

where $x_{t-1} \in \mathbb{R}^{n_{t-1}}$, $c_t \in \mathbb{R}^{n_t}$ and $\xi_t = (A_t, B_t, b_t) \in \mathbb{R}^{q_t \times n_t} \times \mathbb{R}^{q_t \times n_{t-1}} \times \mathbb{R}^{q_t}$.

We choose to distinguish the random cost $\boldsymbol{c}_t$ from the noise $\boldsymbol{\xi}_t$ affecting the constraints. Indeed, our results require $\boldsymbol{\xi}_t$ to be finitely supported (see examples 2.4 and 2.3) while $\boldsymbol{c}_t$ can have a continuous distribution. This separation does not preclude correlation between $\boldsymbol{c}_t$ and $\boldsymbol{\xi}_t$. However, we require $\{(\boldsymbol{c}_t, \boldsymbol{\xi}_t)\}_{t\in[T]}$ to be a sequence of independent random variables to leverage Dynamic Programming, even though some results can be extended to dependent $(\boldsymbol{\xi}_t)_{t\in[T]}$.

4.1.3 Contribution

We rely on a geometric approach, which enlightens the polyhedral structure of MSLP. We first establish exact quantization results in the two-stage case showing that there exists an optimal recourse affine on each cell of a polyhedral complex which is precisely the chamber complex [BS92, RZ96], a fundamental object in combinatorial geometry. A chamber complex is defined as the common refinement of the projections of faces of a polyhedron. In particular, Theorem 4.2 provides a local explicit exact quantization, in which the quantized probabilities and costs are attached to the cones of a polyhedral fan $\mathcal{N}$, and Theorem 4.3 gives a uniform exact quantization result (we refer the reader to [DLRS10, Zie12, Grü13, Fuk16] for background on polyhedral complexes and fans). On each cone $N \in \mathcal{N}$, we replace the distribution of $\boldsymbol{c}\mathbb{1}_{\operatorname{ri} N}$ by a Dirac distribution concentrated on the expected value $\check{c}_N = \mathbb{E}\big[\boldsymbol{c}|\boldsymbol{c} \in \operatorname{ri} N\big]$, and an associated weight $\check{p}_N = \mathbb{P}\big[\boldsymbol{c} \in \operatorname{ri} N\big]$. Further, $\mathcal{N}$ is *universal* in the sense that it does not depend on the distribution of $\boldsymbol{c}$.

In order to extend this result to the multistage case we establish in Lemma 4.12 a Dynamic Programming type equation in the space of polyhedral complexes. We then show an exact quantization result in Theorem 4.15. Again, this quantization is *universal* in the cost distribution.

We apply this polyhedral approach to obtain fixed parameters polynomial time complexity results considering both the exact computation problem and the approximation problem. For distributions that are uniform on polytopes or exponential, we show the MLSP can be solved in a time that is polynomial provided that the horizon T and the dimensions $n_2, \dots, n_T$ of the successive recourses are fixed. The proof relies on the theory of linear programming with oracles [GLS12] as well as on upper bound theorems of McMullen [McM70] and Stanley [Sta75] concerning the number of vertices and the size of a triangulation of a polyhedron. We obtain similar results for the approximation problem. Then the distribution cost can be essentially arbitrary: we only assume that it is given implicitly through an appropriate oracle. This applies in particular to distributions with a smooth density with respect to Lebesgue measure.

In summary, our main contributions are the following:

1. MSLP with arbitrary cost distribution and finitely supported constraints admit an exact quantization result, *i.e.* are equivalent to MSLP with discrete cost distribution;

2. the expected cost-to-go functions of such MSLP are polyhedral and affine on the cells of a universal polyhedral complex (*i.e.* independent of the cost distribution);

3. characterization of the expected cost-to-go function in terms of a weighted and nested extensions of the fiber polytope;

4. fixed-parameter polynomial time tractability results for 2SLP and MSLP.

4.1.4 Structure of the chapter

In Section 4.2 we establish the exact quantization result for 2-stage stochastic linear programming. In Section 4.3, we show that chamber complexes can be propagated through dynamic programming, leading to the exact quantization result for the MSLP. We give an analytical example in Section 4.4. In Section 4.5, we draw the consequences of our results in terms of computational complexity.

4.2 Exact quantization of the 2-stage problem

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\boldsymbol{q} \in L_1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^m)$ be an integrable random vector, and suppose $\xi = (T, W, h)$ is deterministic. We study the cost-to-go function of the 2-stage stochastic linear problem, written as

$$V(x) := \mathbb{E}\left[\hat{V}(x, \boldsymbol{q})\right] \quad \text{with} \quad \hat{V}(x,q) := \begin{array}{ll} \min\limits_{y \in \mathbb{R}^m} & q^\top y \\ \text{s.t.} & Tx + Wy \leqslant h \end{array} \tag{4.3}$$

The dual of the latter problem, for given x and c, is

$$\max_{\mu \in \mathbb{R}^q} \quad (Tx - h)^\top \mu \tag{4.4}$$

$$\text{s.t.} \quad W^\top \mu = -q \tag{4.5}$$

$$\mu \geqslant 0 \tag{4.6}$$

We denote the *coupling constraint polyhedron* of Problem (4.3) by

$$P := \{(x, y) \in \mathbb{R}^{n+m} \mid Tx + Wy \leqslant h\} \tag{4.7}$$

and π the projection of $\mathbb{R}^n \times \mathbb{R}^m$ onto $\mathbb{R}^n$ such that $\pi(x, y) = x$.

The projection of P is the following polyhedron :

$$\pi(P) = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m,\ Tx + Wy \leqslant h\} \tag{4.8}$$

and for any $x \in \mathbb{R}^n$, the *fiber*[1] of P along π is

$$P_x := \{y \in \mathbb{R}^m \mid Tx + Wy \leqslant h\} \tag{4.9}$$

4.2.1 Uniform exact quantization result

The following lemma provides an explicit formula for the cost-to-go function. It shows that an optimal recourse can be chosen as a function of c that is piecewise constant on the normal fan of P_x. This lemma can be interpreted as a reformulation of the basis decomposition Theorem 3.29 to the 2SLP notations.

Lemma 4.1. *Let $x \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$,*

1. *If $x \notin \pi(P)$, then $\hat{V}(x, q) = +\infty$;*
2. *If $x \in \pi(P)$ and $-q \notin \mathrm{Cone}(W^\top)$, then $\hat{V}(x, q) = -\infty$;*
3. *Suppose now that $x \in \pi(P)$ and $-q \in \mathrm{Cone}(W^\top)$. For each cone $N \in \mathcal{N}(P_x)$, let us select in an arbitrary manner a vector q_N in $\mathrm{ri}(-N)$. Then, there exists a vector $y_N(x)$ which achieves the minimum in the expression of $\hat{V}(x, q_N)$ in (4.3). Further, for any selection of such a $y_N(x)$, we have*

$$\hat{V}(x, q) = \sum_{N \in \mathcal{N}(P_x)} \mathbb{1}_{q \in -\mathrm{ri}\, N}\ q^\top y_N(x)\ . \tag{4.10}$$

[1] The term *parametrized polyhedron*, instead of fibers is also used in the literature [CL98, LW97].

Proof. The first point comes from the definitions of $\pi(P)$ in (4.8) and $\hat{V}(x,q)$ in (4.3). If $x \in \pi(P)$ and $-q \notin \mathrm{Cone}(W^\top)$, then the primal problem Eq. (4.3) is feasible and the dual problem is Eq. (4.14a) infeasible. Thus, by strong duality, $\hat{V}(x,q) = -\infty$.

By Eq. (3.32), we have that $\big(\mathrm{rc}(P_x)\big)^\circ = \mathrm{supp}\,\mathcal{N}(P_x)$. Further, by Proposition 3.7 all non empty fibers P_x have the same recession cone $\{y \in \mathbb{R}^m \mid Wy \leqslant 0\}$ whose polar is $\mathrm{Cone}(W^\top)$.

Assume now that $x \in \pi(P)$ and $-q \in \mathrm{Cone}(W^\top) = \mathrm{supp}(\mathcal{N}(P_x))$. Then, there exists $N \in \mathcal{N}(P_x)$ such that $-q \in \mathrm{ri}(N)$. Moreover, for every choice of $c_N \in -\mathrm{ri}(N)$, we have $\arg\min_{y\in P_x} q^\top y = \arg\min_{y\in P_x} q_N^\top y$, by 3.14.(k) . Moreover, there exists $y_N(x)$ such that $N = N_{P_x}\big(y_N(x)\big)$ by definition of a normal cone, thus $y_N(x) \in \arg\min_{y\in P_x} q_N^\top y$; in particular, the latter argmin is non empty. Thus, when $-q \in \mathrm{ri}(N)$, $\hat{V}(x,q) = q^\top y_N(x)$.

Thanks to the partition property of Proposition 3.9, we know that c belongs to the relative interior of precisely one cone in the normal fan of P_x, leading to (4.10). □

Having this property in mind, we make the following assumption:

Assumption 1. The cost $\boldsymbol{q} \in L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^m)$ is integrable with $\boldsymbol{q} \in -\mathrm{Cone}(W^\top)$ almost surely.

Theorem 4.2 (Local exact quantization of the cost distribution)**.** *Let $x \in \pi(P)$ and Assumption 1 holds. Then, for every refinement $\mathcal{R}_x$ of $-\mathcal{N}(P_x)$, we have:*

$$V(x) = \sum_{R\in\mathcal{R}_x} \check{p}_R \hat{V}(x, \check{q}_R) \quad \textit{with} \quad \hat{V}(x,\check{q}_R) := \min_{y\in\mathbb{R}^m} \quad \check{q}_R^\top y + \mathbb{I}_{Tx+Wy\leqslant h} \tag{4.11}$$

where $\check{p}_R := \mathbb{P}\big[\boldsymbol{q} \in \mathrm{ri}(R)\big]$ and $\check{q}_R := \mathbb{E}\big[\boldsymbol{q} \,|\, \boldsymbol{q} \in \mathrm{ri}(R)\big]$ if $\check{p}_R > 0$ and $\check{q}_R := 0$ if $\check{p}_R = 0$.

This is an exact quantization result, since (4.11) shows that $V(x)$ coincides with the value function of a second stage problem with a cost distribution supported by the finite set $\{\check{q}_R \mid R \in \mathcal{R}\}$.

Proof. For $R \in \mathcal{R}$, there exists one and only one $N \in -\mathcal{N}(P_x)$ such that $\mathrm{ri}(R) \subset \mathrm{ri}(N)$, that we denote $N(R)$. Indeed, as $\mathcal{R}$ is a refinement of $-\mathcal{N}(P_x)$, there exists at least one such element, and as $-\mathcal{N}(P_x)$ is a polyhedral complex it is unique.

By Lemma 4.1, under Assumption 1 and since $x \in \pi(P)$,

$$\begin{aligned}
V(x) &= \mathbb{E}\Big[\sum_{N\in\mathcal{N}(P_x)} \mathbb{1}_{\boldsymbol{q}\in -\mathrm{ri}\,N} \boldsymbol{q}^\top y_N(x)\Big] \\
&= \mathbb{E}\Big[\sum_{N\in-\mathcal{N}(P_x)} \sum_{R\in\mathcal{R}|\,\mathrm{ri}(R)\subset\mathrm{ri}(N)} \mathbb{1}_{\boldsymbol{q}\in\mathrm{ri}\,R}\, \boldsymbol{q}^\top y_N(x)\Big] && \text{by the partition property} \\
&= \sum_{R\in\mathcal{R}} \mathbb{E}\big[\mathbb{1}_{\boldsymbol{q}\in\mathrm{ri}\,R}\boldsymbol{q}^\top\big] y_{N(R)}(x) && \text{by linearity} \\
&= \sum_{R\in\mathcal{R}} \check{p}_R \check{q}_R^\top y_{N(R)}(x) \\
&= \sum_{R\in\mathcal{R}} \check{p}_R \min_{y\in\mathbb{R}^m} \check{q}_R^\top y + \mathbb{I}_{Tx+Wy\leqslant h}
\end{aligned}$$

the last equality is obtained by definition of $y_{N(R)}(x)$ as $\check{q}_R \in N(R)$, which leads to Eq. (4.11). □

The local exact quantization Theorem 4.2 together with the normal equivalence on the chamber complex Proposition 3.30 allows us to derive a uniform exact quantization result:

Theorem 4.3 (Uniform exact quantization of the cost distribution)**.** *Let $x \in \pi(P)$ and Assumption 1 holds. Then, for every refinement $\mathcal{R}$ of $-\bigwedge_{\sigma\in\overline{\mathcal{C}(P,\pi)}}\mathcal{N}_\sigma$, we have:*

$$V(x) = \sum_{R\in\mathcal{R}} \check{p}_R \hat{V}(x,\check{q}_R) \quad \textit{with} \quad \hat{V}(x,\check{q}_R) := \min_{y\in\mathbb{R}^m} \quad \check{q}_R^\top y + \mathbb{I}_{Tx+Wy\leqslant h} \tag{4.13}$$

where $\check{p}_R := \mathbb{P}[\boldsymbol{q} \in \mathrm{ri}(R)]$ and $\check{q}_R := \mathbb{E}[\boldsymbol{q} \,|\, \boldsymbol{q} \in \mathrm{ri}(R)]$ if $\check{p}_R > 0$ and $\check{q}_R := 0$ if $\check{p}_R = 0$.

Proof. Let R be a refinement of $-\bigwedge_{\sigma\in\mathcal{C}(P,\pi)}\mathcal{N}_\sigma$. Let $x \in \pi(P)$, there exists $\sigma_x \in \mathcal{C}(P,\pi)$ such that $x \in \mathrm{ri}(\sigma_x)$. We have that $\mathcal{R} \preccurlyeq \bigwedge_{\sigma\in\mathcal{C}(P,\pi)}\mathcal{N}_\sigma \preccurlyeq \mathcal{N}_{\sigma_x} = \mathcal{N}(P_x)$, which, by Theorem 4.2, leads to Eq. (4.13).

We now prove that $\bigwedge_{\sigma\in\mathcal{C}(P,\pi)}\mathcal{N}_\sigma = \bigwedge_{\sigma\in\overline{\mathcal{C}(P,\pi)}}\mathcal{N}_\sigma$. If τ is included in σ, by Proposition 3.32, we have that $\mathcal{N}_\sigma$ refines $\mathcal{N}_\tau$. Thus, $\bigwedge_{\sigma\in\mathcal{C}(P,\pi)}\mathcal{N}_\sigma = \bigwedge_{\sigma\in\overline{\mathcal{C}(P,\pi)}}\mathcal{N}_\sigma$. □

4.2.2 Weighted fiber polyhedron

In this subsection, we give a dual interpretation to the exact quantization theorem for 2-stage stochastic linear programming.

We first notice that for all $q \in -\mathrm{Cone}(W^\top)$ and $x \in \mathbb{R}^n$, we have strong duality for the second stage problem

$$\begin{aligned} \hat{V}(x,q) = \sup_{\lambda\in\mathbb{R}^q} \quad & (Tx-h)^\top\lambda && \text{(4.14a)}\\ \text{s.t.} \quad & W^\top\lambda = -q && \text{(4.14b)}\\ & \lambda \geqslant 0 && \text{(4.14c)} \end{aligned}$$

We denote by $D_q := \{\lambda \in \mathbb{R}^l_+ \,|\, W^\top\lambda + q = 0\}$ the admissible dual set for a fixed cost $q \in -\mathrm{Cone}(W)$. Then, for every $q \in -\mathrm{Cone}(W)$, we have

$$\hat{V}(x,q) = \sup_{\lambda\in D_q} (Tx-h)^\top\lambda \tag{4.15}$$

In [BS92], Billera and Sturmfels defined the *fiber polytope* as the normalized Minkowski integral $\frac{1}{\mathrm{Vol}(Q)}\int_Q P_x dx$ of bounded fibers P_x where x is uniformly distributed on a polytope Q. We now extend the notion of fiber polytope. First, we allow the fibers to be polyhedron with non trivial recession cones and lineality spaces. Secondly, we replace the uniform distribution on a polytope by a probability distribution on a polyhedron. We call this new polyhedron the *weighted fiber polyhedron*. To link this notion with stochastic programming, we give the definition with respect to the dual fibers D_q.

Definition 4.4 (Weighted fiber polyhedron)**.** *Let Assumption 1 holds. The* weighted fiber polyhedron *E of the bundle $(D_q)_{q\in\mathrm{supp}(\boldsymbol{q})}$ is the Minkowski integral of all the fiber at q when q varies according to its probability distribution:*

$$E := \mathbb{E}[D_{\boldsymbol{q}}] = \int D_q \mathbb{P}(dq) = \Big\{ \int \lambda(q)\mathbb{P}(dq) \;\Big|\; \lambda(q) \in D_q \quad \mathbb{P}\text{-}a.s.,\ \lambda \in L^1(\mathbb{P},\mathbb{R}^m,\mathbb{R}^\ell) \Big\} \tag{4.16}$$

The weighted fiber polyhedron synthesises the polyhedral structure of 2SLP with stochastic cost q. In particular, the expected cost-to-go function V is, up to a change of variable, equal to the support function of the weighted fiber polyhedron.

Theorem 4.5. *Let Assumption 1 holds. Then, the expected cost-to-go V defined in* (4.3) *is the composition of the support function of the weighted fiber polyhedron E defined in Definition 4.4 and the affine transformation $a : \begin{cases} \mathbb{R}^m \to \mathbb{R}^\ell \\ x \mapsto Tx - h \end{cases}$*

$$V(x) = \sigma_E \circ a(x) := \sup_{\lambda\in E} (Tx-h)^\top\lambda \tag{4.17}$$

In particular, the affine regions of V are exactly the maximal sets of the polyhedral complex $a^{-1}\big(\mathcal{N}(E)\big)$.

The proof consists in applying the interchangeability theorem (see [RW09, Thm 14.60]) to the dual formulation of the second stage problem.

Proof. Under Assumption 1, we have $q \in \operatorname{Cone}(W^\top)$ almost surely then

$$
\begin{aligned}
V(x) &= \mathbb{E}\big[\hat{V}(x, \boldsymbol{q})\big] && \text{(4.18a)}\\
&= \mathbb{E}\big[\sup_{\lambda \in D_{\boldsymbol{q}}} (Tx-h)^\top \lambda\big] && \text{(4.18b)}\\
&= \mathbb{E}\big[\sup_{\lambda \in \mathbb{R}^\ell} (Tx-h)^\top \lambda - \mathbb{I}_{\lambda \in D_{\boldsymbol{q}}}\big] && \text{by (4.15)} \\
& && \text{(4.18c)}\\
&= \int_{\mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^\ell} \Big((Tx-h)^\top \lambda - \mathbb{I}_{\lambda \in D_q}\Big)\mathbb{P}(dq) && \text{(4.18d)}\\
&= \sup_{\lambda(.) \in L^1(\mathbb{R}^n, \mathbb{R}^\ell)} \int_{\mathbb{R}^n} \Big((Tx-h)^\top \lambda(q) - \mathbb{I}_{\lambda(q) \in D_q}\Big)\mathbb{P}(dq) && \text{by [RW09, Thm 14.60]} \\
& && \text{(4.18e)}\\
&= \sup_{\lambda(.) \in L^1(\mathbb{R}^n, \mathbb{R}^\ell)} (Tx-h)^\top \int_{\mathbb{R}^n} \lambda(q)\mathbb{P}(dq) - \mathbb{I}_{\lambda(q) \in D_q \ \mathbb{P}\text{- a.s.}} && \text{(4.18f)}\\
&= \sup_{\lambda(.) \in L^1(\mathbb{R}^n, \mathbb{R}^\ell) \,\big|\, \lambda(q) \in D_q \ \mathbb{P}\text{- a.s.}} (Tx-h)^\top \int_{\mathbb{R}^n} \lambda(q)\mathbb{P}(dq) && \text{(4.18g)}\\
&= \sup_{\lambda \in E} (Tx-h)^\top \lambda && \text{by definition of } E \\
& && \text{(4.18h)}
\end{aligned}
$$

Indeed, we can apply [RW09, Thm 14.60] since the opposite of the function $(q, \lambda) \mapsto (Tx-h)^\top \lambda - \mathbb{I}_{\lambda \in D_q}$ is a normal integrand (see [RW09, Def 14.27]) and as $L^1(\mathbb{R}^n, \mathbb{R}^\ell)$ is a decomposable space (see [RW09, Def 14.59]) with the measure $\mathbb{P}$. □

We now show that the weighted fiber polyhedron is indeed a polyhedron. To do so, we replace the Minkowski integral by a Minkowski sum of well-chosen fibers, leveraging the normal equivalence on the cells of the chamber complex.

Let D be the dual coupling polyhedron

$$D := \{(\lambda, q) \in \mathbb{R}^\ell \times \mathbb{R}^m \,|\, W^\top \lambda + q = 0, \lambda \geqslant 0\} \tag{4.19}$$

and $\mathcal{C}(D, \pi_q^{\lambda,q}) \subset 2^{-\operatorname{Cone}(W^\top)} \subset 2^{\mathbb{R}^m}$ be the chamber complex after projecting on the space $\mathbb{R}^m$ where q belongs.

We now give an adaptation of [BS92, Theorem 1.5] for the weighted fiber polyhedron.

Theorem 4.6. *The fiber polyhedron can reduced to the finite Minkowski sum*

$$E = \sum_{\gamma \in \mathcal{C}(D, \pi_q^{\lambda,q})} \check{p}_\gamma D_{\check{q}_\gamma} \tag{4.20}$$

where $\check{p}_\gamma := \mathbb{P}[\boldsymbol{q} \in \operatorname{ri}(\gamma)]$ and $\check{q}_\gamma := \mathbb{E}[\boldsymbol{q} \,|\, \boldsymbol{q} \in \operatorname{ri}(\gamma)]$ is the centroid of the cell γ if $\check{p}_\gamma > 0$ and $\check{q}_\gamma$ is an arbitrary point in $\operatorname{ri}(\gamma)$ if $\check{p}_\gamma = 0$.

Proof. By Proposition 3.7, all D_q for $q \in -\operatorname{Cone}(W^\top)$ have the same recession cone C and lineality space L. We can assume, if we restrict all polyhedra to the orthogonal $L^\perp$ of the lineality space, that $\operatorname{Lin}(D_q) = \{0\}$ for all $q \in -\operatorname{Cone}(W^\top)$.

By noticing that $-\operatorname{Cone}(W^\top) = \sqcup_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})} \operatorname{ri}(\gamma)$, we have

$$E = \int_{-\operatorname{Cone}(W^\top)} D_q\mathbb{P}(dq) = \sum_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})} \int_{\operatorname{ri}(\gamma)} D_q\mathbb{P}(dq) \tag{4.21}$$

We now show that $\int_{\operatorname{ri}(\gamma)} D_q\mathbb{P}(dq) = \check{p}_\gamma D_{\check{q}_\gamma}$. If all lineality spaces are reduced to $\{0\}$, we have by Theorem 3.2 that $D_q = \operatorname{Conv}\big(\operatorname{Vert}(D_q)\big) + C$. Since the common recession cone C is convex, we have $\int_{-\operatorname{Cone}(W^\top)} C\mathbb{P}(dq) = C$. We thus only need to prove that $\operatorname{Vert}(\check{p}_\gamma D_{\check{q}_\gamma}) = \operatorname{Vert}\big(\int_{\operatorname{ri}(\gamma)} D_q\mathbb{P}(dq)\big)$. By [BS92, Proposition 1.2], we have that for every measurable $Q \subset \mathbb{R}^m$ and $\psi \in \mathbb{R}^m$, $\int_Q D_q^\psi\mathbb{P}(dq) = \big(\int_Q D_q\mathbb{P}(dq)\big)^\psi$. We then deduce

$$\Big(\int_{\operatorname{ri}\gamma} D_q\mathbb{P}(dq)\Big)^\psi = \int_{\operatorname{ri}\gamma} D_q^\psi\mathbb{P}(dq) = \mathbb{E}\big[D_{\boldsymbol{q}}^\psi \mathbb{1}_{\boldsymbol{q}\in\operatorname{ri}\gamma}\big] = \check{p}_\gamma\mathbb{E}\big[D_{\boldsymbol{q}}^\psi \,|\, \boldsymbol{q}\in\operatorname{ri}\gamma\big]$$

.

For $\gamma \in \mathcal{C}(D,\pi_q^{\lambda,q})$ and every q and q' in $\operatorname{ri}(\gamma)$, D_q and $D_{q'}$ are normally equivalent, i.e. $\mathcal{N}(D_q) = \mathcal{N}(D_{q'})$. In particular, we have by Fig. 3.9 that for $\psi \in \operatorname{ri}(N)$ and $N \in \overline{\mathcal{N}(D_q)}$ each vertex $D_q^{-\psi}$ of D_q depends affinely on q and then get $\mathbb{E}\big[D_{\boldsymbol{q}}^\psi \,|\, \boldsymbol{q}\in\operatorname{ri}\gamma\big] = D_{\check{q}_\gamma}^\psi$. We then have $\operatorname{Vert}(\check{p}_\gamma D_{\check{q}_\gamma}) = \operatorname{Vert}\big(\int_{\operatorname{ri}(\gamma)} D_q\mathbb{P}(dq)\big)$ which implies $\int_{\operatorname{ri}(\gamma)} D_q\mathbb{P}(dq) = \check{p}_\gamma D_{\check{q}_\gamma}$ and thus Eq. (4.20). □

We now show that the normal fan of the weighted fiber polyhedron is refined by the secondary fan, up to a change of sign.

Theorem 4.7. *The normal fan of the weighted fiber polyhedron E is such that*

$$-\,\Sigma\text{-fan}(W^\top) \preccurlyeq \mathcal{N}(E) \tag{4.22}$$

Since for two polyhedra Q and Q', we have $\mathcal{N}(Q+Q') = \mathcal{N}(Q)\wedge\mathcal{N}(Q')$, see e.g. [Zie12, Prop. 7.12]. Then if a polyhedron Q have a coarser normal fan than Q, i.e. $\mathcal{N}(P) \preccurlyeq \mathcal{N}(Q)$, it means that Q has a more simple combinatorial structure than P and that there exists a polyhedron Q' such that $P = Q + Q'$. We say that Q is a Minkowski summand of P. Moreover, the secondary fan is the normal fan of a well-studied polytope called secondary polytope introduced in [GKZ94]. In [BS92], Billera and Sturmfels defined the fiber polytope and showed that the secondary polytope was a particular case of fiber polytope.

Remark 4.8. *We saw in Theorem 4.5 that the affine regions of V were equal to $a^{-1}(\mathcal{N}(E))$. By Theorem 4.7, we then have that the affine regions are refined by $a^{-1}(-\Sigma\text{-fan}(W^\top))$. Finally, in Theorem 3.34, we saw that the chamber complex was a section of the secondary $\mathcal{C}(P,\pi) = a^{-1}(-\Sigma\text{-fan}(W^\top))$. Thus, we will see in Theorem 4.9 that V is affine on every cell of $\mathcal{C}(P,\pi)$. One can see that the more chamber $\gamma \in \mathcal{C}(D,\pi_q^{\lambda,q})$ are such that $\check{p}_\gamma = 0$, the coarsest will be $\mathcal{N}(E)$ and thus the affine regions $a^{-1}(\mathcal{N}(E))$.*

Proof. We first prove that $-\bigwedge_{q\in-\operatorname{Cone}(W^\top)} \mathcal{N}(D_q) = \Sigma\text{-fan}(W^\top)$. We have that

$$\operatorname{supp}\Big(\bigwedge_{q\in-\operatorname{Cone}(W^\top)} -\mathcal{N}(D_q)\Big) = \operatorname{Cone}(W) + \mathbb{R}_+^\ell = \operatorname{supp}\Big(\Sigma\text{-fan}(W^\top)\Big)$$

. Since these collections of sets are both fans, by Lemma 3.16, it is enough to show an inclusion for the collection of maximal sets: $\overline{\Sigma\text{-fan}(W^\top)} \subset -\bigwedge_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})} \overline{\mathcal{N}(D_{q_\gamma})}$.

Let $S \in \overline{\Sigma\text{-fan}(W^\top)}$ by Definition 3.20, we know that there exists a regular triangulation $\mathcal{I}$ such that $S = \Sigma\text{-}C(W^\top,\mathcal{I})$. Thus, by Proposition 3.21, we have

$$\Sigma\text{-}C(W^\top,\mathcal{I}) = \bigcap_{B\in\overline{\mathcal{I}}} C_B \tag{4.23a}$$

$$\text{where } C_B := \{\psi \in \operatorname{Cone}(W) + \mathbb{R}_+^\ell \mid WW_B^{-1}\psi_B \leqslant \psi\} \tag{4.23b}$$

Assume $q \in -\operatorname{Cone}(W_B)$ and denote $\lambda_{B,q}$ the basic point, of the polyhedron in standard form D_q associated to the basis B_I, see Definition 3.24. As $q \in -\operatorname{Cone}(W_B)$, we have $\lambda_{B,q} \in D_q$ i.e. B is an admissible basis for the linear problem $LP^{=}_{W^\top,-q,\psi}$. We now prove that $C_B = -N_{D_q}(\lambda_{B,q})$. If $\psi \in C_B$, we have that the reduced cost $\psi - WW_B^{-1}\psi_B$ is nonnegative. In particular, by Lemma 3.25, B_I is an optimal basis for the problem $LP^{=}_{W^\top,\psi,-q}$ and $\lambda_{B,q}$ is an optimal point. Thus, for all $\lambda \in D_q$, $\psi^\top \lambda_{B,q} \leqslant \psi^\top \lambda$ which is equivalent to $\psi \in -N_{D_q}(\lambda_{B,q})$. Reciprocally, if $\psi \in -N_{D_q}(\lambda_{B,q})$ then the basic point $\lambda_{B,q}$ is an optimal point of $LP^{=}_{W^\top,-q,\psi}$ and the reduced cost $\psi - WW_B^{-1}\psi_B$ is non-negative and thus $\psi \in C_B$.

Since $\mathcal{I}$ is a triangulation of $\operatorname{Cone}(W^\top)$, we have $\bigcup_{B\in\overline{\mathcal{I}}} \operatorname{Cone}(W_B)$. In particular, for every $q \in -\operatorname{Cone}(W^\top)$, there exists $I \in \overline{\mathcal{I}}$ such that $q \in -\operatorname{Cone}(W_B)$. Then, for every $q \in -\operatorname{Cone}(W)$, we associate a point function $\lambda_{\mathcal{I},q} = \lambda_{B,q}$ where we choose $B \in \overline{\mathcal{I}}$ such that $q \in -\operatorname{Cone}(W_B)$.

Finally, we have

$$S = \bigcap_{B\in\overline{\mathcal{I}}} C_B = \bigcap_{B\in\overline{\mathcal{I}}} \bigcap_{q\in-\operatorname{Cone}(W_B)} -N_{D_q}(\lambda_{B,q}) = \bigcap_{q\in-\operatorname{Cone}(W^\top)} -N_{D_q}(\lambda_{\mathcal{I},q}) \tag{4.24}$$

Thus, $S \in -\bigwedge_{q\in-\operatorname{Cone}(W^\top)} \overline{\mathcal{N}(D_q)}$.

We proved that $-\bigwedge_{q\in-\operatorname{Cone}(W^\top)} \mathcal{N}(D_q) = \Sigma\text{-fan}(W^\top)$. Moreover, by the normal equivalence Proposition 3.30 for every $\gamma \in \mathcal{C}(D,\pi_q^{\lambda,q})$ and $q \in \operatorname{ri}(\gamma)$, $\mathcal{N}(D_q) = \mathcal{N}(D_{\check{q}_\gamma})$. Then, $-\bigwedge_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})} \mathcal{N}(D_{\check{q}_\gamma}) = \Sigma\text{-fan}(W^\top)$.

Recall that the normal fan of the Minkowski sum is the common refinement of the normal fan $\mathcal{N}(P+Q) = \mathcal{N}(P) \wedge \mathcal{N}(Q)$, see e.g. [Zie12, Prop. 7.12], and that $\mathcal{N}(\alpha P) = \mathcal{N}(P)$ for every $\alpha > 0$. Then, since $E = \sum_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})\,|\,\check{p}_\gamma>0} \mathcal{N}(D_{\check{q}_\gamma})$ Thus, by Eq. (4.20) and $-\Sigma\text{-fan}(W^\top) = \bigwedge_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})} \mathcal{N}(D_{q_\gamma}) \preccurlyeq \bigwedge_{\gamma\in\mathcal{C}(D,\pi_q^{\lambda,q})\,|\,\check{p}_\gamma>0} \mathcal{N}(D_{\check{q}_\gamma}) = \mathcal{N}(E)$ □

4.2.3 Explicit characterization of expected cost-to-go

As a consequence of the exact quantization Theorem 4.2, we obtain explicit representations for the values and subdifferentials of the expected cost-to-go function V. We also show that V is affine on every cell of the chamber complex for every distribution of the random cost.

Theorem 4.9 (Characterization of the expected cost-to-go function)**.** *Let Assumption 1 holds. For $x \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$, we denote*

$$D_c^{h-Tx} := \operatorname{argmax}\left\{(Tx-h)^\top \lambda : A^\top\lambda = -c, \lambda \geqslant 0\right\}$$

the set of optimal dual solutions of the second stage problem. Then,

$$\forall \sigma \in \mathcal{C}(P,\pi), \quad \forall x, x' \in \operatorname{ri}(\sigma), \quad \forall c \in \operatorname{supp}(\boldsymbol{c}), \qquad D_c^\sigma := D_c^{h-Tx} = D_c^{h-Tx'}.$$

Set

$$\alpha_\sigma := \sum_{N\in-\mathcal{N}_\sigma} B^\top \lambda^\sigma_{\check{c}_N} \quad \text{and} \quad \beta_\sigma := \sum_{N\in-\mathcal{N}_\sigma} -b^\top \lambda^\sigma_{\check{c}_N} \tag{4.25}$$

where λ_c^σ is an element of D_c^σ. Then, we have

$$\forall \sigma \in \mathcal{C}(P,\pi), \quad \forall x \in \sigma, \quad V(x) = \alpha_\sigma^\top x + \beta_\sigma \tag{4.26a}$$

$$\forall x \in \mathbb{R}^n, \quad V(x) = \mathbb{I}_{x\in\pi(P)} + \max_{\sigma\in\overline{\mathcal{C}(P,\pi)}} \alpha_\sigma^\top x + \beta_\sigma \tag{4.26b}$$

In particular, for all distributions of $\boldsymbol{c}$ satisfying Assumption 1, V is affine on each cell of $\mathcal{C}(P,\pi)$, i.e. the cells of the chamber complex are universal *affine regions.*

Moreover, we characterize the subdifferential of the cost-to-go function as

$$\partial V(x) = N_{\pi(P)}(x) + \operatorname{Conv}\left\{(\alpha_\sigma)_{\sigma\in\overline{\mathcal{C}(P,\pi)}\,|\,x\in\sigma}\right\}. \tag{4.27}$$

Proof. By the basis decomposition theorem Theorem 3.27, we have that $D_c^{\psi} = D_c^{\psi'}$ for all ψ and ψ' belonging to the same relative interior of a cone of the secondary fan Σ-fan$(W^\top)$. In particular, by Theorem 3.34, for every x, x' in the same relative interior of a chamber σ, we have $D_c^{h-Tx} = D_c^{h-Tx'}$.

For all $x \in \mathrm{ri}(\sigma) \subset \pi(P)$ and all $c \in \mathrm{supp}(\boldsymbol{c})$, by Lemma 4.1, we have $\hat{V}(x,c) < +\infty$ and then by strong duality, $\hat{V}(x,c) = (Tx-h)^\top \lambda_\sigma^c$. Then by the exact quantization result Eq. (4.11), for all $x \in \mathrm{ri}(\sigma)$,

$$V(x) = \sum_{N \in -\mathcal{N}_\sigma} \check{p}_N \hat{V}(x, \check{c}_N) = \sum_{N \in -\mathcal{N}_\sigma} \check{p}_N (Tx-h)^\top \lambda_\sigma^{\check{c}_N} = \alpha_\sigma^\top x + \beta_\sigma.$$

Further, as V is lower semicontinuous and convex, we deduce *Eq.* (4.26a).

To show Eq. (4.26b), suppose first that $\dim\big(\pi(P)\big) = m$. Then, for $\sigma \in \overline{\mathcal{C}(P,\pi)}$, $x \to \alpha_\sigma^\top x + \beta_\sigma$ is a supporting affine function of V which coincide with V on σ whose dimension is m. Since $\bigcup_{\sigma \in \overline{\mathcal{C}(P,\pi)}} \sigma = \mathrm{supp}(\mathcal{C}(P,\pi)) = \pi(P)$, V is piecewise affine on the polyhedron $\pi(P)$ and equals to $+\infty$ elsewhere. Together with convexity of V, this yields Eq. (4.26b). When $\pi(P)$ is not full dimensional, we get the same result by restraining the ambient space to the affine hull $\mathrm{Aff}\,\big(\pi(P)\big)$. Since $\mathcal{C}(P,\pi)$ does not depend on $\boldsymbol{c}$, for all distributions of $\boldsymbol{c}$ satisfying *Assumption* 1, V is affine on each cell of $\mathcal{C}(P,\pi)$. Finally, the subgradient formula follows from Eq. (4.26). □

Remark 4.10. *Let $\mathcal{V}^{\max}$ be the collection of affine regions of V. Theorem 4.9 implies that the chamber complex $\overline{\mathcal{C}(P,\pi)}$ refines $\mathcal{V}^{\max}$. However, it does not imply that $\overline{\mathcal{C}(P,\pi)} = \mathcal{V}^{\max}$. Indeed, if $\boldsymbol{c} = 0$ $\mathbb{P}$-almost surely, then $\mathcal{V}^{\max} = \{\pi(P)\}$.*

More precisely, for all cost distribution such that Assumption 1 holds, $\mathcal{V}^{\max}$ is the collection of maximal elements of a polyhedral complex $\mathcal{V}$ such that $\mathcal{C}(P,\pi) \preccurlyeq \mathcal{V}$. We gave an exact representation of $\mathcal{V}$ in Theorem 4.5, showing that $\mathcal{V} = a^{-1}\Big(\mathcal{N}(E)\Big)$.

4.3 Exact quantization of the multistage problem

In this section, we show that the exact quantization result established above for a general cost distribution and deterministic constraints carries over to the case of stochastic constraints with finite support and then to multistage programming.

We denote by $\pi_x^{x,y}$ for the projection from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^n$ defined by $\pi_x^{x,y}(x',y') = x'$. The projections $\pi_{x,y}^{x,y,z}$, $\pi_x^{x,y,z}$, $\pi_y^{y,z}$, $\pi_{x_{t-1}}^{x_{t-1},z}$ are defined accordingly. Note that in the notation $\pi_x^{x,y,z}$, x, y and z are part of the notation and not parameters.

4.3.1 Propagating chamber complexes through Dynamic Programming

We next show that chamber complexes are propagated through dynamic programming in a way that is *uniform* with respect to the cost distribution. This is a key tool to extend the exact quantization theorem to the multistage setting. Note that the proof of Theorem 4.3 cannot be extended to the multistage setting as, in this case, the extensive form requires non-anticipativity constraints that cannot be tackled directly.

We start with a monotonicity result for chamber complexes.

Lemma 4.11 (Chamber complex monotonicity with respect to refinement order)**.** *Consider two polyhedral complexes of $\mathbb{R}^d$ and a projection π. If $\mathcal{R} \preccurlyeq \mathcal{S}$ then $\mathcal{C}(\mathcal{R},\pi) \preccurlyeq \mathcal{C}(\mathcal{S},\pi)$.*

Proof. For any $R \in \mathcal{R}$, there exist $S_R \in \mathcal{S}$ such that $R \subset S_R$. Let $x \in \mathrm{supp}\,\mathcal{C}(\mathcal{R},\pi) = \pi(\mathrm{supp}\,\mathcal{R}) = \pi(\mathrm{supp}\,\mathcal{S}) = \mathrm{supp}\,\mathcal{C}(\mathcal{S},\pi)$

$$\begin{aligned}\sigma_{\mathcal{R},\pi}(x) := \bigcap_{R \in \mathcal{R}\,\text{s.t. }\, x \in \pi(R)} \pi(R) &\subset \bigcap_{R \in \mathcal{R}\,\text{s.t. }\, x \in \pi(R)} \pi(S_R) \\ &\subset \bigcap_{S \in \mathcal{S}\text{s.t. }\, x \in \pi(S)} \pi(S) =: \sigma_{\mathcal{S},\pi}(x) \in \mathcal{C}(\mathcal{S},\pi)\end{aligned}$$

□

Recall that, for a polyhedron P and a vector ψ, we denote $P^\psi := \arg\min_{x\in P} \psi^\top x$. Let f be a polyhedral function on $\mathbb{R}^d$, with a slight abuse of notation we denote $\operatorname{epi}(f)^{\psi,1} = \arg\min_{(x,z)\in\operatorname{epi}(f)} \psi^\top x + z$. We denote $\mathcal{F}_{\text{low}}(\operatorname{epi}(f)) := \{\operatorname{epi}(f)^{\psi,1} \mid \psi \in \mathbb{R}^d\}$ the set of *lower faces* of $\operatorname{epi}(f)$. The collection of projections (on $\mathbb{R}^d$) of lower faces of $\operatorname{epi}(f)$ is the coarsest polyhedral complex such that f is affine on each of its cells (see [DLRS10, Chapter 2]). Moreover, we have

$$\pi_{\mathbb{R}^d}(\operatorname{epi}(f)^{\psi,1}) = \arg\min_{x\in\mathbb{R}^d} \psi^\top x + f(x) \tag{4.28}$$

Lemma 4.12. *Let U be a polyhedral function on $\mathbb{R}^m$ and $\mathcal{U} := \pi_y^{y,z}\big(\mathcal{F}_{\text{low}}(\operatorname{epi}(U))\big)$ a coarsest polyhedral complex such that U is affine on each element of $\mathcal{U}$. Let $\xi = (A, B, b)$ be fixed and Assumption 1 holds. Define, for all $x \in \mathbb{R}^n$*

$$Q(x, y) := U(y) + \mathbb{I}_{Ay+Bx\leqslant b} \tag{4.29a}$$
$$V(x) := \mathbb{E}\big[\min_{y\in\mathbb{R}^m} \boldsymbol{c}^\top y + Q(x, y)\big] \tag{4.29b}$$

Let $\mathcal{V} := \mathcal{C}\Big(\mathcal{F}(P) \wedge (\mathbb{R}^n \times \mathcal{U}), \pi_x^{x,y}\Big) \subset 2^{\mathbb{R}^n}$ with $P := \{(x, y) \mid Ay + Bx \leqslant b\}$.

Then, $\mathcal{V} \preccurlyeq \mathcal{C}\Big(\operatorname{epi}(Q), \pi_x^{x,y,z}\Big)$ and V is a polyhedral function which is affine on each element of $\mathcal{V}$.

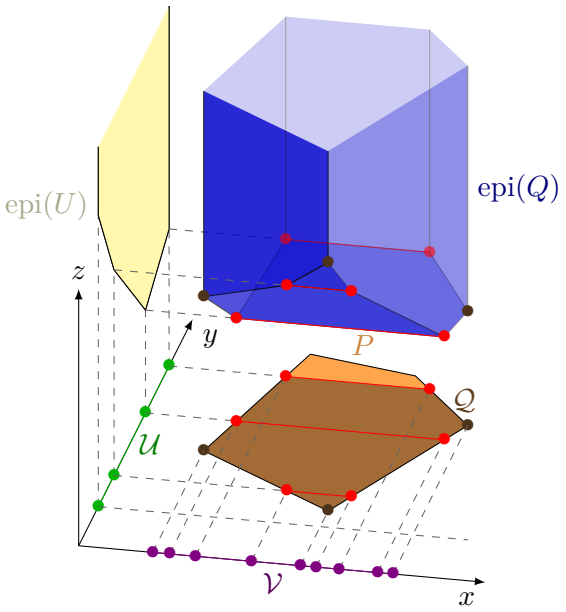


Figure 4.2: An illustration of the proof of Lemma 4.12: the epigraph $\operatorname{epi}(Q)$ of the coupling function in blue in the (x, y, z) space, the epigraph of U in yellow in the (y, z) plane, the affine regions $\mathcal{U}$ of U in green on the y axis, the coupling polyhedron P in orange and brown in the (x, y) plane, the polyhedral complex $\mathcal{Q}$ in red and brown in the (x, y) plane and the chamber complex $\mathcal{V}$ in violet on the x axis.

Proof. We have $\operatorname{epi}(Q) = (\mathbb{R}^n \times \operatorname{epi}(U)) \cap (P \times \mathbb{R}) \subset \mathbb{R}^{n+m+1}$. Since

$$V(x) = \mathbb{E}\big[\min_{y\in\mathbb{R}^m, z\in\mathbb{R}} \boldsymbol{c}^\top y + z + \mathbb{I}_{(x,y,z)\in\operatorname{epi}(Q)}\big],$$

by Theorem 4.9 applied to the problem with variables (y, z) and the coupling polyhedron $\mathrm{epi}(Q)$, V is a polyhedral function affine on each element of $\mathcal{C}(\mathrm{epi}(Q), \pi_x^{x,y,z})$. We now show that $\mathcal{V} \preccurlyeq \mathcal{C}(\mathrm{epi}(Q), \pi_x^{x,y,z})$. As $\mathrm{epi}(Q)$ is the epigraph of a polyhedral function, $\mathcal{Q} := \pi_{x,y}^{x,y,z}(\mathcal{F}_{\mathrm{low}}(\mathrm{epi}(Q))) \subset 2^{\mathbb{R}^{n+m}}$ is a polyhedral complex.

Let $\check{x} \in \pi_x^{x,y,z}(\mathrm{epi}(Q))$, using notation of Definition 3.10,

$$\sigma_{\mathrm{epi}(Q),\pi_x^{x,y,z}}(\check{x}) := \bigcap_{F \in \mathcal{F}(\mathrm{epi}(Q)) \text{ s.t. } \check{x} \in \pi_x^{x,y,z}(F)} \pi_x^{x,y,z}(F) \tag{4.30}$$

$$= \bigcap_{F \in \mathcal{F}_{\mathrm{low}}(\mathrm{epi}(Q)) \text{ s.t. } \check{x} \in \pi_x^{x,y,z}(F)} \pi_x^{x,y,z}(F) \tag{4.31}$$

$$= \bigcap_{F' \in \mathcal{Q} \text{ s.t. } \check{x} \in \pi_x^{x,y}(F')} \pi_x^{x,y}(F') =: \sigma_{\mathcal{Q},\pi_x^{x,y}}(\check{x}) \tag{4.32}$$

Indeed, as $\mathrm{epi}(Q)$ is an epigraph of a polyhedral function, if $F \in \mathcal{F}(\mathrm{epi}(Q))$ such that $\check{x} \in \pi_x^{x,y,z}(F)$ then there exists $G \in \mathcal{F}_{\mathrm{low}}(\mathrm{epi}(Q))$ such that $G \subset F$ and $\check{x} \in \pi_x^{x,y,z}(G)$, allowing us to go from the first to second equality. The third equality is obtained by setting $F' = \pi_{x,y}^{x,y,z}(F)$. Thus, $\mathcal{C}(\mathrm{epi}(Q), \pi_x^{x,y,z}) = \mathcal{C}(\mathcal{Q}, \pi_x^{x,y})$.

We now show that $\mathcal{F}(P) \wedge (\mathbb{R}^n \times \mathcal{U}) \preccurlyeq \mathcal{Q}$. Let $G \in \mathcal{F}(P) \wedge (\mathbb{R}^n \times \mathcal{U})$. There exist $\sigma \in \mathcal{U}$ and $F \in \mathcal{F}(P)$ such that $G = F \cap (\mathbb{R}^n \times \sigma)$. By definition of $\mathcal{F}_{\mathrm{low}}$, there exists $\psi \in \mathbb{R}^m$ such that $\sigma = \pi_y^{y,z}(\mathrm{epi}(U)^{\psi,1})$. We show that $G \subset \pi_{x,y}^{x,y,z}(\mathrm{epi}(Q)^{0,\psi,1}) \in \mathcal{Q}$. Indeed, let $(x, y) \in G = F \cap (\mathbb{R}^n \times \pi_y^{y,z}(\mathrm{epi}(U)^{\psi,1}))$. We have $(x, y) \in F \subset P$ such that $y \in \arg\min_{y' \in \mathbb{R}^m} \{\psi^\top y' + U(y')\}$. Which implies that $(x, y) \in \arg\min \{\psi^\top y' + U(y') \mid (x', y') \in P\}$. This also reads, by Eq. (4.28), as $(x, y) \in \pi_{x,y}^{x,y,z}(\mathrm{epi}(Q)^{0,\psi,1})$. Thus, $G \subset \pi_{x,y}^{x,y,z}(\mathrm{epi}(Q)^{0,\psi,1}) \in \mathcal{Q}$ leading to $\mathcal{F}(P) \wedge (\mathbb{R}^n \times \mathcal{U}) \preccurlyeq \mathcal{Q}$. Finally, by monotonicity, Lemma 4.11 ends the proof. □

Remark 4.13. *In Lemma 4.12, the complex $\mathcal{V}$ is independent of the distribution of $\boldsymbol{c}$. However, for special choices of $\boldsymbol{c}$, V might be affine on each cell of a coarser complex than $\mathcal{V}$. For instance, if $U = 0$ and $\boldsymbol{c} \equiv 0$, we have that $V = \mathbb{I}_{\pi_x^{x,y}(P)}$, V is affine on $\pi_x^{x,y}(P)$. Nevertheless, $\mathcal{V} = \mathcal{C}(P, \pi_x^{x,y})$ is generally finer than $\mathcal{F}(\pi_x^{x,y}(P))$.*

4.3.2 Exact quantization of MSLP

We next show that the multistage program with arbitrary cost distribution is equivalent to a multistage program with independent, finitely distributed, cost distributions. Further, for all step t, there exist affine regions, independent of the distributions of costs, where V_t is affine. Assumption 1 is naturally extended to the multistage setting as follows

Assumption 4.1. The sequence $(\boldsymbol{c}_t, \boldsymbol{\xi}_t)_{2 \leqslant t \leqslant T}$ is independent.[2] Further, for each $t \in \{2, \cdots, T\}$, $\boldsymbol{\xi}_t = (\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{b}_t)$ is finitely supported, and $\boldsymbol{c}_t \in L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_t})$ is integrable with $\boldsymbol{c}_t \in -\mathrm{Cone}(\boldsymbol{A}_t^\top)$ almost surely.

Note that Assumption 4.1 does not require independence between $\boldsymbol{c}_t$ and $\boldsymbol{\xi}_t$. For $t \in [T]$, and $\xi = (A, B, b) \in \mathrm{supp}(\boldsymbol{\xi}_t)$ we define the coupling polyhedron

$$P_t(\xi) := \{(x_{t-1}, x_t) \in \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{n_t} \mid Ax_t + Bx_{t-1} \leqslant b\}, \tag{4.33}$$

and consider, for $x_{t-1} \in \mathbb{R}^{n_{t-1}}$,

$$\widetilde{V}_t(x_{t-1}|\xi) := \mathbb{E}\big[\min_{x_t \in \mathbb{R}^{n_t}} \boldsymbol{c}_t^\top x_t + V_{t+1}(x_t) + \mathbb{I}_{Ax_t + Bx_{t-1} \leqslant b} \mid \boldsymbol{\xi}_t = \xi\big]. \tag{4.34}$$

Then, the cost-to-go function V_t is obtained by

$$V_t(x_{t-1}) = \sum_{\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)} \mathbb{P}[\boldsymbol{\xi}_t = \xi] \widetilde{V}_t(x_{t-1}|\xi) \tag{4.35}$$

[2]The results can be adapted to non-independent $\boldsymbol{\xi}_t$ as long as $\boldsymbol{c}_t$ is independent of $(\boldsymbol{c}_\tau)_{\tau<t}$ conditionally on $(\boldsymbol{\xi}_{\tau \leqslant t})$.

The next two theorems extend the quantization results of Theorem 4.3 to the multistage settings.

Theorem 4.14 (Affine regions independent of the cost)**.** *Assume that $(\boldsymbol{\xi}_t)_{t\in[T]}$ is a sequence of independent, finitely supported, random variables. We define by induction $\mathcal{P}_{T+1} := \{\mathbb{R}^{n_T}\}$ and for $t \in \{2,\dots,T\}$*

$$\mathcal{P}_{t,\xi} := \mathcal{C}\Big((\mathbb{R}^{n_t} \times \mathcal{P}_{t+1}) \wedge \mathcal{F}\big(P_t(\xi)\big)\,,\, \pi_{x_{t-1}}^{x_{t-1},x_t}\Big) \tag{4.36a}$$

$$\mathcal{P}_t := \bigwedge_{\xi_t\in\operatorname{supp}\boldsymbol{\xi}_t} \mathcal{P}_{t,\xi} \tag{4.36b}$$

Then, for all costs distributions $(\boldsymbol{c}_t)_{2\leqslant t\leqslant T}$ such that $(\boldsymbol{c}_t,\boldsymbol{\xi}_t)_{2\leqslant t\leqslant T}$ satisfies Assumption 4.1 and all $t \in \{2,\dots,T\}$, we have $\operatorname{supp}(\mathcal{P}_t) = \operatorname{dom}(V_t)$, and V_t is polyhedral and affine on each cell of $\mathcal{P}_t$.

Proof. We set for all $t \in \{2,\dots,T+1\}$, $\mathcal{V}_t := \pi_{x_{t-1}}^{x_{t-1},z}\big(\mathcal{F}_{\text{low}}\big(\operatorname{epi}(V_t)\big)\big)$ the affine regions of V_t. As $V_{T+1} \equiv 0$ is polyhedral and affine on $\mathbb{R}^{n_T}$, we have $\mathcal{P}_{T+1} = \mathcal{V}_{T+1}$. Assume now that for $t \in \{2,\dots,T\}$, V_{t+1} is polyhedral and $\mathcal{P}_{t+1}$ refines $\mathcal{V}_{t+1}$ (i.e. V_{t+1} is affine on each cell $\sigma \in \mathcal{P}_{t+1}$).

By Lemma 4.12, $\widetilde{V}_t(\cdot|\xi)$, defined in Eq. (4.34), is affine on each cell of $\mathcal{C}(\mathbb{R}^{n_t} \times \mathcal{V}_{t+1} \wedge \mathcal{F}\big(P_t(\xi)\big), \pi_{x_{t-1}}^{x_{t-1},x_t})$ which is refined by $\mathcal{P}_{t,\xi} = \mathcal{C}(\mathbb{R}^{n_t} \times \mathcal{P}_{t+1} \wedge \mathcal{F}\big(P_t(\xi)\big), \pi_{x_{t-1}}^{x_{t-1},x_t})$ by induction hypothesis and Lemma 4.11. Thus, by Eq. (4.35), V_t is affine on each cell of $\mathcal{P}_t$. In particular, V_t is polyhedral and $\mathcal{P}_t := \bigwedge_{\xi_t\in\operatorname{supp}\boldsymbol{\xi}_t} \mathcal{P}_{t,\xi}$ refines $\mathcal{V}_t$. Backward induction ends the proof. □

By Lemma 4.12, we have that $\mathcal{P}_{t,\xi} \preccurlyeq \mathcal{C}(\operatorname{epi}\big(Q_t^\xi\big), \pi_{x_{t-1}}^{x_{t-1},x_t,z})$ where $Q_t^\xi(x_{t-1},x_t) := V_{t+1}(x_t) + \mathbb{I}_{Ax_t+Bx_{t-1}\leqslant b_t}$. In particular, consider $\sigma \in \mathcal{P}_{t,\xi}$, then for all $x_{t-1} \in \operatorname{ri}(\sigma)$, all fibers $\operatorname{epi}(Q_t^\xi)_{x_{t-1}}$ are normally equivalent. We can then define $\mathcal{N}_{t,\xi,\sigma} := \mathcal{N}(\operatorname{epi}(Q_t^\xi)_{x_{t-1}})$ for an arbitrary $x_{t-1} \in \operatorname{ri}(\sigma)$.

The next result shows that we can replace the MSLP problem Eq. (4.2) by an equivalent problem with a discrete cost distribution.

Theorem 4.15 (Exact quantization of the cost distribution, Multistage case)**.** *Assume that $(\boldsymbol{\xi}_t)_{t\in[T]}$ is a sequence of independent, finitely supported, random variables. Then, for all costs distributions such that $(\boldsymbol{c}_t,\boldsymbol{\xi}_t)_{2\leqslant t\leqslant T}$ satisfies Assumption 4.1, for all $t \in [T]$, all $x_{t-1} \in \mathbb{R}^{n_{t-1}}$ and all $\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)$, we have a quantized version of Eq.* (4.34)*:*

$$\widetilde{V}_t(x_{t-1}|\xi) = \sum_{N\in\mathcal{N}_{t,\xi}} \check{p}_{t,N|\xi} \min_{x_t\in\mathbb{R}^{n_t}} \Big\{\check{c}_{t,N|\xi}^\top x_t + V_{t+1}(x_t) + \mathbb{I}_{Ax_t+Bx_{t-1}\leqslant b}\Big\} \tag{4.37}$$

where $\mathcal{N}_{t,\xi} := \bigwedge_{\sigma\in\mathcal{P}_{t,\xi}} -\mathcal{N}_{t,\xi,\sigma}$ and for all $\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)$ and $N \in \mathcal{N}_{t,\xi}$ we denote

$$\check{p}_{t,N|\xi} := \mathbb{P}\big[(\boldsymbol{c}_t,1) \in \operatorname{ri} N \mid \ \boldsymbol{\xi}_t = \xi\big] \tag{4.38a}$$

$$\check{c}_{t,N|\xi} := \begin{cases} \mathbb{E}\big[\boldsymbol{c}_t \mid (\boldsymbol{c}_t,1) \in \operatorname{ri} N, \boldsymbol{\xi}_t = \xi\big] & \text{if } \mathbb{P}\big[\boldsymbol{\xi}_t = \xi, (\boldsymbol{c}_t,1) \in \operatorname{ri} N\big] \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.38b}$$

Proof. Since $\widetilde{V}_t(x_{t-1}|\xi) = \mathbb{E}\big[\min_{x_t\in\mathbb{R}^{n_t},z\in\mathbb{R}} \boldsymbol{c}^\top x_t + z + \mathbb{I}_{(x_{t-1},x_t,z)\in\operatorname{epi}(Q_t^\xi)}\big]$ and the polyhedral complex $\mathcal{P}_{t,\xi}$ refines $\mathcal{C}(\operatorname{epi}\big(Q_t^\xi\big), \pi_{x_{t-1}}^{x_{t-1},x_t,z})$, by applying Theorem 4.3 with variables (x_t,z) and the coupling constraints polyhedron $\operatorname{epi}(Q_t^\xi)$, we deduce that the coefficients $(\check{p}_{t,N|\xi})_{N\in\mathcal{N}_{t,\xi}}$ and $(\check{c}_{t,N|\xi})_{N\in\mathcal{N}_{t,\xi}}$ satisfy

$$\widetilde{V}_t(x_{t-1}|\xi) = \sum_{N\in\mathcal{N}_{t,\xi}} \check{p}_{t,N|\xi} \min_{x_t\in\mathbb{R}^{n_t},z\in\mathbb{R}} \Big\{\check{c}_{t,N|\xi}^\top x_t + z + \mathbb{I}_{(x_{t-1},x_t,z)\in\operatorname{epi}(Q_t^\xi)}\Big\} \tag{4.39}$$

as the deterministic coefficient before z is equal to its conditional expectation. □

In particular, the MSLP problem is equivalent to a finitely supported MSLP as shown in the following result.

For $t_0 \in [T-1]$, we construct the scenario tree $\mathcal{T}_{t_0}$ as follows. A node of depth $t-t_0$ of $\mathcal{T}_{t_0}$ is labelled by a sequence $(N_\tau, \xi_\tau)_{t_0<\tau\leqslant t}$ where $N_\tau \in \mathcal{N}_{\tau,\xi_\tau}$ and $\xi_\tau \in \operatorname{supp}(\boldsymbol{\xi}_\tau)$. In this way, a node of depth $t-t_0$ of $\mathcal{T}_{t_0}$ keeps track of the sequence of realizations of the random variables $\boldsymbol{\xi}_\tau$ for times τ between t_0 and t, and of a selection of cones in $\mathcal{N}_{t,\xi_t}$ at the same times. Note that, by the independence assumption, all the subtrees of $\mathcal{T}_{t_0}$, starting from a node of depth $t-t_0$ are the same as $\mathcal{T}_{t_0+t}$. We denote by $\operatorname{lv}(\mathcal{T}_{t_0})$ the set of leaves of $\mathcal{T}_{t_0}$.

Corollary 4.16 (Equivalent finite tree problem)**.** *Define the quantized probability cost $c_\nu := \check{c}_{t,N_t|\xi_t}$ and probability $p_\nu := \prod_{t_0<\tau\leqslant t} p_{\xi_\tau} \check{p}_{\tau,N_\tau|\xi_\tau}$, for all nodes $\nu = (N_\tau, \xi_\tau)_{t_0<\tau\leqslant t}$. Then, the cost-to-go functions associated with Eq.* (4.1) *are given by*

$$V_{t_0}(\check{x}) = \min_{(x_\nu)_{\nu\in\mathcal{T}_{t_0}}} \quad \sum_{\nu\in\mathcal{T}_{t_0}} p_\nu c_\nu^\top x_\nu \tag{4.40a}$$

$$s.t. \quad Ax_\mu + Bx_\nu \leqslant b \qquad \forall \nu \in \mathcal{T}_{t_0}\setminus \operatorname{lv}(\mathcal{T}_{t_0}), \forall \mu \succcurlyeq \nu \ , \tag{4.40b}$$

for all $2 \leqslant t_0 \leqslant T-1$. Here, $\check{x}$ is the value of x at the root node of $\mathcal{T}_{t_0}$, and the notation $\forall \mu = (\nu, N, A, B, b) \succcurlyeq \nu$ indicates that μ ranges over the set of children of ν.

4.3.3 Dual characterization through nested fiber polyhedra*

We now show that the dual characterization of the expected cost-to-go function through a weighted fiber polyhedron in Theorem 4.5 can be extended to the multistage case. Leveraging dynamic programming equations, we can define at each stage a weighted fiber polyhedron depending on the weighted fiber polyhedron of the next stage. We call them, the *nested* fiber polyhedra.

We denote the dual admissible set at stage t as

$$D_{t,c_t} := \{\mu_t | \mu_t \geqslant 0, A_t^\top \mu_t + c_t = 0\} \tag{4.41}$$

For sake of simplicity, we present the case where the constraints are deterministic.

Assumption 4.2. For each $t \in \{2, \cdots, T\}$, $\boldsymbol{\xi}_t = (\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{b}_t)$ is deterministic and almost surely equal to $\xi_t = (A_t, B_t, b_t)$ Further, the sequence $(\boldsymbol{c}_t)_{2\leqslant t\leqslant T}$ is independent and $\boldsymbol{c}_t \in L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_t})$ is integrable with $\boldsymbol{c}_t \in -\operatorname{Cone}(\boldsymbol{A}_t^\top)$ almost surely.

We now define a *nested* extension of weighted fiber polyhedra and give a dual characterization of the expected cost-to-go function of an MSLP with deterministic constraints. We denote by $\lambda_{[t:T]}$, the vector $(\lambda_t, \lambda_{t+1}, \cdots, \lambda_T)$.

Theorem 4.17. *Assume Assumption 4.2 holds. We define by backward induction the nested fiber polyhedra $E_t \subset \mathbb{R}^{n_t} \times \cdots \times \mathbb{R}^{n_T}$:*

$$F_{T,c_T} := D_{T,c_T} \tag{4.42}$$

$$E_t := \mathbb{E}\big[F_{t,\boldsymbol{c}_t}\big] \qquad \forall t \in [T] \tag{4.43}$$

$$F_{t,c_t} := \big\{(\lambda_{[t:T]}) \,\big|\, \lambda_t \in D_{t,c_t+B_{t+1}^\top \lambda_{t+1}},\ \lambda_{[t+1:T]} \in E_{t+1}\big\} \qquad \forall t \in [T-1] \tag{4.44}$$

Then, at each stage t, the expected cost-to-function can be expressed thanks to the support function of a nested fiber polyhedron. More precisely, for all $t \in \{2, \cdots, T\}$ and all $x_{t-1} \in \mathbb{R}^{n_{t-1}}$, we have

$$V_t(x_{t-1}) = \sigma_{E_t}\Big(B_t x_{t-1} - b_t, -b_{[t+1:T]}\Big). \tag{4.45}$$

Observe that this theorem gives an explicit polyhedral representation of the expected cost-to-go functions as the support functions of the nested fiber polyhedra.

Proof. Note that by backward induction on t, F_{t,c_t} is a polyhedron depending affinely on c_t and that E_t is indeed a polyhedron by Theorem 4.6.

We show the result by backward induction. For $t = T$, the proof is the same as Theorem 4.5. Let $t \leqslant T-1$ and assume the result holds for $t+1$, we have by definition

$$\hat{V}_t(x_{t-1}, c_t) = \inf_{x_t \in \mathbb{R}^{n_t}} c_t^\top x_t + V_{t+1}(x_t) + \mathbb{I}_{A_t x_t + B_t x_{t-1} \leqslant b_t} \tag{4.46}$$

Thinking of the rigth-hand side of this equation as a linear problem and denoting by $V_{t+1}^\star$ the Fenchel transform of V_{t+1}, by strong duality, we also have

$$\hat{V}_t(x_{t-1}, c_t) = \sup_{\lambda_t \in \mathbb{R}_+^{m_t}} (B_t x_{t-1} - b_t)^\top \lambda_t - V_{t+1}^\star(-c_t - A_t^\top \lambda_t) \tag{4.47}$$

Using the induction assumption, we have

$$\begin{aligned}
V_{t+1}^\star(\psi_t) &= \sup_{x_t \in \mathbb{R}_t^n} \quad \psi_t^\top x_t - V_{t+1}(x_t) && \text{(4.48a)}\\
&= \sup_{x_t \in \mathbb{R}_t^n} \quad \psi_t^\top x_t - \sup_{\lambda_{[t+1:T]} \in E_t} \quad (B_{t+1} x_t - b_{t+1})^\top \lambda_{t+1} - \sum_{s=t+2}^{T} b_s^\top \lambda_s && \text{(4.48b)}\\
&= \sup_{x_t \in \mathbb{R}_t^n} \quad \inf_{\lambda_{[t+1:T]} \in E_{t+1}} \quad \psi_t^\top x_t + (b_{t+1} - B_{t+1} x_t)^\top \lambda_{t+1} + \sum_{s=t+2}^{T} b_s^\top \lambda_s && \text{(4.48c)}\\
&= \inf_{\lambda_{[t+1:T]} \in E_{t+1}} \quad \sum_{s=t+1}^{T} b_s^\top \lambda_s + \sup_{x_t \in \mathbb{R}_t^n} \quad x_t^\top (\psi_t - B_{t+1}^\top \lambda_{t+1}) && \text{(4.48d)}\\
&= \inf_{\lambda_{[t+1:T]} \in E_{t+1}} \quad \sum_{s=t+1}^{T} b_s^\top \lambda_s + \mathbb{I}_{B_{t+1}^\top \lambda_{t+1} = \psi_t} && \text{(4.48e)}
\end{aligned}$$

Thus, by putting this new expression into Eq. (4.47), we have

$$\begin{aligned}
\hat{V}_t(x_{t-1}, c_t) &= \sup_{\lambda_t \in \mathbb{R}_+^{m_t}} (B_t x_{t-1} - b_t)^\top \lambda_t - \inf_{\lambda_{[t+1:T]} \in E_{t+1}} \sum_{s=t+1}^{T} b_s^\top \lambda_s + \mathbb{I}_{A_t^\top \lambda_t + c_t + B_{t+1}^\top \lambda_{t+1} = 0} && \text{(4.49a)}\\
&= \sup_{\lambda_{[t:T]}} (B_t x_{t-1} - b_t)^\top \lambda_t - \sum_{s=t+1}^{T} b_s^\top \lambda_s + \mathbb{I}_{\lambda_{[t+1:T]} \in E_{t+1}} + \mathbb{I}_{\lambda_t \in D_{t, c_t + B_{t+1}^\top \lambda_{t+1}}} && \text{(4.49b)}\\
&= \sup_{\lambda_{[t:T]}} (B_t x_{t-1} - b_t)^\top \lambda_t - \sum_{s=t+1}^{T} b_s^\top \lambda_s + \mathbb{I}_{\lambda_{[t:T]} \in F_{t,c_t}} && \text{(4.49c)}
\end{aligned}$$

By interchangeability principle (see the proof of Theorem 4.5), we have

$$\begin{aligned}
\hat{V}(x_{t-1}) &= \mathbb{E}\Big[\sup_{\lambda_{[t:T]} \in F_{t, \mathbf{c}_t}} (B_t x_{t-1} - b_t)^\top \lambda_t - \sum_{s=t+1}^{T} b_s^\top \lambda_s \Big] && \text{(4.50a)}\\
&= \sup_{\lambda_{[t:T]} \in E_t} (B_t x_{t-1} - b_t)^\top \lambda_t - \sum_{s=t+1}^{T} b_s^\top \lambda_s && \text{(4.50b)}\\
&= \sigma_{E_t}(B_t x_{t-1} - b_t, -b_{[t+1:T]}) && \text{(4.50c)}
\end{aligned}$$

□

4.4 An illustrative example

We consider the following second-stage problem, with $n = 1$ and $m = 2$:

$$V(x) = \mathbb{E}\begin{bmatrix} \min\limits_{y\in\mathbb{R}^2} & \boldsymbol{c}^\top y \\ \text{s.t.} & \|y\|_1 \leqslant 1, \quad y_1 \leqslant x \text{ and } y_2 \leqslant x \end{bmatrix}. \tag{4.51}$$

We apply our results, to provide an explicit representation of V.

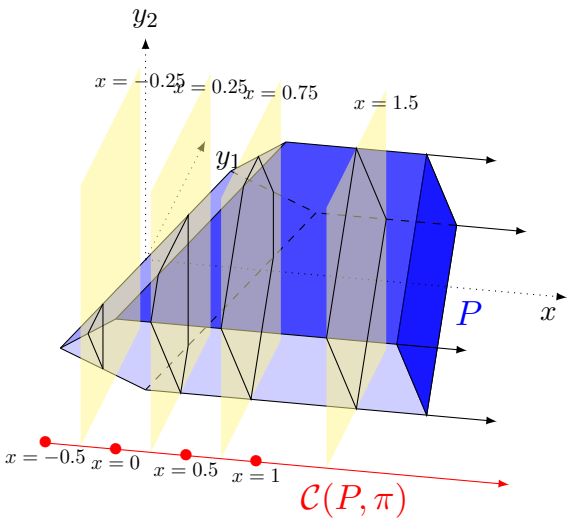
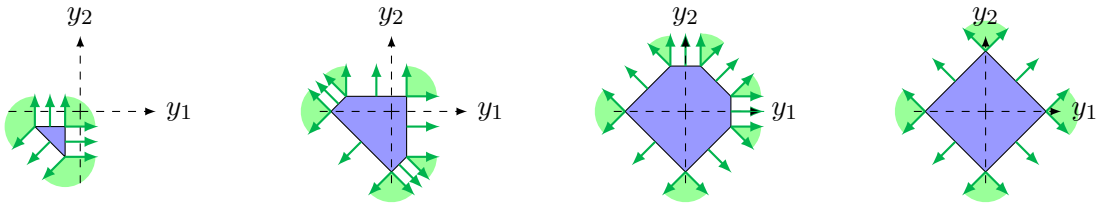


Figure 4.3: The coupling polyhedron P in blue, different cuts and fibers P_x vertical in yellow, and its chamber complex $\mathcal{C}(P, \pi)$ in red on the bottom.

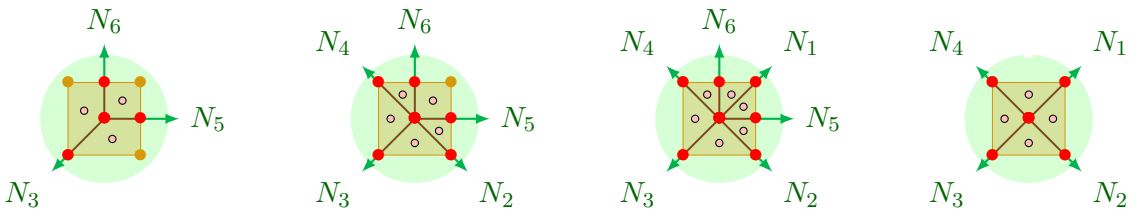


(a) $x = -0.25, \sigma = [-0.5, 0]$ (b) $x = 0.25, \sigma = [0, 0.5]$ (c) $x = 0.75, \sigma = [0.5, 1]$ (d) $x \geqslant 1, \sigma = [1, +\infty)$

Figure 4.4: Fibers P_x in blue and their normal fan $\mathcal{N}(P_x) = \mathcal{N}_\sigma$ in green for different $x \in \mathbb{R}$.

The coupling polyhedron is $P = \{(x, y) \in \mathbb{R} \times \mathbb{R}^2 \,|\, \|y\|_1 \leqslant 1,\, y_1 \leqslant x,\, y_2 \leqslant x\}$ presented in Fig. 4.3, and its V-representation is the collection of vertices $(0, -1, 0)$, $(-0.5, -0.5, -0.5)$, $(0, 0, -1)$, $(1, 1, 0)$, $(0.5, 0.5, 0.5)$, $(1, 0, 1)$ and the ray $(1, 0, 0)$. By projecting the different faces, we see that its projection is the half-line, $\pi(P) = [-0.5, +\infty)$ and its chamber complex $\mathcal{C}(P, \pi)$ is the collection of cells composed of $\{-0.5\}$, $[-0.5, 0]$, $\{0\}$, $[0, 0.5]$, $\{0.5\}$, $[0.5, 1]$, $\{1\}$, $[1, +\infty)$ as presented in Fig. 4.3. As there are 4 different maximal chambers, there are 4 different classes of normally equivalent fibers as shown in Fig. 4.4.

We evaluate $\check{c}_N$ and $\check{p}_N$ for $N \in -\mathcal{N}_\sigma$ using the formulas of Table B.1. For example, when $\boldsymbol{c}$ is uniform on the centered ball for the ∞-norm of radius R, Fig. 4.5 shows the regions of which the areas and centroids need to be computed.



(a) $\sigma = [-0.5, 0]$ (b) $\sigma = [0, 0.5]$ (c) $\sigma = [0.5, 1]$ (d) $\sigma = [1, +\infty)$

Figure 4.5: Exact quantization illustrated. The normal fan $\mathcal{N}_\sigma$ in green with $N_i = W_i^\top \mathbb{R}^+$, $\boldsymbol{c}$ is uniform on the support $Q = -Q = B_\infty(0, R)$ in light orange, the sets $W_i^\top \mathbb{R}^+ \cap Q$ in red. The polyhedral complex $\mathcal{R}_\sigma$ shown in red or orange. The quantized costs $\check{c}_N$ are determined by centroids (small circles in pink).

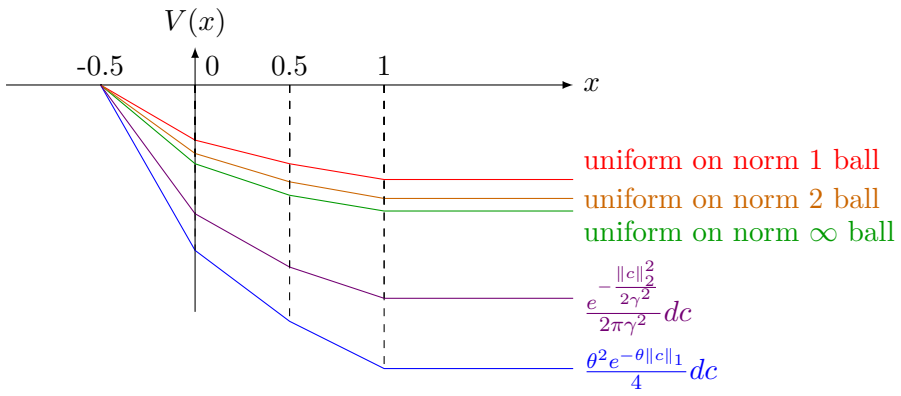


Figure 4.6: Graph of the cost-to-go function V for different distribution of the cost $\boldsymbol{c}$ with $R = \theta = \gamma = 1$.

$d\mathbb{P}(c)$	$-\frac{1}{2} \leqslant x \leqslant 0$	$0 \leqslant x \leqslant \frac{1}{2}$	$\frac{1}{2} \leqslant x \leqslant 1$	$1 \leqslant x$
$\frac{\mathbb{1}_{\|c\|_1 \leqslant R}}{2R^2} dc$	$\frac{-7R}{24}(1+2x)$	$\frac{-R}{24}(7+6x)$	$\frac{-R}{6}(2+x)$	$\frac{-R}{2}$
$\frac{\theta^2 e^{-\theta\|c\|_1}}{4} dc$	$\frac{-7}{8\theta}(1+2x)$	$\frac{-1}{8\theta}(7+6x)$	$\frac{-1}{2\theta}(2+x)$	$\frac{-3}{2\theta}$
$\frac{\mathbb{1}_{\|c\|_\infty \leqslant R}}{4R^2} dc$	$\frac{-R}{12}(5+10x)$	$\frac{-R}{12}(5+4x)$	$\frac{-R}{6}(3+x)$	$\frac{-2R}{3}$
$\frac{e^{-\|c\|_2^2/2\gamma^2}}{2\pi\gamma^2} dc$	$\frac{-\gamma(2+\sqrt{2})(1+2x)}{2\sqrt{2\pi}}$	$\frac{-\gamma(2+\sqrt{2}+2\sqrt{2}x)}{2\sqrt{2\pi}}$	$\frac{-2\gamma(1+(-1+\sqrt{2})x)}{\sqrt{2\pi}}$	$-\frac{2}{\sqrt{\pi}}\gamma$
$\frac{\mathbb{1}_{\|c\|_2 \leqslant R}}{\pi R^2} dc$	$\frac{-R(2+\sqrt{2})(1+2x)}{3\pi}$	$\frac{-R(2+\sqrt{2}+2\sqrt{2}x)}{3\pi}$	$\frac{-4R(1+(-1+\sqrt{2})x)}{3\pi}$	$-\frac{4\sqrt{2}R}{3\pi}$

Table 4.1: Different values of $V(x)$ for different distributions of the cost $\boldsymbol{c}$.

4.5 Complexity

Hanasusanto, Kuhn and Wiesemann showed in [HKW16] that 2-stage stochastic programming is ♯P-hard, by reducing the computation of the volume of a polytope to the resolution of a 2-stage stochastic program. Nevertheless, we show that for a fixed dimension of the recourse space, 2-stage programming is polynomial. Therefore, the status of 2-stage programming seems somehow comparable to the one of the computation of the volume of a polytope – which is also both ♯P-hard and polynomial when the dimension is fixed (see for example [GK94, 3.1.1]). We

also give a similar result for multistage stochastic linear programming.

We now give a summary of our method. A naive approach would be to use directly the exact quantization result Theorem 4.3. However, even in the two-stage case, the latter yields a linear program of an exponential size when only the recourse dimension m is fixed. Indeed, the size of the quantized linear program, $(2SLP)$ is polynomial only when *both* n and m are fixed as $\bigwedge_{\sigma\in\mathcal{C}(P,\pi)} -\mathcal{N}_\sigma$ can have, by McMullen's and Stanley's upper bound theorems ([McM70, Sta75]), an exponential size in n and m, and these bounds are tight. Hence, to handle the case in which only the recourse dimension m is fixed, we need additional ideas. We use the local quantization result Theorem 4.2, observing that when m is fixed, $\mathcal{N}(P_x)$ has a polynomial size. We thus have a polynomial time oracle that gives the values $V(x)$ and a subgradient $g \in \partial V(x)$ by Theorem 4.9. Then, we rely on the theory of linear programming with oracle [GLS12], working in the *Turing model* of computation (a.k.a. *bit model*). In particular, all the computations are carried out with rational numbers. We now provide the needed details of the proof.

4.5.1 Multistage programming with exact oracles

Recall that a polyhedron can be given in two manners. The "H-representation" provides an external description of the polyhedron, as the intersection of finitely many half-spaces. The "V-representation" provides an internal representation, writing the polyhedron as a Minkowski sum of a polytope (given as the convex hull of finitely many points) and of a polyhedral cone (generated by finitely many vectors).

We say that a polyhedron is *rational* if the inequalities in its H-representation are rational or, equivalently, the generators of its V-representation have rational coefficients. We shall say that a (convex) polyhedral function V is *rational* if its epigraph is a rational polyhedron.

Recall that, in the Turing model, the *size* (or encoding length see [GLS12, 1.3]) of an integer $k \in \mathbb{Z}$ is $\langle k\rangle := 1 + \lceil \log_2(|k| + 1)\rceil$; the size of a rational $r = \frac{p}{q} \in \mathbb{Q}$ with p and q coprime integers, is $\langle r\rangle := \langle p\rangle + \langle q\rangle$. The size of a rational matrix or a vector, still denoted by $\langle\cdot\rangle$, is the sum of the sizes of its entries. The size of an inequality $\alpha^\top x \leqslant \beta$ is $\langle\alpha\rangle + \langle\beta\rangle$. The size of a H-representation of a polyhedron is the sum of the sizes of its inequalities and the size of a V-representation of a polyhedron is the sum of the sizes of its generators.

If the dimension of the ambient space is *fixed*, one can pass from one representation to the other one in *polynomial time*. Indeed, the double description algorithm allows one to get a V-representation from a H-representation, see the discussion at the end of section 3.1 in [FP95], and use McMullen's upper bound theorem ([McM70] and [GLS12, 6.2.4]) to show that the computation time is polynomially bounded in the size of the H-representation. A fortiori, the size of the V-representation is polynomially bounded in the size of the H-representation. Dually, the same method allows one to obtain a H-representation from a V-representation. Hence, in the sequel, we shall use the term *size* of a polyhedron for the size of a V or H-representation: when dealing with polynomial-time complexity results in fixed dimension, whichever representation is used is irrelevant. In particular, we define the *size* $\langle N\rangle$ of a rational cone N as the size of a H or V representation of N.

We first observe that the size of the scenario tree arising in the exact quantization result becomes polynomial when suitable dimensions are fixed.

Proposition 4.18. *Let $t \in \{2,\dots,T\}$, and suppose that the dimensions $n_t,\dots,n_T$ and the cardinals $\sharp(\operatorname{supp}\boldsymbol{\xi}_t)$, $\cdots$,$\sharp(\operatorname{supp}\boldsymbol{\xi}_T)$ are fixed. Let $\mathcal{T}$ be the scenario tree constructed in Corollary 4.16. Then, the subtree of $\mathcal{T}$ rooted at an arbitrary node of depth t can be computed in polynomial in $\sum_{s=t}^{T}\sum_{\xi\in\operatorname{supp}(\boldsymbol{\xi}_s)}\langle\xi\rangle$.*

Proof. Recall that the number of chambers of a chamber complex is polynomial when both dimensions are fixed by [VWBC05, 3.9]. Thus, we can compute recursively the (maximal) chambers of the complexes $\mathcal{P}_t$ defined in Theorem 4.14 thanks to the algorithm in [CL98, 3.2] in

polynomial time. We then can compute in polynomial time the fans $\mathcal{N}_t$ defined in Theorem 4.15. □

We recall the theory of linear programming with oracle applies to the class of "well described" polyhedra which are rational polyhedra with an a priori bound on the bit-sizes of the inequalities defining their facets, we refer the reader to [GLS12] for a more detailed discussion of the notions (oracles) and results used here.

Definition 4.19 (first-order oracle)**.** *Let f be a rational polyhedral function. We say that f admits a polynomial time (exact)* first-order oracle*, if there exists an oracle that takes as input a vector x and either returns a hyperplane separating x from* $\mathrm{dom}(f)$ *if $x \notin \mathrm{dom}(f)$ or returns $f(x)$ and $g \in \partial V(x)$ if $x \in \mathrm{dom}(f)$, in polynomial time in $\langle x \rangle$.*

Lemma 4.20. *Let $Q \subset \mathbb{R}^d$ be a polyhedron, $c \in \mathbb{R}^d$ a cost vector and f be a polyhedral function given by a first-order oracle. Furthermore, assume* $\mathrm{epi}(f)$ *and Q are well described. Then, the problem $\min_{x \in Q} c^\top x + f(x)$ can be solved in oracle-polynomial time in $\langle c \rangle + \langle \mathrm{epi}(f) \rangle + \langle Q \rangle$.*

Proof. The case where $\mathrm{dom}(f) = \mathbb{R}^d$ is tackled in Theorem 6.5.19 in [GLS12]. If f has a general domain, we can write $f = \widetilde{f} + \mathbb{I}_{\mathrm{dom}\, f}$ where $\widetilde{f}$ is a polyhedral function with a well described epigraph and such that $\mathrm{dom}\, \widetilde{f} = \mathbb{R}^d$. Then, noting that $\mathrm{epi}(f) = \mathrm{epi}(\widetilde{f}) \cap \mathrm{dom}(f) \times \mathbb{R}$, we can adapt the proof of the latter theorem, using Exercise 6.5.18(a) of [GLS12]. □

We do not require the distribution of the cost $\boldsymbol{c}$ to be described extensively. We only need to assume the existence of the following oracle.

Definition 4.21 (cone-valuation oracle)**.** *Let $\boldsymbol{c} \in L_1(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R}^m)$ be an integrable cost distribution such that, for every* rational *cone N, the quantized probability $\check{p}_N$ and quantized cost $\check{c}_N$ are* rational*. We say that $\boldsymbol{c}$ admits a polynomial time (exact)* cone-valuation oracle*, if there exists an oracle which takes as input a rational polyhedral cone N and returns $\check{p}_N$ and $\check{c}_N$ in* polynomial time *in $\langle N \rangle$.*

Theorem 4.22 (Cone valuation to first-order oracle)**.** *Consider the value functions of problem* (4.1) *defined in Eq.* (4.2)*. Assume that $T, n_2, \dots, n_T$, $\sharp(\mathrm{supp}\, \boldsymbol{\xi}_2)$, $\cdots$,$\sharp(\mathrm{supp}\, \boldsymbol{\xi}_T)$ are fixed integers, and that $(\boldsymbol{c}_t, \boldsymbol{\xi}_t)_{2 \leqslant t \leqslant T}$ satisfies Assumption 4.1. Assume in addition that, every vector $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$ has rational entries and that the probabilities $p_{t,\xi} := \mathbb{P}[\boldsymbol{\xi}_t = \xi]$ are rational numbers. Assume finally that every random variable $\boldsymbol{c}_t$ conditionally to $\{\boldsymbol{\xi}_t = \xi\}$, denoted by $\boldsymbol{c}_{t,\xi}$, admits a polynomial-time cone-valuation oracle (see Definition 4.21).*

Then, for all $t \geqslant 2$, V_t admits a polynomial time first-order oracle.

Proof. We start with the 2-stage case with deterministic constraints. We recall our notation $V(x) := \mathbb{E}[\min_{y \in \mathbb{R}^m} \boldsymbol{c}^\top y + \mathbb{I}_{Ay + Bx \leqslant b}]$. Let $x \in \mathbb{R}^n$ be an input vector. We first check if $x \in \pi(P) = \mathrm{dom}(V)$. By solving the dual of $\min_{y \in \mathbb{R}^m} \{\, 0 \,|\, Ay \leqslant b - Bx\}$, we either find an unbounded ray generated by $\mu \in \mathbb{R}^q$ such that $\mu \geqslant 0$, $\mu^\top A = 0$ and $\mu^\top (b - Bx) < 0$ or a $y \in \mathbb{R}^m$ such that $Ay \leqslant b - Bx$, so that $x \in \pi(P)$. In the former case we have $x \notin \pi(P)$, and we get a cut $\{x' \in \mathbb{R}^n \,|\, \mu^\top B x' = \frac{\mu^\top b + \mu^\top A x}{2}\}$, separating $\pi(P) = \mathrm{dom}(V)$ from x.

So, we now assume that $x \in \pi(P)$, i.e., $V(x) < +\infty$. We next show that we can compute $V(x)$ and a subgradient $\alpha \in \partial V(x)$ in polynomial time. Indeed, the McMullen upper-bound theorem [McM70], in its dual version, guarantees that a polytope of dimension m with f facets has $O(f^{\lfloor m/2 \rfloor})$ faces, see [Sei95]. Since the number of cones in $\mathcal{N}(P_x)$ is equal to the number of faces of P_x which is polynomially bounded in the number of constraints $q \leqslant \langle \xi \rangle$, $\sharp \mathcal{N}(P_x)$ is polynomial in $\langle \xi \rangle$. Thus, since $\boldsymbol{c}$ is given by a cone valuation oracle, we can compute in polynomial time the collection of all quantized costs and probabilities $\check{c}_N$ and $\check{p}_N$, indexed by $N \in -\mathcal{N}(P_x)$. Then, by Theorem 4.9, we can compute $V(x)$ and a subgradient $\alpha \in \partial V(x)$ by solving a linear program for each cone $N \in -\mathcal{N}_\sigma$. All these operations take a polynomial time.

The case of finitely supported stochastic constraints reduces to the case of deterministic constraints dealt with above, using $\mathrm{dom}(V) = \cap_{\xi\in\mathrm{supp}\,\boldsymbol{\xi}}\pi(P(\xi))$ and $V(x) = \sum_{\xi\in\mathrm{supp}\,\boldsymbol{\xi}} p_\xi \widetilde{V}(x|\xi)$ where $\widetilde{V}(x|\xi) := \mathbb{E}\big[\hat{V}(x,\boldsymbol{c},\boldsymbol{\xi}) \mid \boldsymbol{\xi} = \xi\big]$.

We finally deal with the multistage case similarly, using the quantization result Corollary 4.16 in extensive form. Applying Proposition 4.18, the quantized costs and probabilities arising there can be computed by a polynomial number of calls to the cone-valuation oracle. This provides a first order oracle for the expected cost-to-go function V_t. □

We now refine the definition of cone-valuation oracle, to take into account situations in which the distribution of the random cost $\boldsymbol{c}$ is specified by a parametric model. We shall say that such a distribution admits a polynomial-time *parametric cone-valuation oracle* if there is an oracle that takes as input the parameters of the distribution, together with a rational cone N, and outputs the quantized probability $\check{p}_N$ and cost $\check{c}_N$. Especially, we consider the following situations:

1. *Deterministic distribution* equal to a rational cost c. We set $\langle\boldsymbol{c}\rangle := \langle c\rangle$
2. *Exponential distribution on a rational cone* K with rational parameter θ. We set $\langle\boldsymbol{c}\rangle := \langle K\rangle + \langle\theta\rangle$
3. *Uniform distribution on a rational polyhedron* Q such that $\mathrm{Aff}(Q) = \{y \in \mathbb{R}^m \mid \forall j \in J \subset [m], y_j = q_j \in \mathbb{Q}\}$ where J is a subset of $[m]$ and q_j are rational numbers (in particular, Q is full dimensional when $J = \emptyset$). We set: $\langle\boldsymbol{c}\rangle = \langle Q\rangle$
4. *Mixtures of the above distributions*, i.e., convex combination with rational coefficients $(\lambda^k)_{k\in[l]}$ of distributions of random variables $(\boldsymbol{c}_k)_{k\in[l]}$ satisfying 1. 2. or 3. Then, we set $\langle\boldsymbol{c}\rangle = \sum_{k=1}^{l}\langle\boldsymbol{c}_k\rangle + \langle\lambda_k\rangle$.

Theorem 4.23. *Assume that the dimension m is fixed, and that $\boldsymbol{c}$ is distributed according to any of the above laws (deterministic, exponential, uniform, or mixture). Then, the random cost $\boldsymbol{c}$ admits a polynomial-time parametric cone-valuation oracle.*

Proof. 1. *Case of a deterministic distribution.* We first check whether $c \in \mathrm{ri}(N)$, which can be done in polynomial time, see section 6.5 of [GLS12]. Then, if $c \in \mathrm{ri}(N)$, we set $\check{c}_N = c$ and $\check{p}_N = 1$ otherwise $\check{c}_N = 0$ and $\check{p}_N = 0$.

2. *Case of an exponential distribution.* Since the dimension is fixed, for every polyhedron R, we can triangulate $R \cap \mathrm{supp}(\boldsymbol{c})$ and partition it into (relatively open) simplices and simplicial cones $(S_k)_{k\in[l]}$, and by Stanley upper bound theorem, the size l of the triangulation is polynomial in $\langle R\rangle$. By using the Brion formula in Table B.1, we compute in polynomial time $\check{p}_R = \sum_{k=1}^{l}\check{p}_{S_k}$ and $\check{c}_R = \sum_{k=1}^{l}\check{p}_{S_k}\check{c}_{S_k}/\check{p}_R$ if $\check{p}_R = 0$ and $\check{c}_R = 0$ otherwise.

3. *Case of a uniform distribution.* After triangulating (as in the case of an exponential distribution), we may suppose that the support of the distribution is a simplex S, so that $Q = S$. If this simplex S is full dimensional, then its volume is given by a determinantal expression, and so, it is rational (see e.g. [GK94] 3.1). Then, the formulas of Table B.1 yield the result. If this simplex is not full dimensional, we have $\mathrm{Aff}(S) = \{y \in \mathbb{R}^m \mid \forall j \in J, y_j = q_j\}$, a similar formula holds, ignoring the coordinates of y whose indices are in the set J.

4. *Case of mixtures of distributions.* Trivial reduction to the previous cases. □

Remark 4.24. *The conclusion of Theorem 4.23 does not carry over to the uniform distribution on a general polytope of dimension $k < n$. The condition that $\mathrm{Aff}(Q) = \{y \in \mathbb{R}^m \mid \forall j \in J, y_j = q_j\}$ ensures that the orthogonal projection on $\mathrm{Aff}(Q)$ preserves rationality, which entails that the k-dimensional volume of Q is a rational number. In general, this volume is obtained by applying the Cayley Menger determinant formula (see for example [GK94, 3.6.1]), and it belongs to a quadratic extension of the field of rational numbers. For example, if Δ_d is the canonical simplex $\{\lambda \in \mathbb{R}^{d+1}_+ | \sum_{i=1}^{d+1}\lambda_i = 1\}$ then $\mathrm{Vol}(\Delta_d) = \frac{\sqrt{d+1}}{d!}$.*

For the Gaussian distribution, $\check{c}_S$ and $\check{p}_S$ can be determined in terms of solid angles (see [Rib06]) arising in Table B.1. These coefficients generally involve the number π and Euler's Γ function, and thus they are irrational.

Corollary 4.25 (MSLP is polynomial for fixed dimensions)**.** *Consider the problem Eq.* (4.1) *. Assume that $T, n_2, \dots, n_T$, $\sharp(\operatorname{supp} \boldsymbol{\xi}_2)$, $\cdots$,$\sharp(\operatorname{supp} \boldsymbol{\xi}_T)$ are fixed integers, that $(\boldsymbol{c}_t, \boldsymbol{\xi}_t)_{2 \leqslant t \leqslant T}$ satisfies Assumption 4.1. Suppose in addition that, for all $\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)$, $p_{t,\xi} := \mathbb{P}[\boldsymbol{\xi}_t = \xi]$ and ξ are rational and that the random variable $\boldsymbol{c}_t$ conditionally to $\{\boldsymbol{\xi}_t = \xi\}$, denoted by $\boldsymbol{c}_{t,\xi}$, is of the type considered in Theorem 4.23.*

Then, Problem (4.1) *can be solved in a time that is polynomial in the input size $\langle c_1 \rangle + \langle \xi_1 \rangle + \sum_{t=2}^{T} \sum_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} (\langle \boldsymbol{c}_{t,\xi} \rangle + \langle \xi \rangle + \langle p_{t,\xi} \rangle)$.*

Proof. We first show by backward induction that the epigraph $\operatorname{epi}(V_2)$ is well described. The dynamic programming equation Eq. (4.2) allows us to compute a H-representation of $\operatorname{epi}(V_t)$ from a H-representation of $\operatorname{epi}(V_{t+1})$. Indeed, by Theorem 4.15, we have

$$V_t(x_{t-1}) = \sum_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} p_{t,\xi} \sum_{N \in \mathcal{N}_{t,\xi}} \check{p}_{t,N|\xi} \min_{x_t \in \mathbb{R}^{n_t}} Q_{t,N|\xi}(x_t, x_{t-1}) \ , \text{ with}$$
$$Q_{t,N|\xi}(x_t, x_{t-1}) := \check{c}_{t,N|\xi}^{\top} x_t + V_{t+1}(x_t) + \mathbb{I}_{(x_t, x_{t-1}) \in P_t(\xi)} \ .$$

We then have

$$\operatorname{epi}(Q_{t,N|\xi}) = \big(\operatorname{epi}(x_t \mapsto \check{c}_{t,N|\xi}^{\top} x_t) + \operatorname{epi}(V_{t+1}) \big) \ \cap \ (P_t(\xi) \times \mathbb{R})$$
$$\operatorname{epi}(V_t) = \sum_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} p_{t,\xi} \sum_{N \in \mathcal{N}_{t,\xi}} \check{p}_{t,N|\xi} \, \pi_{x_{t-1},z}^{x_{t-1},x_t,z}\big(\operatorname{epi}(Q_{t,N|\xi}) \big) \ ,$$

recalling that $\pi_{x_{t-1},z}^{x_{t-1},x_t,z}$ denotes the projection mapping $(x_{t-1}, x_t, z) \mapsto (x_{t-1}, z)$. Well described polyhedra are stable under the operations of projection, intersection, and Minkowski sum, see in particular [GLS12, 6.5.18]. It follows that $\operatorname{epi}(V_t)$ is well described. Then, the corollary follows from Lemma 4.20, Theorem 4.22 and Theorem 4.23. □

4.5.2 Multistage programming with approximate oracles

We finally consider the situation in which the law of the cost distribution is only known approximately. Hence, we relax the notion of cone-valuation oracle, as follows.

Definition 4.26 (Weak cone-valuation oracle)**.** *Let $\boldsymbol{c} \in L(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R}^m)$ be an integrable cost distribution. We say that $\boldsymbol{c}$ admits a polynomial time* weak cone-valuation oracle*, if there exists an oracle which takes as input a rational polyhedral cone N together with a rational number $\varepsilon > 0$, and returns a rational number $\widetilde{p}_N$ and a rational vector $\widetilde{c}_N$ such that $|\widetilde{p}_N - \check{p}_N| \leqslant \varepsilon$ and $\|\widetilde{c}_N - \check{c}_N\| \leqslant \varepsilon$, in a time that is* polynomial *in $\langle N \rangle + \langle \varepsilon \rangle$.*

Definition 4.27 (Weak first-order oracle)**.** *Let f be a rational polyhedral function. We say that f admits a polynomial time* weak first-order oracle*, if there exists an oracle that takes as input a vector x and either returns a hyperplane separating x from $\operatorname{dom}(f)$ if $x \notin \operatorname{dom}(f)$ or returns a scalar $\widetilde{f}$ and a vector $\widetilde{g}$ such that $|\widetilde{f} - f(x)| \leqslant \varepsilon$ and $d(\widetilde{g}, \partial f(x)) \leqslant \varepsilon$ if $x \in \operatorname{dom}(f)$, in a time which is polynomial in $\langle x \rangle + \langle \varepsilon \rangle$.*

Remark 4.28. *In our definition of weak first order oracle, we require that feasibility ($x \in \operatorname{dom}(f)$) be tested exactly, whereas the value and a subgradient of the function are only given approximately. This is suitable to the present setting, in which the main difficulty resides in the approximation of the function (which may take irrational values for relevant cost distributions).*

We now rely on the theory of linear programming with weak separation oracles developed in [GLS12]. Let $C \subset \mathbb{R}^d$ be a convex set, for $\varepsilon > 0$, let $S(C, \varepsilon) := \{x \in \mathbb{R}^d \mid \|x - y\| \leqslant \varepsilon\}$ and $S(C, -\varepsilon) := \{x \in \mathbb{R}^d \mid B(x, \varepsilon) \subset C\}$ where $B(x, \varepsilon)$ denotes the Euclidean ball centered at x of radius ε. A *weak separation oracle* for a convex set $C \subset \mathbb{R}^d$ takes as argument a vector $x \in \mathbb{R}^d$ and a rational number $\varepsilon > 0$, and either asserts that $x \in S(C, \varepsilon)$ or returns a rational vector $\gamma \in \mathbb{R}^d$, of norm one, and a rational scalar δ, such that $\gamma^\top y \leqslant \gamma^\top x + \varepsilon$ for all $y \in S(C, -\varepsilon)$.

Theorem 4.29 (Weak cone valuation to weak first-order oracle)**.** *Consider the value functions of problem* (4.1) *defined in Eq.* (4.2) *. Assume that $T, n_2, \ldots, n_T$, $\sharp(\operatorname{supp} \boldsymbol{\xi}_2), \cdots, \sharp(\operatorname{supp} \boldsymbol{\xi}_T)$ are fixed integers, and that $(\boldsymbol{c}_t, \boldsymbol{\xi}_t)_{2 \leqslant t \leqslant T}$ satisfies Assumption 4.1. Assume in addition that, every vector $\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)$ has rational entries and that the probabilities $p_{t,\xi} := \mathbb{P}[\boldsymbol{\xi}_t = \xi]$ are rational numbers. Assume finally that the diameters of $\operatorname{dom} V_t$, for $t \geqslant 2$, are bounded by a rational constant R, and that every random variable $\boldsymbol{c}_t$ conditionally to $\{\boldsymbol{\xi}_t = \xi\}$, denoted by $\boldsymbol{c}_{t,\xi}$, admits a polynomial-time weak cone-valuation oracle (see Definition 4.21).*

Then, for all $t \geqslant 2$, V_t admits a polynomial time weak first-order oracle.

Proof. The proof is similar to the one of Theorem 4.22. The main difference is that we need an a priori bound R on the diameter of $\operatorname{dom} V_t$, so that if $d(\widetilde{g}, \partial V_t(x)) \leqslant \varepsilon$, then, using Cauchy-Schwarz inequality, $V_t(y) - V_t(x) \geqslant \widetilde{g} \cdot (y - x) - \varepsilon R$ holds for all $y \in \operatorname{dom} V_t$. Together with and approximation of $V_t(x)$, this allows us to get a weak separation oracle for the epigraph of V_t. □

Corollary 4.30 (Approximate (MSLP) is polynomial-time for fixed recourse dimension m)**.** *Consider Problem* (4.1)*. Assume that $T, n_2, \ldots, n_T$, $\sharp(\operatorname{supp} \boldsymbol{\xi}_2), \cdots, \sharp(\operatorname{supp} \boldsymbol{\xi}_T)$ are fixed integers Assume finally that the diameters of $\operatorname{dom} V_t$, for $t \geqslant 2$, are bounded by a rational constant R, and that for all $\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)$, the random variable $\boldsymbol{c}_t$ conditionally to $\{\boldsymbol{\xi}_t = \xi\}$, denoted by $\boldsymbol{c}_{t,\xi}$, admits a polynomial-time weak cone-valuation oracle.*

Then, there exists an algorithm that either asserts that Problem Eq. (4.1) *is infeasible or find a feasible solution x^* whose cost does not exceed the cost of an optimal solution by more than ε, in polynomial-time in $\langle c_1 \rangle + \langle \xi_1 \rangle + \sum_{t=2}^{T} \sum_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} (\langle \boldsymbol{c}_{t,\xi} \rangle + \langle \xi \rangle + \langle p_{t,\xi} \rangle) + \langle R \rangle$.*

Proof. This follows from Theorem 4.29, using the result analogous to Lemma 4.20 for weak separation oracles, see [GLS12, 6.5.19]. □

Finally, we show that every absolutely continuous cost distribution, with a suitable density function, admits a polynomial-time weak cone-valuation oracle.

Definition 4.31. *We shall say that a density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is* combinatorially tight *if:*

1. *there is a polynomial time algorithm which, given a rational number $\varepsilon > 0$, returns a rational number $r > 0$ such that $\int_{\|x\| > r} f(x) dx \leqslant \varepsilon$.*

2. *there is a polynomial time algorithm, which given a rational vector $x \in \mathbb{R}^n$, and a rational number $\varepsilon > 0$, returns an ε approximation of $f(x)$.*

The terminology is inspired by the notion of tightness from measure theory (analogous to condition 1 in Definition 4.31).

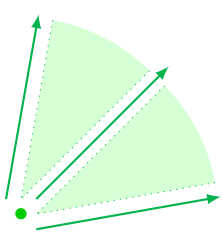
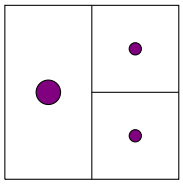
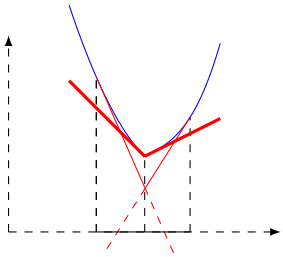
We shall need a classical result on the numerical approximation of multidimensional integrals, which can be found in [DR84]. The *total variation in the sense of Hardy and Krause*, $\|f\|_{\mathrm{BVHK}}$, of a function f on a n dimensional hypercube is defined in [DR84, Def. p.352]. In particular, if f is of regularity class $\mathcal{C}^n$, $\|f\|_{\mathrm{BVHK}}$ is finite. The error made when approximating the integral of a function of n variables by its Riemann sum taken on a regular grid with k points is bounded by $(n\|f\|_{\mathrm{BVHK}})/k^{1/n}$, see the theorem on p 352 of [DR84].

Proposition 4.32. *Suppose that a cost distribution* $\mathbf{c}$ *admits a density function* $f : \mathbb{R}^n \to \mathbb{R}_+$*, that is such that the function* $(1 + \|\cdot\|)f$ *is combinatorially tight and that it has a finite total variation in the sense of Hardy and Krause, bounded by an a priori constant. Suppose that the dimension* n *is fixed. Then,* $\mathbf{c}$ *admits a polynomial-time weak cone valuation oracle.*

Proof. Given a rational cone N, we need to approximate the integrals $\int_N f(c)dc$ and $\int_N cf(c)dc$, up to the precision ε. Using the tightness condition, it suffices to approximate the integrals of the same functions restricted to the domain $N_r := N \cap B_\infty(0, r)$, where $B_\infty(0, r)$ denotes the sup-norm ball of radius r, and the encoding length of r is polynomially bounded in the encoding length of ε. We only discuss the approximation of $\int_{N_r} cf(c)dc$ (the case of $\int_{N_r} f(c)dc$ being simpler). We denote by $\widetilde{c}_{N_r}$ the approximation of $\int_{N_r} cf(c)dc$ provided by taking the Riemann sum of the function $c \mapsto cf(c)$ over the grid $([-r, r))^n \cap ((r/M)\mathbb{Z})^n$, which has $(2M)^r$ points. Then, setting $g := (1 + \|\cdot\|)f$, it follows from the result [DR84, Th. p 352] recalled above that $\|\int_{N_r} cf(c)dc - \widetilde{c}_{N_r}\| \leqslant n\|g\|_{\mathrm{BVHK}}/(2M)$. Hence, for a fixed dimension n, we can get an ε approximation of $\int_N cf(c)dc$ in a time polynomial in the encoding length of ε. □

Remark 4.33. *Proposition 4.32 and Corollary 4.30 entail that, under the previous fixed-parameter restrictions (including dimensions of the recourse spaces), the MSLP problem is polynomial-time approximately solvable for a large class of cost distributions. This applies in particular to distributions like Gaussians, which are combinatorially tight. In this case, condition 1 of Definition 4.31, whereas condition 2 follows from the result of [BB88], implying that the exponential function, restricted to the interval* $(-\infty, 0]$*, can be approximated in polynomial time.*

5 Generalized adaptive partition-based method for two stage stochastic linear problems

Contents

The content of this chapter can be found in the paper [FL22b] published in Operations Research Letters.
$\star$ refers to a section which was not in [FL22b], but which can be found as the section 2 of the preprint [FL22a].

5.1 Introduction

In this chapter, we show a local exact quantization result for 2SLP with generally distributed constraints and fixed recourse. We present this local exact quantization through the notion of adapted partition. This allows us to present an explicit geometric oracle to construct adapted partition and to derive a necessary and sufficient condition for a partition to be adapted even for non-finitely supported distributions. The content of this chapter, without Section 5.5, was published in Operation Research Letters [FL22b] and Section 5.5 can be found in section 3 of the preprint [FL22a].

5.1.1 Setting

We consider the following 2-stage stochastic linear problem with fixed recourse:

$$\min_{x\in\mathbb{R}^n_+} \Big\{ c^\top x + \underbrace{\mathbb{E}\big[Q(x,\boldsymbol{\xi})\big]}_{:=V(x)} \mid Ax = b \Big\}, \tag{2SLP}$$

where the expectation is with respect to $\boldsymbol{\xi} = (\boldsymbol{T}, \boldsymbol{h})$ an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in $\Xi \subset \mathbb{R}^{\ell\times n} \times \mathbb{R}^\ell$, and the recourse cost is

$$Q(x,\xi) := \min_{y\in\mathbb{R}^m_+} \Big\{ q^\top y \mid Tx + Wy = h \Big\}. \tag{5.1}$$

The dual formulation of the recourse problem is

$$Q^D(x,\xi) := \max_{\lambda\in\mathbb{R}^\ell} \Big\{ (h - Tx)^\top \lambda \mid W^\top \lambda \leqslant q \Big\}. \tag{5.2}$$

We define

$$X := \{x \in \mathbb{R}^n_+ \mid Ax = b\}, \tag{5.3a}$$

$$D := \{\lambda \in \mathbb{R}^\ell \mid W^\top \lambda \leqslant q\}. \tag{5.3b}$$

In the rest of the chapter, we assume $D \neq \emptyset$ which implies by duality: $Q(x,\xi) = Q^D(x,\xi)$.

For the sake of simplicity, we assume throughout the chapter that we are in a *relatively complete recourse* setting, that is $X \subset \mathrm{dom}(V)$. Most results can be obtained without this assumption if we add feasibility cuts (see Section 5.3.2).

5.1.2 Literature review

Most results for 2SLP with continuous distributions rely on discretizing the distributions. The Sample Average Approximation (SAA) method samples the costs and constraints. It relies on probabilistic results based on uniform laws of large numbers to give statistical guarantees, see [SDR14, Chap. 5] for details. Obtaining an approximation with satisfying guarantees requires a large number of scenarios. Otherwise, when the support of the random variables are simplices, we can leverage convexity inequalities (like Jensen's and Edmundson-Madansky's) or moments inequalities to construct finite scenario trees such that the discretized problem yields upper or lower bound of the continuous one (see *e.g.* [Kuh06, EZ94b]).

In each of these approaches, we solve an approximate version of the stochastic program, with or without guarantees. In any case, the number of scenarios increases the numerical burden of 2SLP.

In order to alleviate the computations, we can use scenario reduction methods. Some are based on heuristics, aiming at matching properties of the underlying distribution (e.g. matching moments), others are based on adequate distances on the scenario tree (see [DGKR03, HR07]). Alternatively, APM methods iteratively solve an aggregated version of 2SLP over a partition of

the uncertainty space by replacing each subset of scenarios by its weighted mean. We say that a partition is *adapted* to a first-stage control x if the aggregated recourse problem has the same optimal objective value as the recourse problem with the original distribution. After solving an aggregated 2SLP, an APM method call a *(adapted) partition oracle* to define a new (adapted) partition at the current first-stage control. APM were first introduced by Song and Luedtke [SL15], who gave a partition oracle designed for fixed recourse 2SLP with finitely supported random variables. Van Ackooj, de Oliveira and Song [vAdOS18] improved the performance of APM by combining it with level decomposition methods ideas. Finally, Ramirez-Pico and Moreno extended the scope of APMs, under the name GAPM, in [RPM21] to problems with continuous distributions for the right-hand side and technology matrix (and fixed recourse cost vector and matrix). They gave a sufficient condition for a partition to be adapted. They also provided adapted partition oracles for some specific problems.

5.1.3 Contributions

The main contributions of the chapter are the following: i) using polyhedral geometry tools we provide a general adapted partition oracle, ii) we give a new necessary and sufficient condition for a partition to be adapted to $\check{x}$ even in the non-finitely supported case, iii) by casting APM methods as accelerated L-Shaped algorithms where tangeant cones are added instead of tangeant planes (affine cuts), we give convergence and complexity results for APM methods.

5.1.4 Structure of the chapter

Section 5.2 presents the APM framework and a necessary and sufficient condition for a partition to be adapted to $\check{x}$. Section 5.3 uses the link between APM and L-Shaped to obtain convergence and complexity results. Finally, Section 5.4 presents numerical results, while Section 5.5 briefly extends GAPM to non-fixed recourse problem.

5.2 General framework and geometric oracle

In this section, we start by presenting a generic framework for APM algorithms, which depends on partition oracle choice.

We proceed by giving a necessary and sufficient condition for a partition oracle to be adapted, and then a geometric adapted partition oracle.

5.2.1 Partition, refinements and APM framework

A partition $\mathcal{P}$ of Ξ is a collection of non-empty pairwise disjoint subsets covering Ξ, i.e. $\cup_{P\in\mathcal{P}} P = \Xi$, $P \cap P' = \emptyset$ and $P \neq \emptyset$ for $P \neq P' \in \mathcal{P}$. Let P be a measurable subset of Ξ. We denote by $\mathbb{E}\big[\,\cdot\,|P\big]$ the conditional expectation $\mathbb{E}\big[\,\cdot\,|\boldsymbol{\xi} \in P\big]$ and $\mathbb{P}[P]$ the probability $\mathbb{P}[\boldsymbol{\xi} \in P]$. We say that two measurable subsets of $E, F \subset \Xi$ are $\mathbb{P}$-*equivalent*, denoted $E \sim_{\mathbb{P}} F$, if and only if they differ by a $\mathbb{P}$-negligeable set

$$E \sim_{\mathbb{P}} F \iff \mathbb{P}[E \cap F] = \mathbb{P}[E] = \mathbb{P}[F], \tag{5.4}$$

similarly we denote

$$E \subset_{\mathbb{P}} F \iff \mathbb{P}[E \cap F] = \mathbb{P}[E]. \tag{5.5}$$

A $\mathbb{P}$-partition of Ξ is the equivalence class of all partitions that are $\mathbb{P}$-equivalent.

Let $\mathcal{P}$ and $\mathcal{R}$ be two $\mathbb{P}$-partitions of Ξ. We say that $\mathcal{P}$ refines $\mathcal{R}$, denoted $\mathcal{P} \preccurlyeq_{\mathbb{P}} \mathcal{R}$, if

$$\forall P \in \mathcal{P},\ \exists R \in \mathcal{R}, \quad P \subset_{\mathbb{P}} R, \tag{5.6}$$

The *common refinement* of $\mathcal{P}$ and $\mathcal{R}$ is given by

$$\mathcal{P} \wedge \mathcal{R} = \{P \cap R \,|\, P \in \mathcal{P}, R \in \mathcal{R}\}. \tag{5.7}$$

Definition 5.1 (Expected recourse cost of partition)**.** *For $\mathcal{P}$ a $\mathbb{P}$-partition of $\Xi \subset \mathbb{R}^{\ell\times m} \times \mathbb{R}^{\ell}$ we define*

$$V_{\mathcal{P}} : x \mapsto \sum_{P\in\mathcal{P}} \mathbb{P}[P] Q(x, \mathbb{E}[\boldsymbol{\xi}|P]). \tag{5.8}$$

Let $\check{x} \in \mathrm{dom}(V)$. We say that a $\mathbb{P}$-partition $\mathcal{P}$ is adapted to $\check{x}$ *if*

- *$V_{\mathcal{P}}$ is valid,* i.e. *$V_{\mathcal{P}}(x) = V(x) := \mathbb{E}\big[Q(x, \boldsymbol{\xi})\big]$ for all $x \in \mathbb{R}^n$.*
- *and $V_{\mathcal{P}}$ is tight at $\check{x}$* i.e. *$V_{\mathcal{P}}(\check{x}) = V(\check{x}) := \mathbb{E}\big[Q(\check{x}, \boldsymbol{\xi})\big]$.*

The following lemma shows that, by convexity, a finer partition yields a larger expected cost-to-go function.

Lemma 5.2. *Let $\mathcal{P}$ and $\mathcal{R}$ two $\mathbb{P}$-partitions of Ξ then*

$$\mathcal{P} \preccurlyeq_{\mathbb{P}} \mathcal{R} \quad \Longrightarrow \quad V_{\mathcal{P}} \geqslant V_{\mathcal{R}}. \tag{5.9}$$

Moreover,

$$V_{\mathcal{P}\wedge\mathcal{R}} \geqslant \max(V_{\mathcal{P}}, V_{\mathcal{R}}). \tag{5.10}$$

Finally,

$$Q(\cdot, \mathbb{E}[\boldsymbol{\xi}]) \leqslant V_{\mathcal{P}} \leqslant V. \tag{5.11}$$

In particular, in this setting with deterministic recourse matrix W and cost q, for all partition $\mathcal{P}$, $V_{\mathcal{P}}$ is valid. We then only have to prove $V_{\mathcal{P}}(\check{x}) = V(\check{x})$ to prove that $\mathcal{P}$ is adapted to $\check{x}$. However, this would not be the case when we will extend to general cost $\boldsymbol{q}$ in Section 5.5.

Proof. Since $\mathcal{P}$ is a $\mathbb{P}$-partition $\mathbb{1}_{\boldsymbol{\xi}\in R} = \sum_{P\in\mathcal{P}} \mathbb{1}_{\boldsymbol{\xi}R\cap P}$ almost surely. For any measurable set $E \subset \Xi$, $\mathbb{E}[\boldsymbol{\xi}\mathbb{1}_{\boldsymbol{\xi}\in E}] = \mathbb{P}[E]\mathbb{E}[\boldsymbol{\xi}|E]$. We then have,

$$\mathbb{P}[R]\mathbb{E}[\boldsymbol{\xi}|R] = \mathbb{E}[\boldsymbol{\xi}\mathbb{1}_{\boldsymbol{\xi}\in R}] = \sum_{P\in\mathcal{P}} \mathbb{E}[\boldsymbol{\xi}\mathbb{1}_{\boldsymbol{\xi}R\cap P}] \tag{5.12}$$

$$= \sum_{P\in\mathcal{P}} \mathbb{P}[R\cap P]\mathbb{E}[\boldsymbol{\xi}|R\cap P] \tag{5.13}$$

When $\mathbb{P}[R] > 0$, by dividing this equation by $\mathbb{P}[R]$, we obtain that $\mathbb{E}[\boldsymbol{\xi}|R]$ is equal to the convex combination $\sum_{P\in\mathcal{P}} \frac{\mathbb{P}[R\cap P]}{\mathbb{P}[R]} \mathbb{E}[\boldsymbol{\xi}|R\cap P]$. Finally, consider $\check{x} \in X$, the convexity of $\xi \mapsto Q(\check{x}, \xi)$ yields

$$Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|R]) \leqslant \sum_{P\in\mathcal{P}} \frac{\mathbb{P}[P\cap R]}{\mathbb{P}[R]} Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P\cap R])\,. \tag{5.14}$$

Then, if $\mathcal{P} \preccurlyeq_{\mathbb{P}} \mathcal{R}$,

$$V_{\mathcal{R}}(\check{x}) = \sum_{R\in\mathcal{R}} \mathbb{P}[R] Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|R]) \tag{5.15a}$$

$$\leqslant \sum_{P\in\mathcal{P}} \sum_{R\in\mathcal{R}} \mathbb{P}[P\cap R] Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P\cap R]) \tag{5.15b}$$

$$= \sum_{P\in\mathcal{P}} \mathbb{P}[P] Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) = V_{\mathcal{P}}(\check{x}) \tag{5.15c}$$

The last line follows from the fact, that for $P \in \mathcal{P}$, with $\mathbb{P}[P > 0]$ and $\mathcal{P} \preccurlyeq_{\mathbb{P}} \mathcal{R}$, there exists a unique $R \in \mathcal{R}$ such that $\mathbb{P}[P \cap R] = \mathbb{P}[P]$, all other $R \in \mathcal{R}$ being such that $\mathbb{P}[P \cap R] = 0$.

Eq. (5.10) is a direct consequence of $V_{\mathcal{P}\wedge\mathcal{P}'} \geqslant V_{\mathcal{P}}$ and $V_{\mathcal{P}\wedge\mathcal{P}'} \geqslant V_{\mathcal{P}'}$. Thus, $V_{\mathcal{P}\wedge\mathcal{P}'} \geqslant \max(V_{\mathcal{P}}, V_{\mathcal{P}'})$. Coupling this result with $\mathcal{P} \preccurlyeq_{\mathbb{P}} \{\Xi\}$ yields the left inequality of Eq. (5.11) while the other can be found in [RPM21, Prop. 1]. □

With those definitions we present in Algorithm 5.1 a generic framework for APM methods.

$k \leftarrow 0$, $z_U^0 \leftarrow +\infty$, $z_L^0 \leftarrow -\infty$, $\mathcal{P}^0 \leftarrow \{\Xi\}$;
while $z_U^k - z_L^k > \varepsilon$ **do**
  $k \leftarrow k+1$;
  Solve $z_L^k \leftarrow \min_{x\in X} c^\top x + V_{\mathcal{P}^{k-1}}(x)$ and let x^k be an optimal solution ;
  Call the oracle on x^k yielding $\mathcal{P}_{x^k}$;
  $\mathcal{P}^k \leftarrow \mathcal{P}^{k-1} \wedge \mathcal{P}_{x^k}$;
  $z_U^k \leftarrow \min\left(z_U^{k-1}, c^\top x^k + V_{\mathcal{P}^k}(x^k)\right)$;
end

Algorithm 5.1: Generic framework for APM.

5.2.2 Coarsest adapted partition

In this section, we define $\mathcal{R}_{\check{x}}$, a particular $\mathbb{P}$-partition, and prove that it is, in a generic case, the coarsest partition adapted to $\check{x} \in X$, *i.e.*, the only partition adapted to $\check{x}$ that refines $\mathcal{R}_{\check{x}}$ is $\mathcal{R}_{\check{x}}$ itself. Indeed, we are looking for partitions that yield a precise approximation of recourse cost (exact at $\check{x}$ in the adapted case), while having the smallest possible number of elements.

When the distributions have finite support, [SL15] characterized the partitions adapted to $\check{x}$. Building on this result, a sufficient condition for continuous distribution can be found in [RPM21, Prop. 2]. We now prove that, for any distribution, a partition is adapted to $\check{x}$ if and only if it refines the collection $\overline{\mathcal{R}}_{\check{x}}$ defined in (5.17b). Unfortunately, $\overline{\mathcal{R}}_{\check{x}}$ is not necessarily a $\mathbb{P}$-partition, thus we also provide a partition $\mathcal{R}_{\check{x}} \preccurlyeq \overline{\mathcal{R}}_{\check{x}}$ (see Figure 5.2 for an illustration).

Recall that $D = \{\lambda \in \mathbb{R}^\ell \mid W^\top \lambda \leqslant q\}$ and that the normal cone of D at λ is the set $N_D(\lambda) := \{\psi \in \mathbb{R}^\ell \,|\, \psi^\top(\lambda' - \lambda) \leqslant 0, \forall \lambda' \in D\}$. We denote by $\mathrm{ri}(N)$ the relative interior of a cone N. Let $\mathcal{N}(D) := \{N_D(\lambda) \,|\, \lambda \in D\}$ be the normal fan of D, *i.e.*, the (finite) collection of all normal cones of D (see Chapter 3). We denote by $\mathcal{N}(D)^{\max} := \{N \in \mathcal{N}(D) \,|\, \forall N' \in \mathcal{N}(D), N \subset N' \Rightarrow N = N'\}$ the collection of the maximal elements of $\mathcal{N}(D)$ (*i.e.*, full dimensional cones up-to lineality spaces).

Theorem 5.3. *Fix $\check{x} \in \mathrm{dom}(V)$ and N a cone in $\mathbb{R}^m$. We define $E_{N,\check{x}}$ and $\overline{E}_{N,\check{x}}$, subsets of Ξ, as*

$$E_{N,\check{x}} := \{\xi \in \Xi \mid h - T\check{x} \in \mathrm{ri}(N)\} \tag{5.16a}$$
$$\overline{E}_{N,\check{x}} := \{\xi \in \Xi \mid h - T\check{x} \in N\} \tag{5.16b}$$

We define $\mathcal{R}_{\check{x}}$ and $\overline{\mathcal{R}}_{\check{x}}$ as

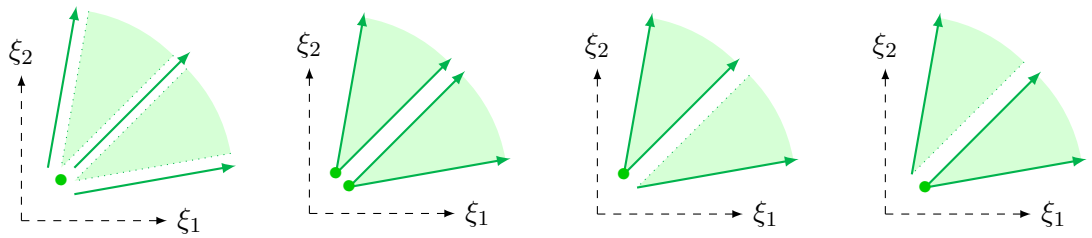
$$\mathcal{R}_{\check{x}} := \{E_{N,\check{x}} \mid N \in \mathcal{N}(D)\} \tag{5.17a}$$
$$\overline{\mathcal{R}}_{\check{x}} := \{\overline{E}_{N,\check{x}} \mid N \in \mathcal{N}(D)^{\max}\}. \tag{5.17b}$$

Then,

$$\mathcal{P} \preccurlyeq_{\mathbb{P}} \mathcal{R}_{\check{x}} \Longrightarrow V_{\mathcal{P}}(\check{x}) = V(\check{x}) \tag{5.18a}$$
$$\mathcal{P} \preccurlyeq_{\mathbb{P}} \overline{\mathcal{R}}_{\check{x}} \iff V_{\mathcal{P}}(\check{x}) = V(\check{x}). \tag{5.18b}$$

Remark 5.4. *When the distribution of $\boldsymbol{\xi}$ is absolutely continuous with respect to the Lebesgue measure of Ξ, $\mathcal{R}_{\check{x}} \sim_{\mathbb{P}} \overline{\mathcal{R}}_{\check{x}}$, thus $\mathcal{R}_{\check{x}}$ is the coarsest partition adapted to $\check{x} \in \mathrm{dom}(V)$.*

If $\boldsymbol{\xi}$ does not admit a density, $\mathcal{R}_{\check{x}}$ is still an adapted partition but not necessarily the coarsest, which might not exist (see Fig. 5.2). Nevertheless, any adapted partition should refine $\overline{\mathcal{R}}_{\check{x}}$. Unfortunately, we cannot use $\overline{\mathcal{R}}_{\check{x}}$ in Algorithm 5.1, as we cannot guarantee that $\overline{\mathcal{R}}_{\check{x}}$ is a $\mathbb{P}$-partition.



(a) $\mathcal{R}_{\check{x}}$ (b) $\overline{\mathcal{R}}_{\check{x}}$ (c) $\mathcal{P}$ (d) $\mathcal{P}'$

Figure 5.2: $\mathcal{R}_{\check{x}}$ is a partition of Ξ into 6 elements, $\overline{\mathcal{R}}_{\check{x}}$ is not a partition, $\mathcal{P}$ and $\mathcal{P}'$ are two distinct coarsest partitions (into 2 elements) with $\mathcal{R}_{\check{x}} \preccurlyeq \mathcal{P} \preccurlyeq \overline{\mathcal{R}}_{\check{x}}$ and $\mathcal{R}_{\check{x}} \preccurlyeq \mathcal{P}' \preccurlyeq \overline{\mathcal{R}}_{\check{x}}$.

Remark 5.5. *Note that Proposition 2 of [RPM21] implies that all partition oracle returning partitions satisfying assumption (7) of [RPM21] must be refinements of $\overline{\mathcal{R}}_{\check{x}}$ by Theorem 5.3. In the finite scenario case, our adaptedness condition is equivalent to Song and Luedtke's condition [SL15].*

We preclude the proof by a technical lemma.

Lemma 5.6. *Consider a set $P \subset \Xi$ such that $\mathbb{P}(P) > 0$, and a first-stage control $\check{x} \in \mathrm{dom}(V)$. Then,*

$$\exists R \in \mathcal{R}_{\check{x}}, \quad P \subset_{\mathbb{P}} R \tag{5.19}$$
$$\implies Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) = \mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P], \tag{5.20}$$
$$\exists \overline{R} \in \overline{\mathcal{R}}_{\check{x}}, \quad P \subset_{\mathbb{P}} \overline{R} \tag{5.21}$$
$$\iff Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) = \mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P]. \tag{5.22}$$

Proof. Since $\exists R \in \mathcal{R}_{\check{x}},\ P \subset_{\mathbb{P}} R$ implies $\exists \overline{R} \in \overline{\mathcal{R}}_{\check{x}},\ P \subset_{\mathbb{P}} \overline{R}$, we only need to prove the second equivalence.

($\Rightarrow$) Let P be such that there exists $N \in \mathcal{N}(D)$ with $P \subset_{\mathbb{P}} \overline{E}_{N,\check{x}}$. By definition of $\mathcal{N}(D)$, there exists a dual point $\lambda_N \in D$ such that N is the normal cone of D at λ_N. By definition of a normal cone, for all $\psi \in N$ and all $\lambda \in D$, $\psi^\top(\lambda - \lambda_N) \leqslant 0$. In other words, $\psi^\top \lambda_N = \max_{\lambda \in D} \psi^\top \lambda$.

As $P \subset \overline{E}_{N,\check{x}}$, for $\mathbb{P}$-almost-all $\xi \in P$, we have $h - T\check{x} \in N$. Recall that $Q(\check{x}, \xi) = \sup_{\lambda \in D}(h - T\check{x})^\top \lambda$, thus, $Q(\check{x}, \xi) = (h - T\check{x})^\top \lambda_N$. Hence,

$$\mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P] = \mathbb{E}[(\boldsymbol{h} - \boldsymbol{T}\check{x})^\top \lambda_N | P] \tag{5.23a}$$
$$= \mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P]^\top \lambda_N = Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) \tag{5.23b}$$

as N is convex and $\mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P] \in N$.

($\Leftarrow$) For $\psi \in \mathbb{R}^\ell$, we denote the face $D^\psi := \mathrm{argmax}_{\lambda \in D} \psi^\top \lambda$. Note that, for all $\psi, \psi' \in \mathrm{ri}(N)$, with $N \in \mathcal{N}(D)$, we have $D^N := D^\psi = D^{\psi'}$.

Assume that there is no $R \in \overline{\mathcal{R}}_{\check{x}}$ such that $P \subset_{\mathbb{P}} R$. Then, for all $R \in \overline{\mathcal{R}}_{\check{x}}$, $\mathbb{P}[P \cap R] < \mathbb{P}[P]$. Since $\mathbb{P}[P] \leqslant \sum_{R \in \overline{\mathcal{R}}_{\check{x}}} \mathbb{P}[P \cap R]$, there exist R_1 and R_2 in $\overline{\mathcal{R}}_{\check{x}}$ such that $\mathbb{P}[P \cap R_1] > 0$ and $\mathbb{P}[P \cap R_2] > 0$. Let $\lambda \in D$ such that $Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) = \mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P]^\top \lambda$ *i.e.*, $\lambda \in D^{\mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P]}$. Let N_1 and $N_2 \in \mathcal{N}(D)^{\max}$ be such that $R_1 = \overline{E}_{N_1,\check{x}}$ and $R_2 = \overline{E}_{N_2,\check{x}}$. Since $N_1 \neq N_2$ are maximal, $D^{N_1} \cap D^{N_2} = \emptyset$. Thus, there exists at least one $i \in \{1, 2\}$ such that $\lambda \notin D^{N_i}$. Then, $\mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P \cap R_i] > \mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P \cap R_i]^\top \lambda$.

Note that $Q(\check{x}, \xi) = \sigma_D(h - T\check{x})$, where σ_D is the support function of the polyhedron D, thus $\xi \mapsto Q(\check{x}, \xi)$ is a polyhedral function. Further, its affine regions are the elements of $\overline{\mathcal{R}}_{\check{x}}$.

By convexity, for any measurable set A, $\mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P \cap A] \geqslant Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P \cap A])$ which is equal to $\max_{\lambda' \in D} \mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P \cap A]^\top \lambda'$. Since $\lambda \in D$, we have $\mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P \cap A] \geqslant \mathbb{E}[\boldsymbol{h} - \boldsymbol{T}\check{x}|P \cap A]^\top \lambda$.

Thus, $\mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P] > Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P])$. □

Proof of Theorem 5.3. By definition $\mathcal{P} \preccurlyeq_{\mathbb{P}} \overline{\mathcal{R}}_{\check{x}}$, if and only if, for all $P \in \mathcal{P}$ there exists a cell $R \in \overline{\mathcal{R}}_{\check{x}}$ such that $P \subset_{\mathbb{P}} R$. By Lemma 5.6 this is equivalent to, for all $P \in \mathcal{P}$, $Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) = \mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P]$. Now, by Jensen's inequality, this equality (for all $P \in \mathcal{P}$) is equivalent to the equality of a convex sum like

$$\sum_{P\in\mathcal{P}} Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P])\mathbb{P}[P] = \sum_{P\in\mathcal{P}} \mathbb{E}[Q(\check{x}, \boldsymbol{\xi})|P]\mathbb{P}[P].$$

Law of total expectation yields (5.18). □

Remark 5.7. *Let $x^\star$ be an optimal solution of*

$$\min_{x\in X} \quad c^\top x + V_{\mathcal{P}^\star}(x) \tag{5.24}$$

where $\mathcal{P}^\star \preccurlyeq_{\mathbb{P}} \mathcal{R}_{x^\star}$. Then, $x^\star$ is also a solution of Problem (2SLP)*. In other words, $\mathcal{P}^\star$ is a* 0*-sufficient partition according to [SL15, Def. 1.2].*

5.3 Comparison with other algorithms and convergence

In this section, we show that the partition-based methods can be seen as an acceleration of the cutting plane method. It then gives us a finite convergence proof with a bound on the number of steps.

5.3.1 Adapted partition and subdifferential

We show that, for any first stage control $x \in X$, if the partition is adapted to x, then the subdifferential of approximate expected recourse cost coincides with the subdifferential of the true expected recourse cost.

Lemma 5.8. *Let $\check{x} \in \mathrm{dom}(V)$ and $\mathcal{P}$ be a refinement of $\mathcal{R}_{\check{x}}$, i.e. $\mathcal{P} \preccurlyeq \mathcal{R}_{\check{x}}$, then*

$$\partial V_{\mathcal{R}_{\check{x}}}(\check{x}) \subset \partial V_{\mathcal{P}}(\check{x}) \subset \partial V(\check{x}) \tag{5.25}$$

Furthermore, if $\check{x} \in \mathrm{ri}(\mathrm{dom}(V))$,

$$\partial V_{\mathcal{R}_{\check{x}}}(\check{x}) = \partial V_{\mathcal{P}}(\check{x}) = \partial V(\check{x}) \tag{5.26}$$

Proof. Let $g \in \partial V_{\mathcal{R}_{\check{x}}}(\check{x})$ then for all x, $V_{\mathcal{R}_{\check{x}}}(x) \geqslant V_{\mathcal{R}_{\check{x}}}(\check{x}) + g^\top(x - \check{x})$. By monotonicity (see (5.9)) $V_{\mathcal{P}}(x) \geqslant V_{\mathcal{R}_{\check{x}}}(x)$ and as $\mathcal{R}_{\check{x}}$ is adapted to $\check{x}$, we have $V_{\mathcal{R}_{\check{x}}}(\check{x}) = V(\check{x}) = V_{\mathcal{P}}(\check{x})$. Thus, $V_{\mathcal{P}}(x) \geqslant V_{\mathcal{P}}(\check{x}) + g^\top(x - \check{x})$ and $g \in \partial V_{\mathcal{P}}(\check{x})$. The proof for the second inclusion is similar.

Let $\check{x} \in \mathrm{ri}(\mathrm{dom}(V))$, we now prove that $\partial V_{\mathcal{R}_{\check{x}}}(\check{x}) = \partial V(\check{x})$. Recall that $D^N = D^\psi = \mathrm{argmax}_{\lambda\in D}\, \psi^\top\lambda$, for $\psi \in \mathrm{ri}(N)$ where $N \in \mathcal{N}(D)$. By [SDR14, Prop 2.8 p.37], $\partial V(\check{x}) = \mathbb{E}[-\boldsymbol{T}^\top D^{\boldsymbol{h}-\boldsymbol{T}\check{x}}] + N_{\mathrm{dom}(V)}(\check{x})$. Thus, since $\check{x} \in \mathrm{ri}(\mathrm{dom}(V))$,

$$\partial V(\check{x}) = \mathbb{E}[-\boldsymbol{T}^\top D^{\boldsymbol{h}-\boldsymbol{T}\check{x}}] \tag{5.27a}$$

$$= \mathbb{E}[\sum_{N\in\mathcal{N}(D)} -\mathbb{1}_{\boldsymbol{h}-\boldsymbol{T}\check{x}\in\mathrm{ri}(N)}\boldsymbol{T}^\top D^N] \tag{5.27b}$$

$$= \mathbb{E}[\sum_{N\in\mathcal{N}(D)} -\mathbb{1}_{\boldsymbol{\xi}\in E_{N,\check{x}}}\boldsymbol{T}^\top D^N] \tag{5.27c}$$

Further,

$$\mathbb{E}\big[\mathbb{1}_{\boldsymbol{\xi}\in E_{N,\check{x}}}\boldsymbol{T}^\top D^N\big] \tag{5.27d}$$
$$= \mathbb{P}\big[E_{N,\check{x}}\big]\mathbb{E}\big[\boldsymbol{T}|E_{N,\check{x}}\big]^\top D^N \tag{5.27e}$$
$$= \mathbb{P}\big[E_{N,\check{x}}\big]\mathbb{E}\big[\boldsymbol{T}|E_{N,\check{x}}\big]^\top D^{\mathbb{E}[\boldsymbol{h}-\boldsymbol{T}\check{x}|E_{N,\check{x}}]} \tag{5.27f}$$
$$\tag{5.27g}$$

And by definition of $\mathcal{R}_{\check{x}}$ in (5.17a) , we get

$$\partial V(\check{x}) = \sum_{P\in\mathcal{R}_{\check{x}}} -\mathbb{P}[P]\mathbb{E}\big[\boldsymbol{T}|P\big]^\top D^{\mathbb{E}[\boldsymbol{h}-\boldsymbol{T}\check{x}|P]} \tag{5.27h}$$
$$= \sum_{P\in\mathcal{R}_{\check{x}}} \mathbb{P}[P]\partial_x Q(\check{x}, \mathbb{E}[\boldsymbol{\xi}|P]) \tag{5.27i}$$
$$= \partial V_{\mathcal{R}_{\check{x}}}(\check{x}) \tag{5.27j}$$

□

5.3.2 Link with L-shaped and Benders decomposition

The classical L-shaped method (see e.g. [BL11, Chapter 5]) is a specification of Benders decomposition to 2SLP with finitely supported distributions. The core idea consists in representing the expected recourse cost in (2SLP), by a lift variable

$$\min_{x\in X,\theta\in\mathbb{R}} \Big\{c^\top x + \theta \mid (x,\theta)\in \operatorname{epi}(V)\Big\}. \tag{5.28}$$

We then relax the epigraphical representation $(x,\theta)\in\operatorname{epi}(V)$, replacing it by a set of valid inequalities called *cuts*, i.e.

$$\min_{x\in X,\theta\in\mathbb{R}} \quad c^\top x + \theta \tag{5.29a}$$
$$\text{s.t.} \quad g^\top x + v \leqslant \theta, \qquad \forall (g,v)\in\mathcal{O}, \tag{5.29b}$$
$$f^\top x \leqslant \overline{f}, \qquad \forall (f,\overline{f})\in\mathcal{F}. \tag{5.29c}$$

More precisely, assume that we have such a relaxation of (2SLP). Let x^k be an optimal first stage control of this relaxation. If it is admissible, meaning that for all scenario ξ there exists an admissible recourse control y_ξ, we compute, through duality, a subgradient $g^k\in\partial V(x^k)$. This yields a new *optimality cut* $\theta \geqslant (g^k)^\top(x-x^k)+V(x^k)$, which is added to $\mathcal{O}$. If x^k is not admissible we can add a *feasibility cut* to $\mathcal{F}$ instead by using dual optimal extreme ray (see [BL11, §5.1.b]). We then solve our strengthened relaxation to obtain x^{k+1}.

The L-Shaped method specifies that the subgradient g^k can be obtained as an average over ξ of subgradients $g^{k,\xi}\in\partial_x Q(x^k,\xi)$. In particular, it means that, to compute the subgradient, we can solve $|\operatorname{supp}(\boldsymbol{\xi})|$ smaller LP instead of a large one.

Remark 5.9 (L-shaped for continuous distribution)**.** *When the distribution are non-finitely supported, we cannot apply naively this method as there is a non-finite number of scenarios. Nevertheless, we can still approximate* $\operatorname{epi}(V)$ *with cuts. We can compute* $\theta = V_{\mathcal{R}_{\check{x}}}(\check{x})$ *and a subgradient* $g\in\partial V_{\mathcal{R}_{\check{x}}}(\check{x})$ *by solving* $|\mathcal{R}_{\check{x}}|$ *linear problems of the form* (5.2) *through exact quantization. By Theorem 5.3,* $\theta = V_{\mathcal{R}_{\check{x}}}(\check{x}) = V(\check{x})$*. Further,* $g\in\partial V_{\mathcal{R}_{\check{x}}}(\check{x})\subset\partial V(\check{x})$ *by Lemma 5.8. Then* (θ,g) *define an optimality cut.*

Lemma 5.8 shows that, at each step k of Algorithm 5.1, we add a collection of valid cuts which are exact at x^k to our collection of cuts. This means that APM methods can be seen as a Bender's decomposition method where we add more than one exact cut per iteration. In particular, when $x^k\in\operatorname{ri}\big(\operatorname{dom}(V)\big)$ we add the whole tangent cone of $\operatorname{epi}(V)$ at x instead of a single cut.

5.3.3 Convergence of APMs

We start by showing that the bounds generated in Algorithm 5.1 are monotonic.

Lemma 5.10. *Assume that the partition oracle used is adapted. For every computed step k we have*

$$z_L^{k-1} \leqslant z_L^k \leqslant \operatorname{val}(2\text{SLP}) \leqslant z_U^k \leqslant z_U^{k-1} \tag{5.30}$$

Proof. Since $\mathcal{P}^k \preccurlyeq_{\mathbb{P}} \mathcal{P}^{k-1}$, by Lemma 5.2, we have, for all $x \in X$,

$$\begin{aligned} c^\top x + V_{\mathcal{P}^{k-1}}(x) &\leqslant c^\top x + V_{\mathcal{P}^k}(x) && \text{(5.31a)} \\ &\leqslant c^\top x + V(x) && \text{(5.31b)} \end{aligned}$$

Minimizing over x yields $z_L^{k-1} \leqslant z_L^k \leqslant \operatorname{val}(2\text{SLP})$. For any k, we have that $\mathcal{P}^k$ is adapted to $x^k \in X$, hence $V_{\mathcal{P}^k}(x^k) = V(x^k)$, thus $\operatorname{val}(2\text{SLP}) \leqslant c^\top x_k + V_{\mathcal{P}^k}(x^k)$. Further, by definition of z_k^U in Algorithm 5.1, $z_k^U = \min_{\kappa \leqslant k} c^\top x^\kappa + V_{\mathcal{P}^\kappa}(x^\kappa)$, yielding $\operatorname{val}(2\text{SLP}) \leqslant z_U^k \leqslant z_U^{k-1}$. □

We now prove finite convergence of any APM.

Theorem 5.11. *Assume that the partition oracle used is adapted. If $X \subset \mathbb{R}^n_+$ has a finite diameter $M \in \mathbb{R}_+$ and $x \mapsto c^\top x + V(x)$ is Lipschitz with constant L then the partition based Algorithm 5.1 finds an ε-solution in at most $\left(\frac{LM}{\varepsilon} + 1\right)^n$ iterations.*

Proof. We adapt the classical proof of Kelley's cutting plane algorithm to APMs. Let $k \in \mathbb{N}$ and $1 < i < k$, we have that $V(x_i) = V_{\mathcal{P}^{k-1}}(x_i) = V_{\mathcal{P}^i}(x_i)$. Let $g \in \partial V_{\mathcal{P}^{k-1}}(x_i) \subset \partial V(x_i)$ such that $\|c + g\|$ is bounded by the Lipschitz constant L then

$$\begin{aligned} z_U^k - z_L^k &\leqslant c^\top x^i + V_{\mathcal{P}^i}(x^i) - (c^\top x^k + V_{\mathcal{P}^{k-1}}(x^k)) && \text{(5.32a)} \\ &= c^\top (x^i - x^k) + V_{\mathcal{P}^{k-1}}(x^i) - V_{\mathcal{P}^{k-1}}(x^k)) && \text{(5.32b)} \\ &\leqslant c^\top (x^i - x^k) - g^\top (x^k - x^i) && \text{(5.32c)} \\ &\leqslant \|c + g\|_2 \|x^i - x^k\|_2 \leqslant L \|x^i - x^k\|_2. && \text{(5.32d)} \end{aligned}$$

Then, for k such that, $\varepsilon < z_k^U - z_k^L$, we have $\varepsilon < L\|x^i - x^k\|$, in particular $\|x^i - x^k\| \geqslant \varepsilon / L$. By definition of M there are at most $\left(\frac{LM}{\varepsilon} + 1\right)^n$ balls of radius ε/L in X. An ε-solution being obtained as soon as two points are in the same ball. □

5.4 Numerical examples

In this section, we detail the actual computation required by Algorithm 5.1 and illustrate the algorithm on numerical examples.

5.4.1 Detailing computation

In the following two sections, we give more details on how to compute the Lines 4 to 7 of Algorithm 5.1.

5.4.1.1 Master problem and subproblems

Once $\mathbb{E}[\boldsymbol{\xi} \,|\, P]$ and $\mathbb{P}[P]$ have been computed for $P \in \mathcal{P}^{k-1}$, by Eq. (5.8) and Eq. (5.1), the problem of Line 4 is reduced to the following linear problem

$$\min_{x \in X, (y_P) \in (\mathbb{R}^m_+)^{\mathcal{P}^{k-1}}} c^\top x + \sum_{P \in \mathcal{P}^k} \mathbb{P}[P] q^\top y_P \tag{5.33a}$$

$$\text{s.t.} \quad \mathbb{E}[\boldsymbol{T}|P]x + W y_P = \mathbb{E}[\boldsymbol{h}|P] \quad \forall P \in \mathcal{P}^k. \tag{5.33b}$$

Moreover, to compute the upper bound in Line 7, we need to solve at most $|\mathcal{P}^k|$ linear problems of dimension m

$$Q(x^k, \mathbb{E}[\boldsymbol{\xi}|P]) := \min_{y_P \in \mathbb{R}^m_+} q^\top y_P \tag{5.34a}$$

$$\text{s.t.} \quad \mathbb{E}[\boldsymbol{T}|P]x^k + W y_P = \mathbb{E}[\boldsymbol{h}|P] \tag{5.34b}$$

5.4.1.2 Refinement, expectation and probabilities

Recall that we can store a polyhedron E, either as a family of constraints (M, β) such that $E = \{x \in \mathbb{R}^d \,|\, Mx \leqslant \beta\}$ (H-representation) or as families of vertices $(v_i)_{i \in I}$ and rays $(r_j)_{j \in J}$ such that $E = \mathrm{Conv}(v_i)_{i \in I} + \mathrm{Cone}(r_j)_{j \in J}$ (V-representation), see Section 3.1. Both representation are implemented polymake, an open source software and julia library [GJ00]. We can switch between representations through algorithms such as the *double description* [FP96].

We can simultaneously compute conditional expectations, probabilities and refinement as detailed in Algorithm 5.2.

Data: $\mathcal{P}^{k-1}$ and $\mathcal{R}_{x^k}$ the partition to refine, second stage distributions $\boldsymbol{T}$ and $\boldsymbol{h}$.
1 Set $\mathcal{P}^k := \emptyset$;
2 **for** $P \in \mathcal{P}^{k-1}$ *and* $R \in \mathcal{R}_{x^k}$ **do**
3 Set $P' := P \cap R$;
4 **if** $\mathbb{P}[P'] > 0$ **then**
5 Store $\mathbb{P}[P']$, $\mathbb{E}[\boldsymbol{T}|P']$ and $\mathbb{E}[\boldsymbol{h}|P']$;
6 Set $\mathcal{P}^k := \mathcal{P}^k \cup \{P'\}$;
7 **end**
8 **end**

Algorithm 5.2: Refinement procedure.

In this algorithm, the computation of probabilities on polyhedra in Line 5 is a $\sharp P$-complete problem in the general case, although, for a large class of distributions, formulas exists (see Appendix B).

5.4.1.3 Explicit partition oracle

In this section, we explain how to compute, for $\check{x} \in X$, $\mathcal{R}_{\check{x}} = \{E_{N,\check{x}} \,|\, N \in \mathcal{N}(D)\}$ where $E_{N,\check{x}} = \{(T, h) \,|\, h - T\check{x} \in \mathrm{ri}(N)\}$.

The computation of the normal fan $\mathcal{N}(D)$, already implemented in polymake, can be done thanks to a double description and active constraint sets. Note that if $N \in \mathcal{N}(D)$, then $E_{N,\check{x}}$ is a relatively open polyhedral cone of Ξ. In particular, if $N := \{\psi \,|\, M\psi \leqslant 0\}$ is given in a non-redundant H-representation where $M \in \mathbb{R}^{p \times l}$, we have $\mathrm{ri}(N) = \{\psi \,|\, M\psi \ll 0\}$. Then $E_{N,\check{x}} = \{\xi \in \Xi \,|\, H^x \xi \ll 0\}$, with $H^x = (-x_1 M \cdots - x_n M \; M)$.

Unfortunately, obtaining an H-representation of the normal cone, from the usual V-representation, requires a double-description which is numerically intractable in large dimension (see McMullen bounds [McM70]).

The double-description can be avoided if the technology matrix $\boldsymbol{T} \equiv T$ is fixed. Indeed, in this case $E_{N,\check{x}} \sim_{\mathbb{P}} \{T\} \times \big(T\check{x} + \mathrm{ri}(N)\big)$. Thus, we can compute, at the beginning of the algorithm, a V-representation of all $N \in \mathcal{N}(D)$, and easily deduce a V-representation of $E_{N,\check{x}}$ by adding Tx to each representant ray.

5.4.2 Numerical examples

We applied Algorithm 5.1with our geometric oracle to the problems LanDs and CV@R of [RPM21]. We obtained the same partition, and thus the same numerical results. Finally, we treat the problem Prod-Mix for which no partition oracle were given in the literature. Our code is available at `https://github.com/maelforcier/GAPM`.

5.4.2.1 Energy planing problem - LandS

We applied numerically our method to the LandS problem and constated that our geometric oracle returned the same partition as [RPM21].

5.4.2.2 Conditional value-at-risk linear problems

For the conditional value-at-risk problem in [RPM21], note that our geometric oracle yields the same partition:

$$Q^D(\check{x}, \xi) := \quad \max_{\lambda \in \mathbb{R}} \quad (-\check{x}^\top r^\xi - \tau)\lambda \tag{5.35a}$$

$$\text{s.t.} \quad 0 \leqslant \lambda \leqslant 1 \tag{5.35b}$$

Here $D = [0,1]$ and $\mathcal{N}(D) = \{\mathbb{R}^-, \{0\}, \mathbb{R}^+\}$ Then, if $\check{x} \neq 0$, $\mathcal{R}_{\check{x}} = \{\{r | \check{x}^\top r > -\tau\}, \{r | \check{x}^\top r = -\tau\}, \{r | \check{x}^\top r < -\tau\}\}$.

5.4.2.3 Prod-Mix

We adapted the problem Prod-mix of `https://stoprog.org/SavedLinks/IBM_StoExt_problems/node4.php` as

$$\min_{x, \boldsymbol{y}} \quad -c^\top x + \mathbb{E}\big[q^\top \boldsymbol{y}\big] \tag{5.36}$$

$$\text{s.t.} \quad \boldsymbol{T}x - \boldsymbol{y} \leqslant \boldsymbol{h} \tag{5.37}$$

$$x, \boldsymbol{y} \geqslant 0, \tag{5.38}$$

where $q^\top = (5, 10)$, $c^\top = (12, 40)$, $\boldsymbol{T}$ follows the uniform law $\begin{pmatrix} \mathcal{U}[3.5, 4.5] & \mathcal{U}[9, 11] \\ \mathcal{U}[0.8, 1.2] & \mathcal{U}[36, 44] \end{pmatrix}$ and $\boldsymbol{h}^\top$ follows the uniform distribution $\big(\mathcal{U}[5970, 6030],\ \mathcal{U}[3979, 4021]\big)$. Algorithm 5.1 gave the results summed up in Table 5.1

To compare our approach with SAA, we solved the same problem 100 times, each with $10'000$ scenarios randomly drawn, yielding a 95% confidence interval centered in -17711, with length 4.4. This statistical confidence interval required 2058s of computation. By APM, an exact gap smaller than 4.4 is obtained after iteration 6, that is in 23s, which is here roughly the time required for solving one SAA. Thus, Algorithm 5.1 can be useful to find accurate values.

The most time-consuming parts of the algorithm are the computations of volumes which take 85% of the total time, because polymake only implement exact computations, which was proven to be $\sharp P$-complete [DF88]. To improve Algorithm 5.1, we could use precise rapid approximation volume algorithms, see e.g. [CV16].

k	z_L^k	z_U^k	$z_U^k - z_L^k$	Total time	$\lvert\mathcal{P}^k\rvert$
1	-18666.67	-16939.71	1726.96	0.57 s	4
2	-17873.01	-17383.73	489.28	2.1 s	9
4	-17744.67	-17709.00	35.67	9.1 s	25
6	-17713.74	-17711.37	2.37	23.7 s	49
8	-17711.71	-17711.56	0.15	50.0 s	81
10	-17711.57	-17711.56	0.01	88.0 s	121

Table 5.1: Results of Algorithm 5.1 for Prod-Mix

5.5 Extension to general cost

We now provide an adapted partition oracle for problems with finitely supported recourse matrix $\boldsymbol{W}$ and cost $\boldsymbol{q}$. The convergence results of Section 5.3 can directly be applied.

By Assumption 6.1, the admissible primal set $\{y \in \mathbb{R}^m \,|\, Tx + Wy = h, y \geqslant 0\}$ is non-empty and compact. Then, by strong duality, we can rewrite Q defined in Eq. (6.16b) as

$$Q(x, W, q, T, h) = \max_{\lambda \in \mathbb{R}^\ell} \{(h - Tx)^\top \lambda \quad | \quad W^\top \lambda \leqslant q\} \tag{5.39}$$

We now define the dual admissible set $D_{W,q}$ as

$$D_{W,q} := \{\lambda \in \mathbb{R}^\ell \mid W^\top \lambda \leqslant q\} \tag{5.40}$$

When W and q are fixed, the value of $Q(\check{x}, W, q, T, h)$ depends on which normal cone $h - T\check{x}$ belongs to. Thus, we finally define

$$E_{N,\check{x}} := \{(T, h) \mid h - T\check{x} \in \operatorname{ri}(N)\} \tag{5.41a}$$
$$\mathcal{R}_{\check{x},W,q} := \{E_{N,\check{x}} \mid N \in \mathcal{N}(D_{W,q})\} \tag{5.41b}$$

We begin with the finitely supported $\boldsymbol{q}$ case as a warm-up.

Remark 5.12 (Finitely supported $\boldsymbol{q}$)**.** *We can show that, when W and q are fixed, $\mathcal{R}_{\check{x},W,q}$ is an adapted partition to $\check{x}$ (see Theorem 5.3). If* $\operatorname{supp}(\boldsymbol{W})$ *and* $\operatorname{supp}(\boldsymbol{q})$ *are finite, we can extend this result to show that $\{\{W\} \times \{q\} \times \mathcal{R}_{\check{x},W,q} \,|\, (W, q) \in \operatorname{supp}(\boldsymbol{W}, \boldsymbol{q})\}$ is an adapted partition to $\check{x}$:*

$$\mathbb{E}\big[Q(\check{x}, \boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})\big] \tag{5.42a}$$
$$= \sum_{(W,q) \in \operatorname{supp}(\boldsymbol{W}, \boldsymbol{q})} \mathbb{E}\Big[Q(\check{x}, W, q, \boldsymbol{T}, \boldsymbol{h}) \;\Big|\; \boldsymbol{W} = W, \boldsymbol{q} = q\Big] \mathbb{P}[\boldsymbol{W} = W, \boldsymbol{q} = q] \tag{5.42b}$$
$$= \sum_{(W,q) \in \operatorname{supp}(\boldsymbol{W}, \boldsymbol{q})} \sum_{R \in \mathcal{R}_{\check{x},W,q}} Q\Big(\check{x}, \mathbb{E}\Big[\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h} \;\Big|\; (\boldsymbol{T}, \boldsymbol{h}) \in R, \boldsymbol{W} = W, \boldsymbol{q} = q\Big]\Big) \mathbb{P}[(\boldsymbol{T}, \boldsymbol{h}) \in R, \boldsymbol{W} = W, \boldsymbol{q} = q] \tag{5.42c}$$

We now extend this result to the case where $\boldsymbol{q}$ has a non-finitely supported distribution. This extension relies on a partition $\mathcal{S}_W$ of $\mathbb{R}^{\ell \times m}$ such that $q \mapsto \mathcal{R}_{\check{x},W,q}$ is constant on each $S \in \mathcal{S}_W$. Actually, this partition $\mathcal{S}_W$ is the collection of relative interiors of secondary cones $\Sigma\text{-fan}^\circ(W)$ (see Definition 3.20).

Let $\mathcal{S}_W$ be the collection of relative interiors of the elements of $\Sigma\text{-fan}(W)$:

$$\mathcal{S}_W := \Sigma\text{-fan}^\circ(W) = \{\operatorname{ri}(S) \mid S \in \Sigma\text{-fan}(W)\}. \tag{5.43}$$

In particular; the elements of $\mathcal{S}_W$ are relatively open (convex) cones of $\mathbb{R}^m$. Further, note that Proposition 3.21 provides constructive representation of Σ-fan(W) and thus of $\mathcal{S}_W$, which paves the way toward explicit computation of $\mathcal{S}_W$.

Lemma 5.13. *Let $W \in \mathbb{R}^{\ell\times m}$ and $S \in \mathcal{S}_W$. For every $q, q' \in S$ we have $\mathcal{R}_{x,W,q} = \mathcal{R}_{x,W,q'}$. Consequently, instead of considering an infinite number of $\mathcal{R}_{x,W,q}$ parametrized by q, we can consider a finite number of $\mathcal{R}_{x,W,S}$ parametrized by $S \in \mathcal{S}_W$ where*

$$\mathcal{R}_{x,W,S} := \mathcal{R}_{x,W,q} \text{ for an arbitrary } q \in S. \tag{5.44}$$

Proof. Let $D := \{(\mu, c) \in \mathbb{R}^\ell \times \mathbb{R}^m \,|\, W^\top\mu \leqslant q\}$ be the dual coupling polyhedron, and $\pi_c^{\mu,c}$ the orthogonal projection of $\mathbb{R}^\ell \times \mathbb{R}^m$ to $\mathbb{R}^m$ Recall that $\mathcal{C}(D, \pi_c^{\mu,c})$ is the chamber complex of D along $\pi_c^{\mu,c}$ (see Definition 3.10). By Theorem 3.34, the chamber complex $\mathcal{C}(D, \pi_q^{\mu,q})$ is equal to the secondary fan Σ-fan(W), and then for all $S \in \mathcal{S}_W$ there exists $\gamma \in \mathcal{C}(D, \pi_q^{\mu,q})$ such that $S = \mathrm{ri}(\gamma)$. By Proposition 3.30, for all $S \in \mathcal{S}_W$, we have the normal equivalence property, i.e. for all $q, q' \in S$, $\mathcal{N}(D_{W,q}) = \mathcal{N}(D_{W,q'})$. Then, since by definition $\mathcal{R}_{x,W,q} = \{E_{N,x} \,|\, N \in \mathcal{N}(W_{W,q})\}$, we have $\mathcal{R}_{x,W,q} = \mathcal{R}_{x,W,q'}$ for all $q, q' \in S$. □

We now leverage this reduction to a finite number of $\mathcal{R}_{x,W,S}$ to define an adapted partition. By using the basis decomposition theorem (see Theorem 3.27), we deduce the following lemma which give regions where we can interchange the function Q with the expectation. This lemma can be seen as an exact quantization result. For sake of simplicity, we make a slight abuse of notation with, for any event $A \in \mathcal{A}$,

$$Q\Big(x, \mathbb{E}\big[(\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{\xi})|A\big]\Big) = Q\Big(x, \mathbb{E}\big[\boldsymbol{W}|A\big], \mathbb{E}\big[\boldsymbol{q}|A\big], \mathbb{E}\big[\boldsymbol{\xi}|A\big]\Big). \tag{5.45}$$

Lemma 5.14. *Let $W \in \mathbb{R}^{\ell\times m}$. Assume that $(\boldsymbol{T}, \boldsymbol{h})$ and $\boldsymbol{q}$ are* independent *random variables, then, for all $S \in \mathcal{S}_W$ and $R \in \mathcal{R}_{\check{x},W,S}$,*

$$Q\Big(\check{x}, W, \mathbb{E}\big[\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h} \mid \boldsymbol{q} \in S, (\boldsymbol{T}, \boldsymbol{h}) \in R\big]\Big) = \mathbb{E}\Big[Q(\check{x}, W, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \mid \boldsymbol{q} \in S, (\boldsymbol{T}, \boldsymbol{h}) \in R\Big] \tag{5.46}$$

Proof. Let $S \in \mathcal{S}_W$ and $R \in \mathcal{R}_{\check{x},W,S}$. There exists $N \in \mathcal{N}(D_{W,q})$ such that and all $(T, h) \in R$, we have $h - T\check{x} \in \mathrm{ri}(N)$ for all $(T, h) \in R$. In particular, by the basis decomposition Theorem 3.27 in standard form, there exists a basis $B \subset [m]$ such that

$$\forall q \in S, \quad \forall (T, h) \in R, \qquad Q(\check{x}, W, q, T, h) = q_B^\top W_B^{-1}(h - T\check{x}) \tag{5.47}$$

Let B such a basis. By independence of $(\boldsymbol{T}, \boldsymbol{h})$ and $\boldsymbol{q}$, we have

$$\begin{aligned}
\mathbb{E}\big[Q(\check{x}, W, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \mid \boldsymbol{q} \in S, (\boldsymbol{T}, \boldsymbol{h}) \in R\big] &= \mathbb{E}\big[\boldsymbol{q}_B^\top W_B^{-1}(\boldsymbol{h} - \boldsymbol{T}\check{x}) \mid \boldsymbol{q} \in S, (\boldsymbol{T}, \boldsymbol{h}) \in R\big] && \text{(5.48a)}\\
&= \mathbb{E}\big[\boldsymbol{q}_B^\top \mid \boldsymbol{q} \in S\big] W_B^{-1} \mathbb{E}\big[\boldsymbol{h} - \boldsymbol{T}\check{x} \mid (\boldsymbol{T}, \boldsymbol{h}) \in R\big] && \text{(5.48b)}\\
&= Q\big(\check{x}, W, \mathbb{E}\big[\boldsymbol{q} \mid \boldsymbol{q} \in S\big], \mathbb{E}\big[\boldsymbol{T}, \boldsymbol{h} \mid (\boldsymbol{T}, \boldsymbol{h}) \in R\big]\big) && \text{(5.48c)}
\end{aligned}$$

Indeed, by convexity of S (resp. R), we have $\mathbb{E}\big[\boldsymbol{q} \mid \boldsymbol{q} \in S\big] \in S$ (resp. $\mathbb{E}\big[\boldsymbol{T}, \boldsymbol{h} \mid (\boldsymbol{T}, \boldsymbol{h}) \in R\big] \in R$). □

By summing over all $W \in \mathrm{supp}(\boldsymbol{W})$, all S in $\mathcal{S}_W$ and $R \in \mathcal{R}_{\check{x},W,S}$ and applying this lemma for every term of the sum, we can now deduce an explicit adapted partition.

Theorem 5.15 (Adapted partition for general second stage cost q)**.** *Assume that* $\boldsymbol{q}$ *and* $\boldsymbol{\xi}$ *are independent conditionally to* $\boldsymbol{W}$ *and that* $\operatorname{supp}(\boldsymbol{W})$ *is finite. We define* $\mathcal{P}_{\check{x}}$ *the following partition*

$$\mathcal{P}_{\check{x}} = \{\{W\} \times S \times R \quad | \quad W \in \operatorname{supp}(\boldsymbol{W}),\, S \in \mathcal{S}_W,\, R \in \mathcal{R}_{\check{x},W,S}\} \tag{5.49}$$

then $\mathcal{P}_{\check{x}}$ *is an adapted partition to* $\check{x}$*.*

Proof. We denote $A_{W,S,R}$ the event $\{\boldsymbol{W} = W, \boldsymbol{q} \in S, (\boldsymbol{T}, \boldsymbol{h}) \in R\}$, we then have

$$\begin{aligned}
V_{\mathcal{P}_{\check{x}}}(\check{x}) &:= \sum_{P \in \mathcal{P}_{\check{x}}} \mathbb{P}\big[(\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \in P\big]\; Q(\check{x}, \mathbb{E}\big[(\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \,|\, (\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \in P\big]) && \text{(5.50a)}\\
&= \sum_{W \in \operatorname{supp}(\boldsymbol{W})} \sum_{S \in \mathcal{S}_W} \sum_{R \in \mathcal{R}_{\check{x},W,S}} \mathbb{P}\big[A_{W,S,R}\big]\; Q(\check{x}, \mathbb{E}\big[(\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \,|\, A_{W,S,R}\big]) && \text{(5.50b)}\\
&= \sum_{W \in \operatorname{supp}(\boldsymbol{W})} \sum_{S \in \mathcal{S}_W} \sum_{R \in \mathcal{R}_{\check{x},W,S}} \mathbb{P}\big[A_{W,S,R}\big]\; Q(\check{x}, W, \mathbb{E}\big[(\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \,|\, A_{W,S,R}\big]) && \text{(5.50c)}\\
&= \sum_{W \in \operatorname{supp}(\boldsymbol{W})} \sum_{S \in \mathcal{S}_W} \sum_{R \in \mathcal{R}_{\check{x},W,S}} \mathbb{P}\big[A_{W,S,R}\big]\; \mathbb{E}\big[Q(\check{x}, W, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \,|\, A_{W,S,R}\big] && \text{(5.50d)}\\
&= \sum_{W \in \operatorname{supp}(\boldsymbol{W})} \sum_{S \in \mathcal{S}_W} \sum_{R \in \mathcal{R}_{\check{x},W,S}} \mathbb{P}\big[A_{W,S,R}\big]\; \mathbb{E}\big[Q(\check{x}, \boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \,|\, A_{W,S,R}\big] && \text{(5.50e)}\\
&= \mathbb{E}\big[Q(\check{x}, \boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})\big] = V(\check{x}) && \text{(5.50f)}
\end{aligned}$$

Eq. (5.50a) comes from the definition of the partitioned expected cost-to-go function $V_{\mathcal{P}_{\check{x}}}$ (see (5.8)), and Eq. (5.50b) from the definition of $\mathcal{P}_{\check{x}}$. The equality (5.50b)=(5.50c) is simply the abuse of notation presented in (5.45). Conditioned by $\boldsymbol{W} = W$, we can use Eq. (5.46) to obtain (5.50c)=(5.50d). Finally, law of total expectation yields (5.50e)=(5.50f).

It remains to prove that $V_{\mathcal{P}}$ is valid , *i.e.* $V_{\mathcal{P}_{\check{x}}} \leqslant V$. For all $W \in \operatorname{supp}(\boldsymbol{W})$ and $S \in \mathcal{S}_W$, we denote $\mathbb{E}_{W,S}$ (resp $\mathbb{P}_{W,S}$) the expectation (resp. the probability) conditional to the event $\{\boldsymbol{W} = W, \boldsymbol{q} \in S\}$. Let $x \in \mathbb{R}^n$. By the law of total expectation, we have

$$V_{\mathcal{P}_{\check{x}}}(x) = \sum_{W \in \operatorname{supp}(\boldsymbol{W})} \sum_{S \in \mathcal{S}_W} \mathbb{P}[\boldsymbol{W} = W, \boldsymbol{q} \in S] \sum_{R \in \mathcal{R}_{\check{x},W,S}} \mathbb{P}_{W,S}[\boldsymbol{\xi} \in R]\; Q(x, W, \mathbb{E}_{W,S}[(\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \mid (\boldsymbol{T}, \boldsymbol{h}) \in R] \tag{5.51}$$

By the independence of $\boldsymbol{q}$ and $(\boldsymbol{T}, \boldsymbol{h})$ and Lemma 5.14 we have

$$Q(x, W, \mathbb{E}_{W,S}[(\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}) \mid (\boldsymbol{T}, \boldsymbol{h}) \in R]) = Q(x, W, \mathbb{E}_{W,S}[\boldsymbol{q}], \mathbb{E}_{W,S}[(\boldsymbol{T}, \boldsymbol{h}) \mid (\boldsymbol{T}, \boldsymbol{h}) \in R]) \tag{5.52}$$

By convexity of $(T, h) \mapsto Q(x, W, q, T, h)$ and Jensen inequality, we have that

$$Q(x, W, \mathbb{E}_{W,S}[\boldsymbol{q}], \mathbb{E}_{W,S}[(\boldsymbol{T}, \boldsymbol{h}) \mid (\boldsymbol{T}, \boldsymbol{h}) \in R]) \leqslant \mathbb{E}_{W,S}[Q(x, W, \mathbb{E}_{W,S}[\boldsymbol{q}], \boldsymbol{T}, \boldsymbol{h}) \mid (\boldsymbol{T}, \boldsymbol{h}) \in R)] \tag{5.53}$$

Now, for an event A, note that we have, by applying the law of total expectation and

Lemma 5.14 twice, and with the abuse of notation Eq. (5.45),

$$\mathbb{E}_{W,S}[Q(x,W,\mathbb{E}_{W,S}[\boldsymbol{q}],\boldsymbol{T},\boldsymbol{h})|A] \tag{5.54a}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{E}_{W,S}[\mathbb{1}_{(\boldsymbol{T},\boldsymbol{h})\in R}\, Q(x,W,\mathbb{E}_{W,S}[\boldsymbol{q}],\boldsymbol{T},\boldsymbol{h})|A] \tag{5.54b}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{P}_{W,S}[(\boldsymbol{T},\boldsymbol{h})\in R]\mathbb{E}_{W,S}[Q(x,W,\mathbb{E}_{W,S}[\boldsymbol{q}],\boldsymbol{T},\boldsymbol{h}) \mid A\cap(\boldsymbol{T},\boldsymbol{h})\in R] \tag{5.54c}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{P}_{W,S}[(\boldsymbol{T},\boldsymbol{h})\in R]Q\left(x,\mathbb{E}_{W,S}[(W,\mathbb{E}_{W,S}[\boldsymbol{q}],\boldsymbol{T},\boldsymbol{h}) \mid A\cap(\boldsymbol{T},\boldsymbol{h})\in R]\right) \quad \text{by Lemma 5.14} \tag{5.54d}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{P}_{W,S}[(\boldsymbol{T},\boldsymbol{h})\in R]Q\big(x,\mathbb{E}_{W,S}[(W,\boldsymbol{q},\boldsymbol{T},\boldsymbol{h}) \mid A\cap(\boldsymbol{T},\boldsymbol{h})\in R]\big) \quad \text{by Eq. (5.45)} \tag{5.54e}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{P}_{W,S}[(\boldsymbol{T},\boldsymbol{h})\in R]\mathbb{E}_{W,S}[Q(x,W,\boldsymbol{q},\boldsymbol{T},\boldsymbol{h}) \mid A\cap(\boldsymbol{T},\boldsymbol{h})\in R] \quad \text{by Lemma 5.14} \tag{5.54f}$$

$$= \sum_{R\in\mathcal{R}_{x,W,S}} \mathbb{P}_{W,S}[(\boldsymbol{T},\boldsymbol{h})\in R]\mathbb{E}_{W,S}[Q(x,\boldsymbol{W},\boldsymbol{q},\boldsymbol{T},\boldsymbol{h}) \mid A\cap(\boldsymbol{T},\boldsymbol{h})\in R] \quad \text{under } \boldsymbol{W}=W \tag{5.54g}$$

$$= \mathbb{E}_{W,S}[Q(x,\boldsymbol{W},\boldsymbol{q},\boldsymbol{T},\boldsymbol{h})|A] \tag{5.54h}$$

By replacing A by $(\boldsymbol{T},\boldsymbol{h})\in R$, for $R\in\mathcal{R}_{\check{x},W,S}$ to Eq. (5.54h), we have

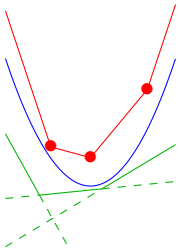
$$\mathbb{E}_{W,S}[Q(x,W,\mathbb{E}_{W,S}[\boldsymbol{q}],\boldsymbol{T},\boldsymbol{h}) \mid (\boldsymbol{T},\boldsymbol{h})\in R] = \mathbb{E}_{W,S}[Q(x,\boldsymbol{W},\boldsymbol{q},\boldsymbol{T},\boldsymbol{h}) \mid (\boldsymbol{T},\boldsymbol{h})\in R] \tag{5.55}$$

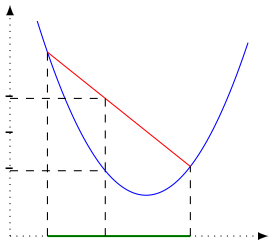
Combining(5.51),(5.52) and (5.53), we now get

$$V_{\mathcal{P}_{\check{x}}}(x) \leqslant \sum_{W\in\mathrm{supp}(\boldsymbol{W})}\sum_{S\in\mathcal{S}_W} \mathbb{P}[\boldsymbol{W}=W,\boldsymbol{q}\in S]\sum_{R\in\mathcal{R}_{\check{x},W,S}} \mathbb{P}_{W,S}[\boldsymbol{\xi}\in R]\mathbb{E}_{W,S}[Q(x,\boldsymbol{W},\boldsymbol{q},\boldsymbol{T},\boldsymbol{h}) \mid (\boldsymbol{T},\boldsymbol{h})\in R] \tag{5.56}$$

By the law of total expectation, we see that the right term is equal to $V(x)$. Thus, $V_{\mathcal{P}_{\check{x}}} \leqslant V(x)$. □

6 Trajectory following dynamic programming methods for multistage stochastic problems





Contents

6.1 Introduction

In this chapter, we present a framework called Trajectory Following Dynamic Programming (TFDP), that encompasses SDDP algorithm and most of its variants to solve non necessarily linear nor convex MSP. By proving an upper bound on the number of iterations, we deduce complexity results that are new for most of these algorithms. The contents of this chapter can be found in the preprint [FL22a].

More precisely, considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define a sequence of random variables, called *noises*, $(\boldsymbol{\xi}_t)_{t\in[T]}$, where $[T]$ stands for $\{1, \dots, T\}$ and T is the *horizon* of the problem. Assuming that each $\boldsymbol{\xi}_t$ has finite support of size n_c, an MSP problem admits an equivalent deterministic formulation with $O(n_c^T)$ variables. There are multiple algorithms (for a recent introduction to the topic we recommend [RW21]), each with various extensions and a rich literature, that exploit the special structure of the equivalent deterministic formulation, among which L-Shaped method [VSW69, LL93], and its extension to MSP *i.e.* nested Bender's decomposition [Lou80, Bir85], or progressive hedging algorithm [RW91].

However, each of these algorithms are numerically limited to small horizon T. For larger horizon we need some additional assumptions on the noises. If they have a limited memory (i.e. that $(\boldsymbol{\xi}_t, \boldsymbol{\xi_{t+1}}, \dots, \boldsymbol{\xi}_{t+\tau})$ is a Markov chain - for adequate indices) this open the door to Dynamic Programming methods, among which the Stochastic Dual Dynamic Programming (SDDP) algorithm [PP91] algorithm, and its variants (e.g [BDZ17, ZAS19, ACdC20, PWB20]). All these algorithms compute a state trajectory and then follow it to update approximations of cost-to-go functions. We call them Trajectory Following Dynamic Programming (TFDP) algorithms.

6.1.1 Problem setting

We present here the general setting of multistage stochastic problem (MSP) we are considering in the chapter. We also introduce three assumptions that are assumed to hold true throughout the chapter.

All random variables (noises $\boldsymbol{\xi}_t$ or states $\boldsymbol{x}_t$) are assumed to be valued, for some adequate integer n_t, in $\mathbb{R}^{n_t}$ endowed with its Borel σ-algebra. To model the constraint of our stochastic problem, we consider for $t \in [T]$, the following Borel-measurable set-valued applications $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Xi_t \rightrightarrows \mathbb{R}^{n_t}$ where $\Xi_t := \mathrm{supp}(\boldsymbol{\xi}_t) \subseteq \mathbb{R}^n$. We further assume, for simplicity, that the first noise is deterministic, that is $\Xi_1 = \{\xi_1\}$. For notational consistency we introduce $x_0 \in \mathbb{R}^{n_0}$ as a parameter, and $\boldsymbol{x}_0$ as the random variable with support $\{x_0\}$. We define recursively a sequence of *reachable set* by

$$X_0^r = \{x_0\} \tag{6.1a}$$

$$X_t^r = \bigcup_{x_{t-1}\in X_{t-1}^r} \bigcup_{\xi\in\Xi_t} \mathcal{X}_t(x_{t-1}, \xi) \qquad \forall t \in [T]. \tag{6.1b}$$

Finally, we consider a sequence of *loss functions* $(\ell_t)_{t\in[T]}$ where $\ell_t : \mathbb{R}^{n_t} \times \Xi_t \to \mathbb{R} \cup \{+\infty\}$.

Assumption 6.1 (Compatibility of constraints). We make the following assumptions, for all $t \in [T]$,

i) ℓ_t is a proper normal integrand;

ii) for all $x_t \in X_t^r$, the random variable $\ell_t(x_t, \boldsymbol{\xi}_t)$ is integrable (in particular $\ell_t(x_t, \boldsymbol{\xi}_t) < +\infty$ $\mathbb{P}$-almost surely);

iii) for all $x_{t-1} \in X_{t-1}^r$ and almost all $\xi_t \in \Xi_{t-1}$, $\mathcal{X}_t(x_{t-1}, \xi_t)$ is a non-empty compact subset of $\mathbb{R}^{n_t}$.

Finally, we say that $(\boldsymbol{x}_t)_{t\in[1:T]}$ is an *admissible policy* if it is a sequence of random variables such that, for all $t \in [T]$, $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ $\mathbb{P}$-almost surely, and $\boldsymbol{x}_t$ is measurable with respect to $\sigma(\{\boldsymbol{\xi}_\tau\}_{\tau\in[t]})$. We denote X^{ad} the set of admissible policies. Then, the multistage stochastic problem (MSP) consists in minimizing over the set of admissible policies the sum of losses, that is

$$\min_{\boldsymbol{x}\in X^{ad}} \quad \mathbb{E}\Big[\sum_{t=1}^{T} \ell_t(\boldsymbol{x}_t, \boldsymbol{\xi}_t)\Big] \tag{MSP}$$

Assumption 6.1 ensures that (MSP) is well-posed and admits an optimal solution. It also guarantees that we are in a *relatively complete recourse* setting in the sense that any sequence of variable $(\boldsymbol{x}_\tau)_{\tau\leqslant t}$ satisfying $\boldsymbol{x}_\tau \in X_\tau(\boldsymbol{x}_{\tau-1}, \boldsymbol{\xi}_\tau)$, for $\tau \leqslant t$ can be completed into an admissible policy $(\boldsymbol{x}_\tau)_{\tau\leqslant T}$ such that $\mathbb{E}\big[\sum_{t=1}^{T} \ell_t(\boldsymbol{x}_t, \boldsymbol{\xi}_t)\big] < +\infty$.

As we are considering Dynamic Programming methods, the following *stagewise independence* assumption is assumed to holds true.

Assumption 6.2 (Stagewise independence). $(\boldsymbol{\xi}_t)_{t\in[T]}$ is a sequence of independent exogeneous random variables, *i.e.* such that the law of $\boldsymbol{\xi}_t$ is independent of all decisions variables.

Leveraging Assumption 6.2, we can rewrite Problem (MSP) in the following equivalent nested form

$$\min_{x_1\in X_1(x_0,\xi_1)} \ell_1(x_1, \xi_1) + \mathbb{E}\Big[\min_{x_2\in X_2(x_1,\boldsymbol{\xi}_2)} \ell_2(x_2, \boldsymbol{\xi}_2) + \mathbb{E}\big[\cdots + \mathbb{E}\big[\min_{x_T\in\mathcal{X}_T(x_{T-1},\boldsymbol{\xi}_T)} \ell_T(x_T, \boldsymbol{\xi}_T)\big]\big]\Big] \tag{6.2}$$

which can be tackled by Dynamic Programming. To this end we introduce the following (backward) *Bellman operators*. For a measurable proper l.s.c function $\widetilde{V} : \mathbb{R}^{n_t} \to \mathbb{R} \cup +\infty$, we denote the Bellman operator of Problem (MSP) applied to $\widetilde{V}$ by

$$\hat{\mathcal{B}}_t(\widetilde{V}) = \begin{cases} \mathbb{R}^{n_t} \times \Xi_{t+1} & \to \mathbb{R} \cup \{+\infty\} \\ (x_t, \xi_{t+1}) & \mapsto \min\limits_{x_{t+1}\in\mathcal{X}_{t+1}(x_t,\xi_{t+1})} \ell_{t+1}(x_{t+1}, \xi_{t+1}) + \widetilde{V}(x_{t+1}) \end{cases} \tag{6.3a}$$

Further, note that for $\widetilde{V}$ l.s.c. and finite valued on X_t^r, $\hat{\mathcal{B}}_t(\widetilde{V})$ is also a normal integrand. We then define,

$$\mathcal{B}_t(\widetilde{V}) : x_t \mapsto \mathbb{E}[\hat{\mathcal{B}}_t(\widetilde{V})(x_t, \boldsymbol{\xi}_{t+1})] \tag{6.3b}$$

With this notation we define by induction the *expected cost-to-go* functions $V_t : \mathbb{R}^{n_{t-1}} \to \mathbb{R}$

$$V_T :\equiv 0 \tag{6.4a}$$

$$V_t := \mathcal{B}_t(V_{t+1}) \qquad \forall t \in \{0, .., T-1\}. \tag{6.4b}$$

Finally, as the law of $\boldsymbol{\xi}_1$ is a dirac on ξ_1, the value of Problem (MSP) is simply $V_0(x_0)$.

Remark 6.1 (Stepwise control)**.** *For notational simplicity we chose to consider loss function ℓ_t that only depends on the next state*[1] *x_t. However, it is worth keeping in mind that these loss functions are often defined as the partial minimum of another normal integrand, i.e.*

$$\ell_t(x_t, \xi) = \inf_{y\in\mathbb{R}^m} \widetilde{\ell}(x_t, y, \xi).$$

In theory, the same problem can be tackled by extending the state vector x to also contains the step decisions y. However, this is misleading: the theoretical complexity is exponential in the dimension of x, which is in line with the curse of dimensionality of Dynamic Programming. Thus extending the state to include y falsely seems to imply an increase in the number of iterations required by trajectory following algorithms to converge. For example, in long term electricity management problem it is standard to have decisions y of dimension a few thousands (thermal generation, transmission on lines...) while the actual state x (hydroelectric storage) is of dimension a few dozen at most.

[1] Cost dependence on x_{t-1} is not considered here simply for notational convenience.

We end the presentation of our setting with a non-trivial assumption.

Assumption 6.3 (Lipschitz). For $t \in [T]$, we assume that[2]

i) X_t^r has a diameter smaller than $D_t < +\infty$;

ii) the expected cost-to-go function V_t is L_t-Lipschitz.

Both part of Assumption 6.3 are strong requirement, needed for the convergence results, while still being natural in most settings. Part i) is satisfied for example if Assumption 6.1 holds, $\mathcal{X}_t(x_{t-1}, \cdot)$ is Lipschitz for all $x_{t-1} \in X_{t-1}^r$ and all Ξ_t are bounded. Part ii) is satisfied under Assumption 6.1 in the linear case, or through an *extended relatively complete recourse* assumption (see [GLP15]) which requires that state x_t that are slightly outside of X_t^r are still admissible.

6.1.2 Review of known convergence results

The SDDP algorithm, and its brethren, called in this manuscript Trajectory Following Dynamic Programming, relies on the dynamic programming equation (6.4). The main idea consists in iteratively refining lower (and sometimes upper) approximations of the expected cost-to-go functions V_t. More precisely, at each iteration, they decide, in a forward phase, trial points at which the approximations should be refined. Then, in a backward phase, they construct *cuts*, that are functions that under approximate V_t. These cuts are as close as possible to the true expected cost-to-go functions around the trial points. The lower approximations are finally defined as the maximum of computed cuts. This is detailed in Section 6.2.

To our knowledge, almost all prior works make the following assumption or consider an approximated problem which satisfies this assumption.

Assumption (FSN) (Finitely supported noise). The support of the random process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is finite.

The first proven convergence result of SDDP algorithm is due to Philpott and Guan [PG08]. In this paper, the authors consider the linear setting. Using Assumption (FSN) they prove that the number of (affine) cuts that can be generated is finite. Then, leveraging the fact that each scenario is sampled an infinite number of time, they prove the almost-sure convergence in a finite number of iterations, without any bound on this number. Later convergence results by [GLP15] (then reformulated and adapted to the risk averse setting in [Gui16]) showed convergence in a non-linear, convex setting. Again, the proof argues that each scenario are selected randomly an infinite number of time. A technical lemma coupled with Borel-Cantelli's yields almost-sure asymptotic convergence.

Instead of random sampling, some deterministic sampling, have been proposed. The problem-child algorithm [BDZ17], which maintains both an upper and a lower approximation, proved convergence by showing that the gap between upper and lower bound is non-increasing with the iteration. This algorithm has been extended to convex-concave framework, using saddle-cuts [BDZ18], e.g. allowing for stagewise-dependent objective uncertainty in [DDB20], or risk-averse problem [GTW19]. In other cases, deterministic sampling are considered as a first step for proving the convergence of the randomized version [PG08, Lan20, ZS20].

The above papers all rely on affine, often called Benders', cuts. Some variants of SDDP, handled by our framework, uses other types of cuts and also have proven asymptotic convergence. Zou et. al. presented a version of SDDP for binary variables, which has an asymptotic convergence proven in [ZAS19] for the convex case, although the proof can be directly adapted to the Lipschitz case. In addition to traditional Benders' cuts it relies on integer, Lagrangian and strengthened Benders' cut, recalled in Section 6.A. Stochastic Lipschitz Dynamic Programming

[2] We do not necessarily require the knowledge of the diameters or Lipschitz constants.

(SLDP) by Ahmed et. al. [ACdC20] uses concave L_1 cuts instead of affine cuts for any Lipschitz V_t. MIDAS by Philpott et al. [PWB20] uses step-function cuts for quasi-monotonous V_t and also fall in this category.

Other works have been dedicated to improve the numerical efficiency of the algorithm. Some methods alleviate the computational burden of each iteration, like Guigues in [Gui20] which considers inexact cuts, or Bandarra and Guigues in [BG21] that present cut selection strategies, which delete some cuts from the representation of V_t. Other methods, like regularization approaches [AP18, VAdOS19, BFFdO20], try to reduce the number of required iterations. To our knowledge, if they sometimes preserve asymptotic convergence, none of these approaches provably reduce the number of iterations required to reach an ε-solution. These extensions are either handled by our framework or discussed in Section 6.2.2.

It is worth noting that the convergence arguments that relies on each possible scenarios being sampled an infinite number of time are mainly theoretical arguments: due to the sheer number of scenarios, in most applications, the algorithms sample only a very small subset of scenarios (and probably never twice the same).

In two recent papers ([Lan20, ZS19]) new approaches were developed, focusing on the state space akin to the complexity proof of Kelley's cutting plane algorithm. They independently obtained the first explicit bound on the number of iterations required to obtain an ε-solution. To this end, they fix the error ε and define some saturated points in the state space. These points are such that the gap between the approximated value and the true value is controlled. Then, leveraging Lipschitz continuity of the value function they control the error in a ball around the saturated points. As the reachable sets are compact, only a finite number of such balls exists. They then each provide a deterministic algorithm with proven convergence, and use it as a proxy to bound the expected number of iterations. Interestingly, the complexity of the deterministic algorithm is polynomial in the horizon T while the sampled algorithm complexity is exponential in T as it requires a given event to happen at each stage simultaneously.

All the convergence proofs recalled here rely on reachable set compactness, relatively complete recourse and finitely supported noise assumption. They then fall into two categories: either they directly use the Lipschitz continuity[3] of V_t, or argues that there exists only a finite number of possible cuts (*e.g.,* [PG08, ZAS19]). Our framework cover all these variants, but the convergence proof presented in Section 6.3, which is built on [Lan20], does not require the finitely supported noise assumption (FSN). It is instead replaced, for the randomized algorithms, by a dedicated nested Hoeffding lemma (see Section 6.C.1). This is another step toward understanding the practical convergence of these TFDP algorithms.

Further, without finitely supported noise assumption (FSN), the standard approach consists in first discretizing the noise and then solving the discretized problem. A common method consists in sampling the problem through the Sample Average Approximation (SAA) approaches. The statistical guarantees of this approach are discussed by Shapiro in [Sha11]. Other sampling strategies are numerically discussed in [HdMDMF11, Löh16]. While never used, to our knowledge, in the context of TFDP algorithms, there are ways of discretizing the noise distribution in order to guarantee that the value of the discretized model under (or over) estimate the value of the true problem, especially in the convex setting, see [BW86, Kuh06, MAB14, MP18].

An alternative approach could consist in finding a finitely supported, stagewise independent distribution that minimize the nested-distance [PP12], to provide a good representation of the true problem. Other approaches exist, like [Fra96, CS05], which uses convexity tools and information relaxation to construct bounds. These approaches seem more relevant for problems with non-independent noises.

Finally, SDDP has been extended to various problem settings to handle risk aversion (e.g.[STdCS13, GTW19, DM20]), infinite horizon (e.g. [SD20]), partially observable problems [DMP20]... We briefly discuss extensions to risk averse setting in the last section, other extensions are outside

[3]Actually MIDAS has a slightly milder requirement (see [PWB20, Eq. (17)]).

the scope of this manuscript.

6.1.3 Contributions and structure of the chapter

Our main contributions are the following:

- we provide a flexible framework (including inexact or regularized computations) for TFDP algorithms for finite horizon, risk neutral problems, that encompass at least 14 variants of SDDP summed up in Table 6.1;
- we provide geometric tools to extend those algorithms to non-finitely supported uncertainties, without sampling or approximations in the linear case;
- we give a convergence speed result with an upper bound on the (expected) number of iterations to reach an ε-solution for these algorithms (which is new for most of those variants) that does not require the finite support assumption;
- we explain how to adapt those results to the minimax case. Some risk averse or robust cases are seen as a special case.

The remains of the chapter is as follows. Section 6.2 introduces the general framework, discusses some extensions, and the classical ways of obtaining exact or approximated cuts. Section 6.3 provides the main convergence and complexity results. Finally, Section 6.4 briefly reviews extensions to some robust and risk averse settings. Technical proofs and definitions can be found in the appendix.

Algorithm's name	Paper	Node selection: Choice $\boldsymbol{\xi}_t^k$	$\mathcal{F}_t$	$\underline{V}_t^k$	$\overline{V}_t^k$	Hypothesis	Complexity known
SDDP	[PP91]	Random sampling	Exact	Benders cuts	V_t	Convex	✓
EDDP	[Lan20]	Explorative	Exact	Benders cuts	V_t	Convex	✓
APDP	[SS22]	Random sampling	Exact	Adaptive partition	V_t	Linear	×
SDDiP	[ZAS19]	Random sampling	Exact	Lagrangian or integer cuts	V_t	Mixed Integer Linear	×
MIDAS	[PWB20]	Random sampling	Exact	Step cuts	V_t	Monotonic Mixed Integer	×
SLDP	[ACdC20]	Random sampling	Exact	Reverse norm cuts	V_t	Non-Convex	×
	[BDZ17]	Problem child	Exact	Benders cuts	Epigraph as convex hull	Convex	×
	[BDZ18]	Problem child	Exact	Benders × Epigraph	Hypograph × Benders	Convex-Concave	×
RDDP	[GTW19]	Deterministic	Exact	Benders cuts	Epigraph as convex hull	Robust	×
ISDDP	[Gui20]	Random sampling	Inexact	Inexact Lagrangian cuts	V_t	Convex	×
TDP	[ACT20]	Problem child	Exact	Benders cuts	Min of quadratic	Convex	×
	[ZS19]	Random or Problem	Regularized	Generalized conjugacy cuts	Norm cuts	Mixed Integer Convex	✓
NDDP	[ZS20]	Random or Problem	Regularized	Benders cuts	Norm cuts	Distributionally Robust	✓
DSDDP	[LCC+20]	Random sampling	Exact	Benders cuts	Fenchel transform	Linear	×

Table 6.1: Synthesis of algorithms following the same framework

6.2 Trajectory Following Dynamic Programming framework

Various extensions of the Stochastic dual dynamic programming (SDDP) algorithm have been developed for different set of assumptions. In this section, we first present a generic algorithmic framework for TFDP algorithms (see Algorithm 6.1) for risk-neutral multistage stochastic program that encompasses multiple known algorithms (see Table 6.1). These algorithms consider under and over approximations of the expected cost-to-go functions. The under-approximations are defined as the maximum of basic functions called *cuts* (some classical cuts are recalled in Section 6.A). The upper-approximations are more diverse and not always computed. The remains of the section detail how to obtain cuts, in particular Section 6.2.3 presents the finitely supported case, while Section 6.2.4 builds approximated cuts in the convex case.

6.2.1 Algorithm

The flexible framework of Algorithm 6.1 defines improving lower approximations $\underline{V}_t^k$ (resp. upper approximations[4] $\overline{V}_t^k$) of the expected cost-to-go functions V_t. Each iteration k of the algorithm consists in a forward phase to determine where to refine the approximations, followed by a backward phase to actually refine the approximations.

During the forward phase, we generate a trajectory $x_1^k, \cdots, x_{T-1}^k$. Each x_t^k is chosen as an (almost) optimal decision at time t starting from state x_{t-1}^k, knowing that the random variable $\boldsymbol{\xi}_t$ takes the value ξ_t^k, and considering that the cost-to-go is given by the lower approximation $\underline{V}_t^{k-1}$. This is encapsulated in the forthcoming notion of *forward Bellman operator*. We denote $\gamma\text{-}\arg\min_{x\in X} f(x)$ the set of $x \in X$ such that $f(x) \leqslant \inf_{x\in X} f(x) + \gamma$. We now define, for any Lipschitz (on X_t^r) function $\widetilde{V}$, the set $\mathcal{X}_{\gamma,t}^\sharp(\widetilde{V})$ of γ-optimal solution of the parametrized stage t problem with cost-to-go function $\widetilde{V}$, that is,

$$\mathcal{X}_{\gamma,t}^\sharp(\widetilde{V}) : (x,\xi) \mapsto \gamma\text{-}\operatorname*{arg\,min}_{y\in\mathcal{X}_t(x,\xi)} \ell_t(y,\xi) + \widetilde{V}(y).$$

Since $\ell_t + \widetilde{V}$ is a normal integrand, [RW09, Corollary 14.33] guarantees measurability of $\mathcal{X}_{\gamma,t}^\sharp(\widetilde{V})$. Further, as $\mathcal{X}_{\gamma,t}^\sharp(\widetilde{V})$ is compact, there exists ([RW09, Cor 14.6]) a measurable selection of $\xi \mapsto \mathcal{X}_{\gamma,t}^\sharp(\widetilde{V})(x,\xi)$ for all $x \in X_t^r$.

The following definition mathematically formalizes the selection choice[5].

Definition 6.2 (Forward operator)**.** *We say that $\mathcal{F}_t$ is a γ_{t+1}^F-*forward operator *if, for all functions $\widetilde{V} : \mathbb{R}^{n_{t+1}} \to \mathbb{R} \cup \{+\infty\}$, Lipschitz on X_{t+1}^r and $x \in X_t^r$, $\mathcal{F}_t(\widetilde{V})(x,\cdot)$ is a measurable selection of $\mathcal{X}_{\gamma_{t+1}^F,t+1}^\sharp(\widetilde{V})(x,\cdot)$.*

During the backward phase, we refine the approximations $\underline{V}_t^k$ and $\overline{V}_t^k$, they are both assumed to be Lipschitz on X_t^r. Further, the lower approximations $\underline{V}_t^k$ is defined as the maximum of a finite number of *cuts*: $\underline{V}_t^k = \max_{\kappa\leqslant k} f_t^k$.

For Algorithm 6.1 to converge we make the following assumption on the approximations computed.

Assumption 6.4 (Admissible approximations). The computed cuts f_t^k of $\mathcal{B}_t(\underline{V}_{t+1}^k)$ at x_t^k satisfy:

i) f_t^k is $\underline{\gamma}_t$-tight, i.e. $f_t^k(x_t^k) \geqslant \mathcal{B}_t(\underline{V}_{t+1}^k)(x_t^k) - \underline{\gamma}_t$

ii) f_t^k is valid, i.e. $f_t^k \leqslant \mathcal{B}_t(\underline{V}_{t+1}^k)$

[4] In some common cases, the upper approximations are chosen as $\overline{V}_t^k = V_t$ but never evaluated.

[5] This choice is comparable to selecting a stage solver which always return the same solution among the set of optimal solutions.

iii) $\overline{V}_t^k$ is $\overline{L}_t$-Lipschitz

On the other hand the upper approximations $\overline{V}_t^k$, not necessarily computed, shall satisfy the following properties:

iv) $\overline{V}_t^k(x_t^k) \leqslant \mathcal{B}_t(\overline{V}_{t+1}^k)(x_t^k) + \overline{\gamma}_t$ (tightness)

v) $\overline{V}_t^k \geqslant \mathcal{B}_t(\overline{V}_{t+1}^k)$ (validity)

vi) $\overline{V}_t^k \leqslant \overline{V}_t^{k-1}$ (monotonicity)

vii) $\overline{V}_t^k$ is $\overline{L}_t$-Lipschitz

Data: Random variables $\boldsymbol{\xi}_t$, cost function at each step ℓ_t, constraints set-valued function X_t, initial state x_0, γ_t^F-forward operators $\mathcal{F}_t$.

1 $\underline{V}_t^0 \equiv -\infty$ and $\overline{V}_t^0 \equiv +\infty$ for $t \in [T]$;
2 **for** $k \in \mathbb{N}$ **do**
  `/* Forward phase */`
3 Set $x_0^k = x_0$;
4 **for** $t = 1 : T-1$ **do**
5 Choose $\xi_t^k \in \mathrm{supp}(\boldsymbol{\xi}_t)$;
6 Let $x_t^k = \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi_t^k)$;
7 **end**
  `/* Backward phase */`
8 Set $\underline{V}_T^k \equiv \overline{V}_T^k \equiv 0$;
9 **for** $t = T-1 : -1 : 1$ **do**
10 Find a $\underline{L}_t$-Lipschitz on X_t^r, valid and $\underline{\gamma}$-tight cut f_t^k of $\mathcal{B}_t(\underline{V}_{t+1}^k)$ at x_t^k, i.e. such that $f_t^k(x_t^k) \geqslant \mathcal{B}_t(\underline{V}_{t+1}^k)(x_t^k) - \underline{\gamma}_t$ and $f_t^k \leqslant \mathcal{B}_t(\underline{V}_{t+1}^k)$;
11 Set $\underline{V}_t^k = \max(\underline{V}_t^{k-1}, f_t^k)$;
12 Define $\overline{V}_t^k$ satisfying Assumption 6.4, Items iv) to vii) ;
13 **end**
14 **end**

Algorithm 6.1: A general framework for TFDP algorithms

For the algorithm to be well-defined we need to guarantee the existence of cuts and upper approximation satisfying previous assumption, as formally assumed now:

Assumption 6.5. For every $t \in [T]$ and $k \in \mathbb{N}^\star$, there exists at least one cut f_t^k of $\mathcal{B}_t(\underline{V}_{t+1}^k)$ satisfying Assumption 6.4.

This assumption is for example ensured through relatively complete recourse in the linear setting [PP91], through extended relatively complete recourse in the convex setting [GLP15], through relatively complete continuous recourse in the binary setting [ZAS19], and relatively complete recourse and Lipschitz assumption in the Lipschitz setting of [ACdC20].

Remark 6.3 (Asymmetry of upper and lower approximations)**.** *The framework is not symmetrical in its treatment of the upper and lower cost-to-go approximations. Indeed, Line 6 should not be done with the upper approximations*[6] *as it would restrict the exploration of the state space. For example, assume that $\overline{V}_t$ are (slightly Lipschitz-regularized) indicator functions of a single point, then the forward phase would always produce the same trajectory, and the upper bound would not be updated.*

Further, multiple TFDP algorithm do not actually compute $\overline{V}_t$, simply setting it to the true expected cost-to-go V_t (for iterations bounds).

[6] The upper approximations $(\overline{V}_t^k)_{t\in[T]}$ still provide an admissible policy through the forward Bellman operators which has interesting properties, see [LCC+20].

We have a first monotonicity result.

Lemma 6.4. *Under Assumptions 6.1 to 6.5, for all $k \in \mathbb{N}$, $t \in [T-1]$ and $x \in \mathbb{R}^{n_t}$ we have*

$$\underline{V}_t^{k-1}(x) \leqslant \underline{V}_t^k(x) \leqslant \mathcal{B}_t(\underline{V}_{t+1}^k)(x) \leqslant V_t(x) \tag{6.5a}$$

$$V_t(x) \leqslant \mathcal{B}_t(\overline{V}_{t+1}^k)(x) \leqslant \overline{V}_t^k(x) \leqslant \overline{V}_t^{k-1}(x) \tag{6.5b}$$

In particular, the gap can only decrease

$$0 \leqslant \overline{V}_t^k(x) - \underline{V}_t^k(x) \leqslant \overline{V}_t^{k-1}(x) - \underline{V}_t^{k-1}(x) \tag{6.6}$$

Proof. Direct by double induction on t and k and monotonicity of the Bellman operator. □

Remark 6.5 (The standard SDDP algorithm)**.** *The most common TFDP algorithm is the stochastic dual dynamic programming (SDDP). It was originally designed by Pinto and Pereira ([PP91]) for multistage stochastic linear problems. In SDDP, the value of the noise ξ_t^k, chosen in Line 5, is drawn randomly on* $\mathrm{supp}(\boldsymbol{\xi}_t)$ *which is assumed to be finite. The lower approximations are defined as maximum of affine cuts. For each $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$, computing $\hat{\mathcal{B}}_t(\underline{V}_{t+1}^{k-1})(x_t^k, \xi)$ consists in solving a LP, and standard linear programming duality yields a subgradient $\hat{\alpha}_t^k(\xi) \in \partial\hat{\mathcal{B}}_t(\underline{V}_{t+1}^{k-1})(x_t^k, \xi)$ and value $\hat{\beta}_t^k(\xi) = \hat{\mathcal{B}}_t(\underline{V}_{t+1}^{k-1})(x_t^k, \xi)$. Taking the expectation, we set $\alpha_t^k = \mathbb{E}\big[\hat{\alpha}_t^k(\boldsymbol{\xi}_t)\big]$ and $\beta_t^k = \mathbb{E}\big[\hat{\beta}_t^k(\boldsymbol{\xi}_t)\big]$, to define the so-called Benders' cut*

$$f_t^k : x_t \mapsto \alpha_t^{k\top}(x_t - x_t^k) + \beta_t^k. \tag{6.7}$$

Under relatively complete recourse assumption, the cuts can be assumed to be $\underline{L}_t$-Lipschitz. Further, in this simple setting, all errors are null: $\underline{\gamma}_t = \overline{\gamma}_t = \gamma_t^F = 0$. Finally, no upper bound are computed and the complexity results of Section 6.3 are obtained by taking $\overline{V}_t^k = V_t^k$.

Algorithm 6.1 is a flexible framework, and some lines remain to be detailed, which we now discuss.

Node selection choice in Line 5 Most TFDP algorithms choose ξ_t^k by drawing it randomly according to the law of the random variable $\boldsymbol{\xi}_t^k$. The forward phase can then be seen as a Monte Carlo method for finding a trajectory $x_1^k, \cdots, x_T^k$. Then, it is also possible to choose ξ_t^k thanks to quasi Monte-Carlo methods.

Another way of choosing ξ_t^k consists in picking the $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$ that maximizes a certain criterion. In [BDZ17], Baucke, Downward and Zakeri suggested to chose ξ_t^k such that x_t^k maximizes the gap between the upper and lower approximations, *i.e.*, $\overline{V}_t^k(x_t^k) - \underline{V}_t^k(x_t^k)$. They called this choice of ξ_t^k, the *problem child* node selection. In [Lan20], Lan presented the Explorative Dual Dynamic Programming algorithm, where ξ_t^k is chosen so that x_t^k is the most distinguishable point, i.e. such that x_t^k is far from the previous computed points, see Eq. (6.29b), we speak of *explorative* node selection.

The proofs of convergence are harder to derive when ξ_t^k is chosen randomly, and the best upper bound known on the number of iterations of these algorithms are exponential in the horizon T. In comparison, when ξ_t^k is chosen deterministically as the problem child or as the most distinguishable point, the number of iterations is bounded by a polynomial in T. However, random sampling is often more efficient in practice (and easier to implement). We discuss the complexity results in Section 6.3.

Forward operator choice in Line 6 In most algorithms, we assume that $\gamma_t^F = 0$ for all $t \in [T-1]$, thus $\mathcal{F}_{t-1}(\widetilde{V})(x, \cdot)$ is a measurable selection of $\arg\min_{y \in \mathcal{X}_t(x,\cdot)} \ell_t(y, \cdot) + \widetilde{V}(y)$. There has also been proposition to use inexact cuts [Gui20] to alleviate computational burden of each iteration.

Further, there have been various propositions to regularize the SDDP algorithm, see [AP18, VAdOS19, GLT20]. They mostly boil down to choosing a different forward operator, *e.g.,* by adding a regularization term, which can be seen as γ_t^F-forward operator with $\gamma_t^F \neq 0$. For example, one can choose $\mathcal{F}_{t-1}(\widetilde{V})(x,\xi)$ as a proximity operator $\operatorname{prox}_{\ell_t(\cdot,\xi)+\widetilde{V}(\cdot),\alpha}(\check{y})$ which by definition is equal to $\arg\min_{y\in\mathcal{X}_t(x,\xi)} \ell_t(y,\xi)+\widetilde{V}(y)+\alpha\|y-\check{y}\|_2^2$. In that case, if $X_t(x,\xi)$ has a finite diameter D, for $y = \mathcal{F}_{t-1}(\widetilde{V})(x,\xi)$, we have $\ell_t(y,\xi)+\widetilde{V}(y) \leqslant \min_{y'\in\mathcal{X}_t(x,\xi)} \ell_t(y',\xi)+\widetilde{V}(y')+\alpha D$. Then, $\mathcal{F}_t$ is an αD-forward operator.

Finally, it is important that the algorithm use a single γ_t^F-forward operator. Indeed, if the set of γ_t^F-optimal solutions $\mathcal{X}^\sharp_{\gamma_t^F,t}(\widetilde{V})(x,\xi)$ is not reduced to a single point, the convergence results only holds for the points selected by the forward operator. This remark is not only theoretical and have implications in practice: to be safe one should use the same solver (and parameters) during the training phase and exploitation phase of the algorithm. For example, consider a problem with two equivalent storage and that only one of them is required to provide an optimal solution. Consider two forward operators, the first one, $\mathcal{F}^1_{t-1}$, prefers using the first storage while the second, $\mathcal{F}^2_{t-1}$ prefers using the second storage. Now assume that the algorithm ran until convergence with $\mathcal{F}^1_{t-1}$ yielding the approximations $\underline{V}_t^\infty$. Then, $\underline{V}_t^\infty$ correctly evaluates the value of the first storage, but has no information on the second. Consequently, a trajectory given by $\mathcal{F}^2_{t-1}(\underline{V}_t^\infty)$ might be far from optimal. A discussion of this fact, and practical answers, can be found in [Dow18, §2.7].

Cuts f_t^k choice in Line 10 We need to compute cuts f_t^k to approximate $\mathcal{B}_t(\underline{V}^k_{t+1})$ in the neighborhood of x^k_{t-1}. Recall that in Eq. (6.3b), $\mathcal{B}_t$ is defined as an expectation of parametric Bellman operators $\mathcal{B}_t(\underline{V}^k_{t+1}) = \mathbb{E}[\hat{\mathcal{B}}_t(\underline{V}^k_{t+1})(\cdot,\boldsymbol{\xi}_t)]$ Eq. (6.3a). Then, we can compute the average cut f_t^k thanks to parametric cuts $\hat{f}^k_{t,\xi}$. In the finitely supported case in Section 6.2.3, we show that we can compute the average cut f_t^k directly by taking $f_t^k = \mathbb{E}[\hat{f}^k_{t,\boldsymbol{\xi}}]$ whereas in the convex, non finitely-supported case, we present in Section 6.2.4 methods to approximate $\mathbb{E}[f^k_{t,\boldsymbol{\xi}}]$. Finally, exact methods for linear problems are developed in Section 6.2.5. Furthermore, depending on the problem structure, there exist several types of parametrized cuts $\hat{f}^k_{t,\xi}$ in the literature. We recall them in Section 6.A.

Upper approximations $\overline{V}_t^k$ choice in Line 12 In most TFDP algorithms, no upper bound function is computed. In that case, we just set $\overline{V}_t^k \equiv V_t$ in the convergence proof. However, some algorithms rely on the computation of these upper bounds, for example for computing a problem-child node selection. In the convex case, assume that we have, for $t \in [T]$, some points $(x_t^\kappa, \overline{v}_t^\kappa)_{\kappa\in[k]}$ that are in the epigraph of V_t. Now define $\overline{V}_t^k$ such that $\operatorname{epi}(\overline{V}_t^k) = \operatorname{Conv}\left((x_t^\kappa,\overline{v}_t^\kappa)_{\kappa\in[k]}\right) + \{(x,z)\in\mathbb{R}^{n_t}\times\mathbb{R} \mid \overline{L}_t\|x\|_1 \leqslant z\} \subseteq \operatorname{epi}(V_t)$, see Fig. 6.2. Then, $\overline{V}_t^k$ is an upper-approximation $\overline{V}_t^k$ of V_t on X_t^r. Computing points $(x_t^\kappa,\overline{v}_t^\kappa)_{\kappa\in[k]}$ in the epigraph of $\hat{V}_t$ can be done either throughout the algorithm as in the problem-child approach [BDZ17], or in batch backward in time for a given set of trajectories as suggested by [PdMF13]. Upper approximation functions can also be obtained through duality see [LCC+20, dCL21].

6.2.2 Extensions of the framework

Although we tried to present a general framework, for the sake of simplicity, Algorithm 6.1 does not integrate every variants of SDDP. We now discuss how this framework can be extended and if the complexity results and proofs are still valid with these new extensions.

Multiple forward phases. In practice, SDDP is often implemented with multiple forward phases, *i.e.*, at iteration k we compute N forward phases $(x_t^{k,i})_{t\in[T-1],i\in[N]}$, in parallel. Conse-

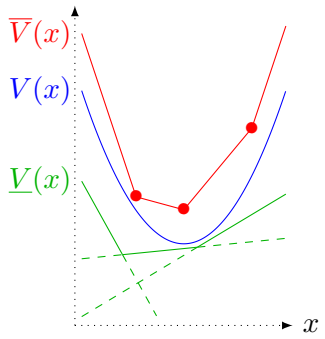


Figure 6.2: An example of upper and lower approximations

quently, in the backward phase we compute, for each time step $t \in [T-1]$, N tight and valid cuts $(f_t^{k,i})_{i\in[N]}$. This variation is included in the framework of Algorithm 6.1 by considering that the cut f_t^k is the maximum over $i \in [N]$ of all cuts $f_t^{k,i}$. The complexity results follow directly (in iteration number).

Multicut. In the finitely supported case, instead of computing an average cut f_t^k of the expected cost-to-go function V_t, it is possible to store for each $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$ a cut $\hat{f}_{t,\xi}$ of the cost-to-go function $\hat{V}_t(\cdot, \xi)$. Unlike the single cut case where $\underline{V}_t^k(\cdot) = \max_{\kappa\leqslant k} f_t^\kappa(\cdot)$, in the multicut case, we compute approximation functions as $\underline{V}_t^k(\cdot) = \sum_{\xi\in\mathrm{supp}(\boldsymbol{\xi}_t)} \mathbb{P}[\xi] \max_{\kappa\leqslant k} \hat{f}_{t,\xi}^\kappa(\cdot)$. Up to a slight reinterpretation, by considering a global cut $f_t(\cdot) = \sum_{\xi\in\mathrm{supp}(\boldsymbol{\xi}_t)} \mathbb{P}[\xi] \max_{\kappa\leqslant k} \hat{f}_{t,\xi}(\cdot)$, this variation is covered by our framework.

However, with continuous random variables, the notion of multiple cuts is not well-defined.

Cut computation in forward. Another variation of SDDP consists in computing the cuts during the forward phase (and no backward phase). In this variant, the cuts do not approximate $\mathcal{B}_t(\underline{V}_{t+1}^k)$ and $\mathcal{B}_t(\overline{V}_{t+1}^k)$ in the neighborhood of x_t^k, but approximate $\mathcal{B}_t(\underline{V}_{t+1}^{k-1})$ and $\mathcal{B}_t(\overline{V}_{t+1}^{k-1})$ in the same neighborhood. Although this variant is not handled by the framework, all proofs can be adapted in a straightforward manner. More precisely, we only need to adapt the forthcoming proof of Lemma 6.28. In particular, in the proof of Lemma 6.28, we obtain directly Eq. (6.57c) and Eq. (6.58c), without using the monotonicity, because we approximate $\mathcal{B}_t(\underline{V}_{t+1}^{k-1})$ and $\mathcal{B}_t(\overline{V}_{t+1}^{k-1})$.

Cut selection. After many iterations, the number of cuts can slow down the new iterations. To speed up SDDP iterations, another idea is to delete some cuts. For example, we can decide to delete only the dominated cuts, *i.e.*, the cuts that do not affect the values of the approximations $\underline{V}_t^k$. The monotonicity property and the complexity results are still valid in this setting. Unfortunately, finding which cut is dominated is time-consuming which do not make this method numerically efficient. Instead, we often use some heuristics to delete cuts which are probably dominated. However, these heuristics do not guarantee that we have the monotonicity property of approximations. Then, the complexity and convergence results seems harder to obtain. See [BG21] for an asymptotic convergence result on SDDP with cut selection.

Adaptive partition based methods In [SL15], Song and Luedtke presented the adaptive partition based methods (APM) to solve 2-stage linear problems by partitioning the set of scenarios. It was then adapted to the multistage case in [SS22] where Siddig and Song proposed an adaptive partition based SDDP, in both case under the finitely supported noise assumption (FSN). Recall, see Chapter 5, that the idea of APM is to replace the expected cost-go-function $V = \mathbb{E}[\hat{V}(\cdot, \boldsymbol{\xi})]$ by a partitioned expected cost-to-go function $V_{\mathcal{P}} = \sum_{P\in\mathcal{P}} \mathbb{P}[\boldsymbol{\xi} \in P] \hat{V}(\cdot, \mathbb{E}[\boldsymbol{\xi} \,|\, \boldsymbol{\xi} \in$

$P]$) where $\mathcal{P}$ is a partition of the uncertainty set Ξ. A partition $\mathcal{P}$ is said to be *tight at* $\check{x}$, if $V_{\mathcal{P}}(\check{x}) = V(\check{x})$, *valid* if $V_{\mathcal{P}}(x) \leqslant V(x)$ for all $x \in \mathbb{R}^{n_t}$ and *adapted to* $\check{x}$ if it is valid and tight at $\check{x}$. Then, when $\mathcal{P}$ is a partition adapted to $\check{x}$, we can see the partitioned expected cost-to-go function $V_{\mathcal{P}}$ as a valid and tight cut of V at $\check{x}$. Such cuts represent the tangeant cone of $\mathrm{epi}(\mathcal{B}_t(\underline{V}^k_{t+1}))$ at x where Benders' cut represent a single tangeant plane (see Lemma 5.8). APM methods were extended to general distribution in [RPM21]. In Chapter 5, we provided a necessary and sufficient condition for a partition to be adapted (without Assumption (FSN)) as well as a geometric method to obtain a valid and adapted partition. In particular, Adaptive Partition-based Dynamic Programming (APDP) algorithm of [SS22] is a TFDP algorithm falling in the framework of Algorithm 6.1. It can be adapted to the non-finitely supported case through the discussion in Section 6.2.5. As Ramirez-Pico and Moreno called GAPM the generalisation of APM methods to general distributions, we will name the GAPDP the extension of APDP algorithm to general distributions.

6.2.3 Cuts with finitely supported distribution

We now focus on finding a cut in Line 10 of Algorithm 6.1. More precisely, we want to approximate $\mathcal{B}_t(\underline{V}^k_{t+1})$ in the neighborhood of x^k_{t-1}. Recall that $\mathcal{B}_t$ is defined as an expectation of parametric Bellman operators $\hat{\mathcal{B}}_t$ (see Eq. (6.3)). When the distribution of $\boldsymbol{\xi}_t$ is finitely supported, computing a cut of $\hat{\mathcal{B}}_t(\widetilde{V})(\cdot, \xi)$ for each elements $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$ automatically yields a cut for $\mathcal{B}_t(\widetilde{V})$.

Proposition 6.6. *Assume that $\boldsymbol{\xi}_t$ is finitely supported with $p_\xi := \mathbb{P}[\boldsymbol{\xi}_t = \xi]$, for all $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$, then*

$$\mathcal{B}_t(\widetilde{V})(x) = \sum_{\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)} p_\xi \hat{\mathcal{B}}_t(\widetilde{V})(x, \xi) \tag{6.8a}$$

For every $\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)$, assume that $\hat{f}_\xi$ is a valid and $\underline{\hat{\gamma}}_{t,\xi}$-tight cut of function $\hat{\mathcal{B}}_t(\widetilde{V})(\cdot, \xi)$ at $\check{x}$, then we have

$$f := \sum_{\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)} p_\xi \hat{f}_\xi \text{ is a valid and } \underline{\gamma}_t\text{-tight cut of } \mathcal{B}_t(\widetilde{V}) \text{ at } \check{x} \text{ with } \underline{\gamma}_t := \sum_{\xi \in \mathrm{supp}(\boldsymbol{\xi}_t)} p_\xi \underline{\hat{\gamma}}_{t,\xi} \tag{6.8b}$$

In this finitely-supported distribution setting, it remains to find cuts of the function $\hat{\mathcal{B}}_t(\widetilde{V})(., \xi)$. There exist several tight and valid cuts depending on the structure of ℓ_t and X_t. We present classical cuts of the literature in Section 6.A where we detail under which conditions these cuts are tight and valid and show how to compute them.

6.2.4 Approximated cuts in the convex case

In this section, we now turn to obtaining approximated cuts leveraging convexity. We present a method based on the inequalities of Jensen and Edmundson-Madansky, adapting the results of Birge and Wets [BW86] to our setting, see also [KM$^+$76, 4.7].

We start by recalling two well-known useful convex inequalities illustrated in Fig. 6.3.

Proposition 6.7 (Jensen's and Edmunson-Madansky inequalities)**.** *Let $g : \mathbb{R}^\ell \mapsto \mathbb{R}$ be a convex function and $\boldsymbol{\xi}$ be a random variable. Assume that there exists a polytope*[7] *$\Xi \subset \mathbb{R}^\ell$ containing* $\mathrm{supp}(\boldsymbol{\xi})$.

For any $\xi \in \Xi$ we denote $S_\Xi(\xi)$ the set of barycentric coordinates of ξ, that is the set of coefficients $(\widetilde{\sigma}_{\Xi,v})_{v \in \mathrm{Vert}\,\Xi}(\xi) \in [0,1]^{|\mathrm{Vert}\,\Xi|}$ such that $\xi = \sum_{v \in \mathrm{Vert}(\Xi)} \widetilde{\sigma}_{\Xi,v}(\xi) v$ and $\sum_{v \in \mathrm{Vert}(\Xi)} \widetilde{\sigma}_{\Xi,v}(\xi) =$

[7]The results can be extended to the case where Ξ is an unbounded polyhedron. We must then consider a set $\mathrm{Ray}(\Xi)$ of extreme rays of the recession cone of Ξ (see [EZ94a])

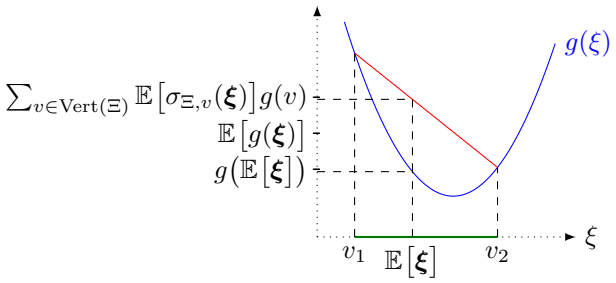


Figure 6.3: An illustration of Jensen and Edmundson-Madansky inequalities

1. *Let $\xi \mapsto (\sigma_{\Xi,v}(\xi))_{v\in\mathrm{Vert}\,\Xi}$ be any measurable selection*[8] *of $S_\Xi(\xi)$. We have the following inequality:*

$$g(\mathbb{E}[\boldsymbol{\xi}]) \leqslant \mathbb{E}[g(\boldsymbol{\xi})] \leqslant \sum_{v\in\mathrm{Vert}(\Xi)} \mathbb{E}[\sigma_{\Xi,v}(\boldsymbol{\xi})]g(v) \tag{6.9}$$

Moreover, if g is Lipschitz with constant L and Ξ has a diameter D, the gap is at most LD:

$$\sum_{v\in\mathrm{Vert}(\Xi)} \mathbb{E}[\sigma_{\Xi,v}(\boldsymbol{\xi})]g(v) \leqslant g(\mathbb{E}[\boldsymbol{\xi}]) + LD \tag{6.10}$$

Proof. Left-hand side of Eq. (6.9) is the classical Jensen inequality. Let $\xi \in \Xi$, as $(\sigma_{\Xi,v})_{v\in\mathrm{Vert}\,\Xi}(\xi)$ are barycentric coordinates, we have, by convexity of g, $g(\xi) \leqslant \sum_{v\in\mathrm{Vert}(\Xi)} \sigma_{\Xi,v}(\xi)g(v)$. Taking the expectation leads to the right-hand side of Eq. (6.9) called Edmundson-Madansky inequality.

Assume now that Ξ has diameter D. Since Ξ is convex, $\mathbb{E}[\boldsymbol{\xi}] \in \Xi$, thus for all $v \in \mathrm{Vert}\,\Xi$, $\|\mathbb{E}[\boldsymbol{\xi}] - v\| \leqslant D$. Further, as g is Lipschitz, we have $\|g(v) - g(\mathbb{E}[\boldsymbol{\xi}])\| \leqslant LD$. Taking the convex combination yields Eq. (6.10). □

These inequalities can be refined. Let $\mathcal{P}$ be a finite collection of almost surely disjoint polyhedra covering $\mathrm{supp}(\boldsymbol{\xi})$, i.e. $\mathrm{supp}(\boldsymbol{\xi}) \subset \cup_{P\in\mathcal{P}} P$ and $\mathbb{P}[P \cap P'] = 0$ if $P \neq P' \subset \mathcal{P}$. Then, by the law of total expectation, $\mathbb{E}[g(\boldsymbol{\xi})] = \sum_{P\in\mathcal{P}} \mathbb{P}[P]\mathbb{E}[g(\boldsymbol{\xi})|P]$. Applying Jensen and Edmundson inequalities to each term of this sum, we get

$$\sum_{P\in\mathcal{P}} \mathbb{P}[P]g(\mathbb{E}[\boldsymbol{\xi}|P]) \leqslant \mathbb{E}[g(\boldsymbol{\xi})] \leqslant \sum_{P\in\mathcal{P}} \mathbb{P}[P] \sum_{v\in\mathrm{Vert}(P)} \mathbb{E}[\sigma_{P,v}(\boldsymbol{\xi})]g(v) \tag{6.11}$$

In particular, if all polyhedra $P \in \mathcal{P}$ have a diameter smaller than d, the gap can be bounded by Ld.

We now get back to the problem of Line 10 of Algorithm 6.1 where we want to approximate $\mathcal{B}_t(\underline{V}^k_{t+1})$ in the neighborhood of x^k_{t-1}. Recall that $\mathcal{B}_t$ is defined as an expectation of parametric Bellman operators $\hat{\mathcal{B}}_t$ (see Eq. (6.3)). Unlike in Section 6.2.3 where the random variable where finitely supported, we cannot write the expected cut as a finite sum of parametric cuts. However, the Jensen and Edmundson-Madansky inequalities allows us to derive approximate cuts and upper bound functions.

Proposition 6.8. *Assume that ℓ_t is a jointly convex function with Lipschitz constant L. Let $\mathcal{P}$ be a finite collection of almost surely disjoint polyhedra covering* $\mathrm{supp}(\boldsymbol{\xi})$*, such that any $P \in \mathcal{P}$ has a diameter smaller than $d \in \mathbb{R}_+$. Denote for each $P \in \mathcal{P}$, $p_P := \mathbb{P}[P]$ and $\xi_P := \mathbb{E}[\boldsymbol{\xi}|P]$.*

For every $P \in \mathcal{P}$, assume that $\underline{f}_P$ is a valid and $\underline{\gamma}_{t,P}$-tight cut of the parametric function $\hat{\mathcal{B}}_t(\widetilde{V})(\cdot, \xi_P)$ at $\check{x}$, then by defining $\underline{\gamma}_t := Ld + \sum_{P\in\mathcal{P}} p_P\underline{\gamma}_{t,P}$, we have

$$\underline{f} := \sum_{P\in\mathcal{P}} p_P\underline{f}_P \text{ is a valid and } \underline{\gamma}_t\text{-tight cut of } \mathcal{B}_t(\widetilde{V}) \text{ at } \check{x} \tag{6.12}$$

[8] Such a selection always exists. Indeed, if Ξ is a simplex, barycentric coordinates are uniquely defined through a linear application. Then, any triangulation of Ξ define a measurable selection as piecewise linear applications.

For every $P \in \mathcal{P}$ *and* $v \in \operatorname{Vert} P$, *assume that* $\overline{f}_{v,P}$ *satisfies* $\overline{f} \geqslant \hat{\mathcal{B}}_t(\widetilde{V})(\cdot, v)$ *and* $\overline{f}(\check{x}) \leqslant \hat{\mathcal{B}}_t(\widetilde{V})(\check{x}, v) + \overline{\gamma}_{t,P}$, *then by defining* $\overline{\gamma}_t := Ld + \sum_{P\in\mathcal{P}} p_P \sum_{v\in\operatorname{Vert}(P)} \mathbb{E}\left[\sigma_{P,v}(\boldsymbol{\xi})\right]\overline{\gamma}_{t,P}$, *we have*

$$\overline{f} := \sum_{P\in\mathcal{P}} p_P \sum_{v\in\operatorname{Vert}(P)} \mathbb{E}\left[\sigma_{P,v}(\boldsymbol{\xi})\right]\overline{f}_{P,v} \text{ satisfies } \overline{f} \geqslant \mathcal{B}_t(\widetilde{V}) \text{ and } \overline{f}(\check{x}) \leqslant \mathcal{B}_t(\widetilde{V})(\check{x}) + \overline{\gamma}_t \tag{6.13}$$

This result can also be adapted to "saddle" cost functions, i.e. functions that are convex in some coordinates of ξ and concave in the other coordinates of ξ, by using both inequalities according to the sign of convexity, see e.g. [Kuh06, §4].

6.2.5 Exact SDDP in the linear case with general distributions

In this section, we consider the particular case of multistage linear stochastic programming i.e. Problem (MSP) where, for all $t \in [T]$, $\boldsymbol{\xi}_t = (\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{b}_t, \boldsymbol{c}_t)$, $\ell_t(x_t, \xi_t) = c_t^\top x_t$ is linear and $\mathcal{X}_t(x_{t-1}, \xi_t) = \{x_t \,|\, A_t x_t + B_t x_{t-1} = b_t,\ x_t \geqslant 0\}$ is a polyhedron. We first reformulate the stage problem (6.3a) as a standard two-stage linear program. Then, we show that we can construct exact cuts thanks to adapted partitions, fsuch as the partition defined in Theorem 5.15.

We make the following assumptions:

Assumption (LS) (Linear setting). For $t \in [T]$ we have $\ell_t(x_t, \xi_t) = c_t^\top x_t$ and $\mathcal{X}_t(x_{t-1}, \xi_t) = \{x_t \in \mathbb{R}^{n_t} \,|\, A_t x_t + B_t x_{t-1} = b_t,\ x_t \geqslant 0\}$. Further, the random variable $\boldsymbol{\xi}_t = (\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{b}_t, \boldsymbol{c}_t)$ and the approximated expected cost-to-go functions $\underline{V}_t^k$ satisfy

1. $\boldsymbol{A}_t$ has a finitely supported distribution;
2. $\boldsymbol{c}_t$ and $(\boldsymbol{B}_t, \boldsymbol{b}_t)$ are independent[9];
3. the lower expected cost-to-go function $\underline{V}_t^k$ are defined as the maximum of affine cuts, *i.e.*, we have $(\alpha_t^l)_{l\leqslant k}$ and $(\beta_t^l)_{l\leqslant k}$ such that

$$\underline{V}_t^k(x_t) = \max_{l\leqslant k}\ \alpha_t^{l\top} x_t + \beta_t^l \tag{6.14}$$

Under Assumption (LS), Bellman operator defined in (6.3a) applied to $\underline{V}_t^k$ reads

$$\begin{aligned}
\hat{\mathcal{B}}_{t-1}(\underline{V}_t^k)(x_{t-1}, \xi_t) = \min_{x_t, z}\quad & c_t^\top x_t + z && = \min_{x_t, z^+, z^-, r}\quad && c_t^\top x_t + z^+ - z^- \\
\text{s.t.}\quad & A_t x_t + B_t x_{t-1} = b_t, && \text{s.t.}\quad && A_t x_t + B_t x_{t-1} = b_t, \\
& \alpha_t^{\kappa\top} x_t + \beta_t^\kappa \leqslant z,\ \forall \kappa \leqslant k && && \alpha_t^{\kappa\top} x_t + \beta_t^l + r = z^+ - z^-,\ \forall \kappa \leqslant k \\
& x_t \geqslant 0 && && x_t, z^+, z^-, r \geqslant 0
\end{aligned}$$

Then, for any $t \in [T]$ and $k \in \mathbb{N}$, setting

$$x := x_t,\ y := (x_t, z^+, z^-, r),\ \boldsymbol{W} := \begin{pmatrix} \boldsymbol{A}_t & 0 & 0 & 0 \\ \alpha_t^1 & -1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_t^k & -1 & 1 & 1 \end{pmatrix},\ \boldsymbol{q} := \begin{pmatrix} \boldsymbol{c}_t \\ 1 \\ -1 \\ 0 \end{pmatrix},\ \boldsymbol{T} := \begin{pmatrix} \boldsymbol{B}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } \boldsymbol{h} := \begin{pmatrix} \boldsymbol{b}_t \\ -\beta_t^1 \\ \vdots \\ -\beta_t^k \end{pmatrix}$$

We obtain

$$V(x) = \mathbb{E}\big[Q(x, \boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})\big] = \mathbb{E}\big[\hat{\mathcal{B}}_{t-1}(\underline{V}_t^k)(x_{t-1}, \boldsymbol{\xi}_t)\big] = \mathcal{B}_t(\underline{V}_{t+1}^k)(x_{t-1}). \tag{6.16a}$$

$$\text{with } Q(x, W, q, T, h) := \min_y \left\{ q^\top y \quad | \quad Tx + Wy = h, y \geqslant 0 \right\}. \tag{6.16b}$$

[9] Independence can be replaced by finite support assumption on one of the random variables. More generally, we can consider a finitely supported random variable $\boldsymbol{M}_t$ such that $\boldsymbol{c}_t$ and $(\boldsymbol{B}_t, \boldsymbol{b}_t)$ are independent *conditionally* to $\boldsymbol{M}_t$.

Under Assumption (LS) we have that i) $\boldsymbol{W}$ is finitely supported, and ii) $\boldsymbol{q}$ and $(\boldsymbol{T}, \boldsymbol{h})$ are independent. In particular, under Assumption (LS) we can apply Theorem 5.15 to find an explicit adapted partition. We now show that this adapted partition provides a valid and tight cut.

Proposition 6.9. *Let $\mathcal{P}$ be an adapted partition to $\check{x}$. For all $P \in \mathcal{P}$, let $\hat{f}_P$ be a valid and tight cut of $Q(\cdot, \mathbb{E}[\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}|P])$ at $\check{x}$. Then $f := \sum_{P\in\mathcal{P}} \mathbb{P}[P]\hat{f}_P$ is a valid and tight cut of V at $\check{x}$.*

Proof. We have

$$f(x) = \sum_{P\in\mathcal{P}} \mathbb{P}[P]\hat{f}_P(x) \tag{6.17a}$$
$$\leqslant \sum_{P\in\mathcal{P}} \mathbb{P}[P]Q(x, \mathbb{E}[\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}|P]) \qquad \text{since } \hat{f}_P \text{ is valid} \tag{6.17b}$$
$$= V_{\mathcal{P}}(x) \leqslant V(x) \qquad \text{since } V_{\mathcal{P}} \text{ is valid.} \tag{6.17c}$$

Thus, f is valid. Moreover,

$$f(\check{x}) = \sum_{P\in\mathcal{P}} \mathbb{P}[P]\hat{f}_P(\check{x}) \tag{6.18a}$$
$$= \sum_{P\in\mathcal{P}} \mathbb{P}[P]Q(\check{x}, \mathbb{E}[\boldsymbol{W}, \boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}|P]) \qquad \text{since } \hat{f}_P \text{ is tight} \tag{6.18b}$$
$$= V_{\mathcal{P}}(\check{x}) = V(\check{x}) \qquad \text{since } V_{\mathcal{P}} \text{ is tight.} \tag{6.18c}$$

Thus, f is tight. □

Remark 6.10. *As we saw in Section 6.2.2, this explicit adapted partition provides a new method to find tight and valid cuts in Line 10 of Algorithm 6.1 without having an approximation error,* i.e., $\underline{\gamma}_t = 0$*, in the linear case (under Assumption (LS)). Moreover, this explicit adapted partition allows to extend the scope of APM methods. Siddig and Song presented in [SS22] an algorithm combining ideas of APM methods and SDDP in the finitely supported case (APDP). The explicit adapted partition in Theorem 5.15 paves the way to a generalized APDP algorithm for non-finitely supported random variables (GAPDP).*

6.3 Complexity results

In this section, we give convergence and complexity results for various instances of Algorithm 6.1. In Section 6.3.1, we first define the notion of *effective iteration* and deduce an upper bound on the number of effective iterations required by Algorithm 6.1 to get an ε-solution. We then distinguish between deterministic and randomized selection processes for the choice in Line 5 of the algorithm. For deterministic selection processes, namely the problem-child and explorative node selections, we show in Section 6.3.2 that all iterations are effective. Finally, when the node selection is randomized, we show in Section 6.3.3 the existence of a positive probability for an iteration to be effective. We then deduce a complexity bound on the expected number of iterations.

6.3.1 Bounding the number of effective iterations

We first recall that the value of Problem (MSP) can be written in a more concise form, by using the nested problem in Eq. (6.2) and the definition of expected cost-to-go function in Eq. (6.4), and keeping in mind that ξ_1 is deterministic:

$$\text{val}\,(\text{MSP}) = \min_{x_1\in\mathcal{X}_1(x_0,\xi_1)} \ell_1(x_1, \xi_1) + V_1(x_1) \tag{6.19}$$

Our aim is to show that, for some iteration k, the solution x_1^k is a ε-solution of Eq. (6.19), and the lower-bound $\underline{V}_0(x_0)$ is ε-tight. Unfortunately, Assumptions 6.1 to 6.5 are not enough to ensure convergence of Algorithm 6.1: we need a further assumption on the node selection process.

Regardless of node selection, we define the notion of *effective iteration*. Recall that γ_t^F, $\underline{\gamma}_t$, $\overline{\gamma}_t$ are errors in forward Bellman operator and approximation update (see Algorithm 6.1) at time $t \in [T]$, and $\underline{L}_t$ (resp. $\overline{L}_t$) are Lipschitz bounds on the cuts (resp. upper-approximation) at time t. In the remains of the section we consider a sequence $(\overline{V}_t^k, \underline{V}_t^k, x_k^t)_{t\in[T],k\in\mathbb{N}}$ produced by Algorithm 6.1.

Definition 6.11 (effective iteration)**.** *For every $t \in [T-1]$, let $\delta_t > 0$ and $\eta_t \geqslant 0$. By backward induction, we define*

$$\varepsilon_{T-1} := \underline{\gamma}_{T-1} + \overline{\gamma}_{T-1} \tag{6.20a}$$

$$\varepsilon_t := \varepsilon_{t+1} + (\overline{L}_{t+1} + \underline{L}_{t+1})(\delta_{t+1} + \eta_{t+1}) + \gamma_{t+1}^F + \underline{\gamma}_t + \overline{\gamma}_t \qquad \forall t \in [T-2] \tag{6.20b}$$

$$\varepsilon_0 := \varepsilon_1 + (\overline{L}_1 + \underline{L}_1)(\delta_1 + \eta_1) + \gamma_1^F \tag{6.20c}$$

For $t \in [T-1]$ and $k \in \mathbb{N}$, we say that
x_t^k is ε_t-saturated*, if $\overline{V}_t^k(x_t^k) - \underline{V}_t^k(x_t^k) \leqslant \varepsilon_t$ and*
x_t^k is δ_t-distinguishable *if $\|x_t^k - x_t^\kappa\| > \delta_t$ for all $\kappa < k$ such that x_t^κ is ε_t-saturated.*
We say that an iteration $k \in \mathbb{N}$ is effective *if it generates either a ε_1-saturated point, which is also a ε_0-solution to Problem* (6.19)*, or a new ε_t-saturated and δ_t-distinguishable point for at least one $t \in [T]$,* i.e.*,*

$$x_1^k \text{ is } \varepsilon_1\text{-saturated and } \ell_1(x_1^k, \xi_1) + V_1(x_1^k) - \mathrm{val}\,(\mathrm{MSP}) \leqslant \varepsilon_0 \tag{6.21a}$$

$$OR \quad \exists t \in [T-1], \quad x_t^k \text{ is } \varepsilon_t\text{-saturated and } \delta_t\text{-distinguishable.} \tag{6.21b}$$

We now give an upper bound on the number of effective iterations of Algorithm 6.1 to find an ε_0- optimal solution.

Theorem 6.12 (bound on effective iterations number)**.** *Let Assumptions 6.1 to 6.5 be satisfied and $t \in [T-1]$, assume that $\delta_t \in [0, D_t]$ and $\eta_t \in \mathbb{R}_+$ are given and ε_t defined by* (6.20)*. Let*

$$\overline{K} := \sum_{t=1}^{T-1} \left(\frac{D_t}{\delta_t} + 1\right)^{n_t} \tag{6.22}$$

After at most $\overline{K} + 1$ effective *iterations we have a ε_1-lower bound of Problem* (MSP)*:*

$$\underline{V}_0^k(x_0) = \ell_1(x_1^k, \xi_1) + \underline{V}_1^k(x_1^k) \geqslant \mathrm{val}\,(\mathrm{MSP}) - \varepsilon_1 \tag{6.23}$$

Further, there exists, among those $\overline{K} + 1$ effective iterations, at least one such that x_1^k is an ε_0-solution to Problem (6.19)*:*

$$\ell_1(x_1^k, \xi_1) + V_1(x_1^k) \leqslant \mathrm{val}\,(\mathrm{MSP}) + \varepsilon_0 \tag{6.24}$$

Proof. For $t \in [T-1]$, there are at most $(\frac{D_t}{\delta_t} + 1)^{n_t}$ disjoint balls[10] of diameter δ_t in a ball of diameter $D_t + \delta_t$ (see [ZS19, A.3.2]). In particular, we cannot compute more than $(\frac{D_t}{\delta_t} + 1)^{n_t}$, δ_t-distinguishable points at step t. Thus, after $\overline{K} = \sum_{t=1}^{T-1}(\frac{D_t}{\delta_t} + 1)^{n_t}$ effective iterations, for all

[10] We consider here balls for the euclidean norm $\|\cdot\|_2$, but the result is still valid with the p-norm $\|\cdot\|_p$ for every $p \in [1, +\infty]$.

$t \in [T]$, it is impossible to compute a new δ_t-distinguishable point. Then, as the iteration k is effective and we cannot have (6.21b), we have (6.21a) and in particular x_1^k is an ε_0-solution Moreover, x_1^k is ε_1-saturated. Then,

$$\begin{aligned}
\ell_1(x_1^k, \xi_1) + \underline{V}_1^k(x_1^k) &\geqslant \ell_1(x_1^k, \xi_1) + \overline{V}_1^k(x_1^k) - \varepsilon_1 && \text{(6.25a)}\\
&\geqslant \ell_1(x_1^k, \xi_1) + V_1(x_1^k) - \varepsilon_1 && \text{(6.25b)}\\
&\geqslant \min_{x_1 \in \mathcal{X}_1(x_0, \xi_1)} \ell_1(x_1, \xi_1) + V_1(x_1) - \varepsilon_1 && \text{(6.25c)}\\
&= \mathrm{val}\,(\mathrm{MSP}) - \varepsilon_1 && \text{(6.25d)}
\end{aligned}$$

□

Remark 6.13. *Finally, although the theorems of this section state that we find an ε_0-optimal solution at stage 1, we have no guarantee that the approximations $\underline{V}_t^k$ converges to V_t. We cannot hope that these approximations converge to the true expected cost-to-go functions far from the optimal and reachable trajectories.*

Nevertheless, by considering the sets of points that are δ_t-close to every optimal and reachable trajectories, we could hope to have a convergence of strategies generated by $\mathcal{F}(\underline{V}_t^k)$ on those sets. If we add a finite diameter of the support of $\boldsymbol{\xi}_t$ and a Lipschitz assumptions for $\boldsymbol{\xi}_t$, we are confident that the proof can be adapted. However, the general case looks harder and might require different ideas for proving complexity results for the convergence of strategies at every stages.

For a class of specific (deterministic) implementations of Algorithm 6.1 each iteration is effective, in which case we can directly bound the number of iterations required to obtain an ε_0-optimal solution.

6.3.2 Deterministic node selection

In this section, we present sufficient condition for an iteration to be effective. Consequently, for two algorithms with deterministic node selections (namely problem-child node selection [BDZ17] and explorative node selection [Lan20]), we show that each iteration is effective, yielding a complexity result.

We first define the distance to the set of ε_t-saturated points.

Definition 6.14. *Let $t \in [T-1]$ and $k \geqslant 1$.*

We denote $\boldsymbol{y}_t^k$ the random variable

$$\boldsymbol{y}_t^k := \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \boldsymbol{\xi}_t) \tag{6.26}$$

We denote by d_t^k the distance function to the set of ε_t-saturated points until iteration k :

$$d_t^k(x) := \min_{\kappa < k | x_t^\kappa \text{ is } \varepsilon_t\text{-saturated}} \|x - x_t^\kappa\| \tag{6.27}$$

In particular, x_t^k is δ_t-distinguishable if and only if $d_t^k(x_t^k) > \delta_t$.

The following technical lemma, whose proof can be found in Section 6.B, shows that if the new state x_t^k (resulting from the choice of ξ_t^k) is either i) far enough from the set of saturated points, or ii) yielding a large enough gap, then iteration k is effective.

Lemma 6.15. *Let Assumptions 6.1 to 6.5 be satisfied and assume that, for all $t \in [T-1]$, $\delta_t \in [0, D_t]$, $\eta_t \in \mathbb{R}_+$ are given and ε_t defined by (6.20). Let $k \in \mathbb{N}^\star$. If, for all $t \in [T-1]$, at least one of the following inequalities is satisfied*

$$\mathbb{E}\,[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)] \leqslant \overline{V}_t^{k-1}(x_t^k) - \underline{V}_t^{k-1}(x_t^k) + (\overline{L}_t + \underline{L}_t)\eta_t \tag{6.28a}$$

$$\mathbb{E}\big[d_t^k(\boldsymbol{y}_t^k)\big] \leqslant d_t^k(x_t^k) + \eta_t \tag{6.28b}$$

then, iteration k is effective.

In [Lan20], Lan suggested to choose x_t^k as the most distinguishable point in a new algorithm called Explorative Dual Dynamic Programming (EDDP). We then speak of *explorative* node selection. The following lemma shows that both these selections lead to effective iterations.

Lemma 6.16. *Let Assumptions 6.1 to 6.5 holds and assume that, for all $t \in [T]$ $\delta_t \in [0, D_t]$, $\eta_t = 0$, are given and ε_t is defined by* (6.20)*.*

We say that we have a problem-child *node selection if for all $k \in \mathbb{N}^\star$, and $t \in [T-1]$, ξ_t^k is chosen such that it maximizes the current gap,* i.e.*,*

$$\xi_t^k \in \operatorname*{argmax}_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} \overline{V}_t^{k-1}\big(\mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi)\big) - \underline{V}_t^{k-1}\big(\mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi)\big) \tag{6.29a}$$

We say that we have an explorative *node selection if for all $k \in \mathbb{N}^\star$, and $t \in [T-1]$, ξ_t^k is chosen such that x_t^k maximizes the distance to previous ε_t-saturated points,* i.e.*,*

$$\xi_t^k \in \operatorname*{argmax}_{\xi \in \operatorname{supp}(\boldsymbol{\xi}_t)} d_t^k\big(\mathcal{F}_{t-1}(\underline{V}_{t+1}^{k-1})(x_{t-1}^k, \xi)\big) \tag{6.29b}$$

Then, with a problem-child *or an* explorative *node selection method, each iteration of Algorithm 6.1 is effective.*

Proof. It is a consequence of Lemma 6.15. Indeed, as $\eta_t = 0$, $x_t^k = \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi_t^k)$ and $\boldsymbol{y}_t^k = \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \boldsymbol{\xi}_t)$, and since the maximum is greater than the expected value, Eq. (6.28a) implies Eq. (6.29a) and Eq. (6.28b) implies Eq. (6.29b). □

Lemma 6.16 implies that every iteration of these deterministic node selection method is effective. Coupled with Theorem 6.12 we easily obtain complexity bounds, for example as follows.

Corollary 6.17. *Let Assumptions 6.1 to 6.5 holds and assume that every iteration of Algorithm 6.1 is effective (*e.g.*, problem-child or explorative node selection). Further, for simplicity, let the total error be $\gamma_\Sigma := \sum_{t=1}^{T-1} \underline{\gamma}_t + \overline{\gamma}_t + \gamma_t^F$ and choose n, D, L such that, for all $t \in [T-1]$, $n_t \leqslant n$, $D_t = D$, $\overline{L}_t = \underline{L}_t = L$. Then, for every $\varepsilon > \gamma_\Sigma$, sufficiently small (e.g. such that $\varepsilon \leqslant 2DL + \gamma_\Sigma$), Algorithm 6.1 finds an ε-first stage solution x_1^k within at most $\overline{K}_\varepsilon$ iterations where*

$$\overline{K}_\varepsilon := \left(\frac{2DL}{\varepsilon - \gamma_\Sigma}\right)^n (T-1)^{n+1} \tag{6.30}$$

Proof. We set $\delta_t := \frac{\varepsilon - \gamma_\Sigma}{2L(T-1)}$ and $\eta_t := 0$ for all $t \in [T-1]$. Then, as $\varepsilon \leqslant 2DL + \gamma_\Sigma$ we have $\delta_t \leqslant D = D_t$. Moreover, ε_0, defined in Eq. (6.20), satisfies

$$\varepsilon_0 = \sum_{t=1}^{T-1} \Big[(\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \underline{\gamma}_t + \overline{\gamma}_t + \gamma_t^F\Big] = (T-1)2L\delta + \gamma_\Sigma = \varepsilon. \tag{6.31}$$

With this setting, we have that $\overline{K}$, as defined in (6.22), satisfies

$$\overline{K} \leqslant (T-1)\left(\frac{D}{\delta} + 1\right)^n = T\left(\frac{2DL\,(T-1)}{\varepsilon - \gamma_\Sigma} + 1\right)^n$$

Now, as ε is assumed to be small enough to have $2DL/(\varepsilon - \gamma_\Sigma) \leqslant 1$ (i.e. $\varepsilon \leqslant 2DL + \gamma_\Sigma$) we get

$$\overline{K} \leqslant (T-1)\left(\frac{2DL(T-1)}{\varepsilon - \gamma_\Sigma}\right)^n = \overline{K}_\varepsilon.$$

By assumption all iterations are effective and Theorem 6.12 ends the proof. □

Remark 6.18. *Note that the maximum in* (6.29a) *(resp.* (6.29b)*) is easily obtained under finitely supported noise Assumption (FSN). Indeed, we can compute the gap (resp. the distance) for every ξ_t in the support of $\boldsymbol{\xi}_t$ and keep ξ_t^k maximizing the gap.*

However, without finite noise Assumption (FSN), we just need to find a ξ_t^k leading to a gap worse than the expected gap (see Lemma 6.15), and not necessarily a maximizer. This paves the way for a deterministic node selection, with non finitely supported random variables.

6.3.3 Randomized algorithms

When the choice of ξ_t^k is made randomly, there is no guarantee that the iteration will be effective. However, through a technical, yet necessary to deal with dependence issue, nested Hoeffding lemma's shown in Section 6.C.1, we show that there is a positive probability p for an iteration to be effective. Then, by comparing the time to obtain an effective iteration to a geometric random variable of probability of success p in Section 6.C.2, we deduce a bound on the expected number of iteration required to get an ε-optimal solution.

Remark 6.19 (Notational difficulty of randomized algorithm on stochastic problem)**.** *We are now considering a stochastic algorithm for solving the MSP Problem* (MSP)*. Thus, there are two sources of randomness: the intrinsic $(\boldsymbol{\xi}_t)_{t\in[T]}$ and the node selection $\xi_t^k = \widetilde{\xi}_t^k$. To distinguish both, we denote in bold random variables that are $(\boldsymbol{\xi}_t)_{t\in[T]}$ measurable, with a tilde random variables that are $(\widetilde{\xi}_t^k)_{t\in[T],k\in\mathbb{N}^*}$ measurable (and with both if they are neither).*

For example the trajectory determined during the forward phase $(\widetilde{x}_t^k)_{t\in[T]}$ only depends on the past node selections, whereas the tentative points $\widetilde{\boldsymbol{y}}_t^k$ depends both on the past node selections and the actual realization of $\boldsymbol{\xi}_t$.

Under Assumption (FSN), this discussion is usually avoided by representing the dependence on $(\boldsymbol{\xi}_t)_{t\in[T]}$ with a (finite) scenario tree, and indexing the variables by the tree nodes.

Let $(\mathcal{A}^k)_{k\in\mathbb{N}^\star}$ be the filtration such that $\mathcal{A}^k := \sigma(\widetilde{\xi}_t^\kappa)_{t\in[T-1],\kappa\in[k]}$, and $\mathcal{A}^\infty = \bigcup_{k\in\mathbb{N}} \mathcal{A}^k$. In particular, a random variable measurable with $\mathcal{A}^k$ knows all node selection up to iteration k, that include, for example, $\underline{V}_t^k$ for all $t \in [T]$.

Lemma 6.20. *Let Assumptions 6.1 to 6.5 be satisfied and assume that, for all $t \in [T-1]$, $\delta_t \in [0, D_t]$, $\eta_t \in \mathbb{R}_+$ are given and ε_t defined by* (6.20)*. Further, assume that in Algorithm 6.1, Line 5, we draw ξ_t^k randomly according to the distribution of $\boldsymbol{\xi}_t$, and independently of all other $\widetilde{\xi}_\tau^\kappa$ as well as $(\boldsymbol{\xi}_\tau)_{\tau\in[T-1]}$.*

Then, for all iteration $k \in \mathbb{N}$ of Algorithm 6.1 and all event $A^{k-1} \in \mathcal{A}^{k-1}$ such that $\mathbb{P}[A^{k-1}] > 0$, we have

$$\mathbb{P}\Big[\textit{Iteration } k \textit{ is effective.} \,\Big|\, A^{k-1}\Big] \geqslant \prod_{t=1}^{T} \left(1 - e^{\frac{-2\eta_t^2}{D_t^2}}\right) \tag{6.32}$$

Proof. Let $\mathcal{A}^\infty := \sigma(\widetilde{\xi}_t^\kappa)_{t\in[T-1],\kappa\in\mathbb{N}^\star}$ and $k \in \mathbb{N}^\star$. By Lemma 6.15, we have

$$\mathbb{P}[\text{Iteration } k \text{ is effective.} \,|\, A^{k-1}] \geqslant \mathbb{P}\Big[\forall t \in [T-1],\, \mathbb{E}[\widetilde{d}_t^k(\widetilde{\boldsymbol{y}}_t^k) \,|\, \mathcal{A}^\infty] < \widetilde{d}_t^k(\widetilde{x}_t^k) + \eta_t \,|\, A^{k-1}\Big] \tag{6.33}$$

For $t \in [T-1]$, let $\mathcal{A}_t^k := \sigma(\mathcal{A}^{k-1}, (\widetilde{\xi}_k^t)_{\tau\in[t]})$. We have that $\sigma(\widetilde{d}_t^k(\widetilde{\boldsymbol{y}}_t^k)) \subset \sigma(\mathcal{A}_{t-1}^k, \boldsymbol{\xi}_t)$ from which we deduce that $\mathbb{E}[\widetilde{d}_t^k(\widetilde{\boldsymbol{y}}_t^k) \,|\, \mathcal{A}^\infty] = \mathbb{E}[\widetilde{d}_t^k(\widetilde{\boldsymbol{y}}_t^k) \,|\, \mathcal{A}_{t-1}^k]$. We define the events $E_t^k := \{\omega \in \Omega \mid \mathbb{E}[\widetilde{d}_t^k(\widetilde{\boldsymbol{y}}_t^k) \mid \mathcal{A}_{t-1}^k] < \widetilde{d}_t^k \widetilde{x}_t^k + \eta_t\}$. Thus, $\mathbb{P}[\text{Iteration } k \text{ is effective} \,|\, A^{k-1}] \geqslant \mathbb{P}[\bigcap_{t=1}^{T-1} E_t^k \,|\, A^{k-1}]$ By applying Lemma 6.30 with the random variables $(\widetilde{\xi}_t^k)_{k\in\mathbb{N},t\in[T-1]}$, the filtration $(\mathcal{A}_t^k)_{k\in\mathbb{N},t\in[T-1]}$

and the measurable function $f_t^k : ((\xi_\tau^\kappa)_{\tau\in[T-1],\kappa\in[k-1]}, (\xi_\tau^k)_{\tau\in[t]}) \mapsto d_t^k(x_t^k)$ taking its value in $[0, D_t]$, we have $\mathbb{P}\Big[\bigcap_{t=1}^{T-1} E_t^k \,\Big|\, A^{k-1}\Big] \geqslant \prod_{t=1}^{T}\left(1 - e^{\frac{-2\eta_t^2}{D_t^2}}\right)$ which gives Eq. (6.32). □

We now give a complexity results for all TFDP algorithms (following framework of Algorithm 6.1) where the choice of ξ_t^k is made randomly.

Theorem 6.21. *Let Assumptions 6.1 to 6.5 be satisfied and assume that in Line 5, we draw ξ_t^k randomly according to the distribution of $\boldsymbol{\xi}_t$, and independently from the previous ξ_τ^κ.*

Further, for simplicity, let the total error be $\gamma_\Sigma := \sum_{t=1}^{T-1} \underline{\gamma}_t + \overline{\gamma}_t + \gamma_t^F$ and choose n, D, L such that, for all $t \in [T-1]$, $n_t \leqslant n$, $D_t = D$, $\overline{L}_t = \underline{L}_t = L$.

*Then, for $\varepsilon > \gamma_\Sigma$, sufficiently small (*e.g., *such that $\varepsilon \leqslant 4DL + \gamma_\Sigma$), the expected number of iterations of Algorithm 6.1 required to find an ε-solution x_1^k to problem* (6.19), i.e. *, such that $\ell_1(x_1^k, \xi_1) + V_2(x_1^k) \leqslant \operatorname{val}(\mathrm{MSP}) + \varepsilon$ is bounded by $(T-1)\left(\frac{4DL(T-1)}{\varepsilon-\gamma_\Sigma}\right)^{n+2(T-1)}$.*

Proof. We set for all $t \in [T-1]$, $\delta_t = \eta_t = \frac{\varepsilon-\gamma_\Sigma}{4L(T-1)}$. Then, as $\varepsilon \leqslant 4DL + \gamma_\Sigma$ we have $\eta_t = \delta_t \leqslant D = D_t$. Moreover, ε_0, defined in Eq. (6.20), satisfies

$$\varepsilon_0 = \sum_{t=1}^{T-1}\Big[(\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \underline{\gamma}_t + \overline{\gamma}_t + \gamma_t^F\Big] = (T-1)2L \times 2\frac{\varepsilon - \gamma_\Sigma}{4L(T-1)} + \gamma_\Sigma = \varepsilon.$$

Let $\widetilde{K}$ the (random) number of iterations needed to compute $\overline{K}_\varepsilon := \sum_{t=1}^{T-1}\left(1 + \frac{D_t}{\delta_t}\right)^{n_t} \leqslant \left(\frac{4DL}{\varepsilon-\gamma_\Sigma}\right)^n (T-1)^{n+1}$ effective iterations, then by Theorem 6.12, Algorithm 6.1 finds an ε-solution after at most $\widetilde{K}$ iterations. Let $p := \prod_{t=1}^{T}\left(1 - \exp\left(\frac{-2\eta_t^2}{D_t^2}\right)\right)$, by Lemma 6.20, for $A^{k-1} \in \mathcal{A}^{k-1}$, we have $\mathbb{P}\big[\text{Iteration } k \text{ is effective} \,\big|\, A^{k-1}\big] \geqslant p$. Thus, by Lemma 6.31, we have $\mathbb{E}\big[\widetilde{K}\big] \leqslant \frac{K_\varepsilon}{p}$.

Moreover, since as $x \mapsto \frac{x}{1-e^{-x}}$ is an increasing function on $(0, \frac{1}{2}]$, then for all $x \in (0, 1]$, we have $\frac{1}{1-e^{-x}} \leqslant \frac{1}{1-e^{-1}} \times \frac{1}{x} \leqslant 1.6 \times \frac{1}{x}$. Thus, as $\frac{2\eta_t^2}{D_t^2} \leqslant 1$, we have that $\frac{1}{p} \leqslant \prod_{t=1}^{T-1} 1.6 \times \frac{D_t^2}{2\eta_t^2} \leqslant \left(\frac{4DL(T-1)}{\varepsilon-\gamma_\Sigma}\right)^{2(T-1)}$. We then obtain $\mathbb{E}\big[\widetilde{K}\big] \leqslant \frac{K_\varepsilon}{p} \leqslant \left(\frac{4DL}{\varepsilon-\gamma_\Sigma}\right)^n (T-1)^{n+1} \times \left(\frac{4DL(T-1)}{\varepsilon-\gamma_\Sigma}\right)^{2(T-1)} = (T-1)\left(\frac{4DL(T-1)}{\varepsilon-\gamma_\Sigma}\right)^{n+2(T-1)}$.

□

Remark 6.22 (Stochastic dominance and comparison with finitely supported noise)**.** *The proof of Theorem 6.21 actually give more information on the (random) number of iteration $\widetilde{K}$ after which we obtain an ε-solution: $\widetilde{K}$ is stochastically dominated by a random variable with a negative binomial distribution representing the number of trials to obtain $\overline{K}_\varepsilon$ successes with probability of success p, (see Lemma 6.31).*

Further, under finitely supported noise assumption (FSN), the probability of choosing the problem child ξ_t^k (cf Lemma 6.16) for each $t \in [T-1]$ is lower bounded by $\prod_{t=1}^{T-1}\frac{1}{|\operatorname{supp}(\boldsymbol{\xi}_t)|}$. Then, Lemma 6.20 still holds after replacing the right hand side probability of success by $\prod_{t=1}^{T-1}\frac{1}{|\operatorname{supp}(\boldsymbol{\xi}_t)|}$. We can then deduce other complexity bounds under (FSN). For example, in [Lan20], assuming that $|\operatorname{supp}(\boldsymbol{\xi}_t)| \leqslant N$, for all $t \in [T-1]$, the probability of having an effective iteration is bounded by $\frac{1}{N^{T-1}}$.

6.4 Extension to risk-averse setting

We now briefly discuss extensions involving a maximization problem in the dynamic programming equation, arising for example from multistage risk-averse, robust or distributionally robust

problems. Algorithm 6.1 can be adapted to such problems, by changing the definitions of the Bellman operators.

Further, in the risk neutral case, Algorithm 6.1 is not symmetrical in its treatment of lower and upper approximations. As noted in Remark 6.3, for a minimization problem, in Algorithm 6.1, the forward phase Line 6 should be done using the lower approximations $\underline{V}_t^k$. More generally, one should use an *outer approximation* (that is under approximation for min sub-problems and upper approximations for max sub-problems) during the forward phase to be able to explore the state space. Thus, for those min-max problems the computation of upper-approximations $\overline{V}_t^k$ is not optional.

Minimax problems. Baucke, Downward and Zakeri, in [BDZ17], presented a convergent problem-child algorithm to solve stochastic minimax dynamic programs. Although our framework of Algorithm 6.1 do not handle such minimax problem, we can extend it to do so. More precisely, we consider a problem where the decision maker chooses $x_t \in \mathcal{X}_t(x_{t-1}, y_{t-1}, \xi_t)$, and then an adversary chooses $y_t \in \mathcal{Y}_{t-1}(x_{t-1}, y_{t-1}, x_t, \xi_t)$. Thus, the Bellman operators are now defined as

$$\mathcal{B}_{t-1}(\widetilde{V})(x_{t-1}, y_{t-1}) = \mathbb{E}\big[\min_{x_t \in \mathcal{X}_t(x_{t-1}, y_{t-1}, \boldsymbol{\xi}_t)} \max_{y_t \in \mathcal{Y}_{t-1}(x_{t-1}, y_{t-1}, x_t, \boldsymbol{\xi}_t)} \ell_t(x_t, y_t, \boldsymbol{\xi}_t) + \widetilde{V}(x_t, y_t)\big]. \quad (6.34)$$

The reachable sets then become

$$X_0^r = \{x_0\} \qquad Y_0^r = \{y_0\} \quad (6.35a)$$

$$X_t^r = \bigcup_{x_{t-1} \in X_{t-1}^r} \bigcup_{y_{t-1} \in Y_{t-1}^r} \bigcup_{\xi_t \in \Xi_t} \mathcal{X}_t(x_{t-1}, y_{t-1}, \xi_t) \qquad \forall t \in [T]. \quad (6.35b)$$

$$Y_t^r = \bigcup_{x_{t-1} \in Y_{t-1}^r} \bigcup_{y_{t-1} \in Y_{t-1}^r} \bigcup_{x_t \in X_t^r} \bigcup_{\xi_t \in \Xi_t} \mathcal{Y}_t(x_{t-1}, y_{t-1}, x_t, \xi_t) \qquad \forall t \in [T]. \quad (6.35c)$$

In the forward phase, as in Algorithm 6.1, the γ_t^F-optimal solution x_t^k should be chosen thanks to the approximation $\underline{V}_t^{k-1}$. However, as we maximize over y_t, y_t^k must be a γ_t^F-optimal solution of the step problem with the approximation $\overline{V}_t^{k-1}$:

$$x_t^k = \mathcal{F}_{t-1}^{\min}(\underline{V}_t^{k-1})(x_{t-1}, y_{t-1}, \xi_t^k) \in \gamma_t^F\text{-}\operatorname*{arg\,min}_{x_t \in \mathcal{X}_t(x_{t-1}, y_{t-1}, \xi_t^k)} \max_{y_t \in \mathcal{Y}_{t-1}(x_{t-1}, y_{t-1}, x_t, \xi_t^k)} \ell_t(x_t, y_t, \xi_t^k) + \underline{V}_t^{k-1}(x_t, y_t) \quad (6.36a)$$

$$y_t^k = \mathcal{F}_{t-1}^{\max}(\overline{V}_t^{k-1})(x_{t-1}, y_{t-1}, x_t^k, \xi_t^k) \in \gamma_t^F\text{-}\operatorname*{argmax}_{y_t \in \mathcal{Y}_{t-1}(x_{t-1}, y_{t-1}, x_t^k, \xi_t^k)} \ell_t(x_t, y_t, \xi_t^k) + \overline{V}_t^{k-1}(x_t, y_t) \quad (6.36b)$$

Assuming that the reachable sets X_t^r and Y_t^r have finite dimensions d_x and d_y and diameter D, and that the objective function are L-Lipschitz, the convergence and complexity results still hold developing on the ideas of [ZS19]. The upper bound on the number of effective iterations then becomes $K_\varepsilon := \left(\frac{2DL}{\varepsilon - \gamma_\Sigma}\right)^{d_x + d_y} (T-1)^{d_x + d_y + 1}$.

Robust Closely related, in [GTW19], Georghiou, Tsoukalas and Wiesemann presented the Robust Dual Dynamic Programming algorithm (RDDP) to solve multistage robust optimization problems. In such problems, instead of minimizing the expectation like in Eq. (MSP), we minimize considering the worst case scenario $\xi_t \in \Xi_t$. In this setting, the Bellman operator reads

$$\mathcal{B}_{t-1}(\widetilde{V}) = \max_{\xi_t \in \Xi_t} \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \ell_t(x_t, \xi_t) + \widetilde{V}(x_t). \quad (6.37)$$

Note that this robust setting can be seen as a particular case of minimax problems where we have deterministic random variables. Indeed, if we invert the order of max and min, either by changing the indices or by taking the opposite, and Eq. (6.37) can be written as Eq. (6.34) where ξ_t of (6.37) plays the role of y_t and the $\boldsymbol{\xi}_t$ of (6.34) are deterministic parameter. The upper bound on the number of effective iterations then becomes $K_\varepsilon := \left(\frac{2DL}{\varepsilon-\gamma_\Sigma}\right)^{d_x+d_\xi}(T-1)^{d_x+d_\xi+1}$.

Risk averse Multistage stochastic problems in the risk averse setting are MSP where the expectation is replaced by a multiperiod risk measure. In the nested coherent risk measure framework we present conditions under which Algorithm 6.1 can be adapted.

Let ρ be a coherent risk measure (see [ADEH99, ADE$^+$07] or [SDR14, Def 6.4]) the Bellman operator in the risk averse setting reads

$$\mathcal{B}_{t-1}(\widetilde{V}) : x_{t-1} \mapsto \rho\Big(\min_{x_t \in \mathcal{X}_t(x_{t-1},\boldsymbol{\xi}_t)} \ell_t(x_t,\boldsymbol{\xi}_t) + \widetilde{V}(x_t)\Big). \tag{6.38}$$

We recall a classical Fenchel representation theorem for proper, lower semicontinuous, law-invariant, coherent risk measure (see [SDR14, Thm 6.5]). For every random variable $\boldsymbol{z} \in L_1(\Omega,\mathcal{A},\mathbb{P},\mathbb{R})$, we have

$$\rho(\boldsymbol{z}) = \max_{\boldsymbol{y}\in\mathfrak{A}_\rho} \mathbb{E}_\mathbb{P}[\boldsymbol{y}\boldsymbol{z}] \tag{6.39}$$

where $\mathfrak{A}_\rho := \{\boldsymbol{y} \in L_\infty(\Omega,\mathcal{A},\mathbb{P},\mathbb{R}) \mid \mathbb{E}[\boldsymbol{y}] = 1, \quad \boldsymbol{y} \geqslant 0 \ a.s., \quad \mathbb{E}[\boldsymbol{y}\boldsymbol{z}'] \leqslant \rho(\boldsymbol{z}), \quad \forall \boldsymbol{z}' \in L_1(\Omega,\mathcal{A},\mathbb{P},\mathbb{R})\}$.

With this representation, we get

$$\mathcal{B}_{t-1}(\widetilde{V}) = \max_{\boldsymbol{y}\in\mathfrak{A}_\rho} \mathbb{E}_\mathbb{P}\Big[\min_{x_t \in \mathcal{X}_t(x_{t-1},\boldsymbol{\xi}_t)} \boldsymbol{y}\,\ell_t(x_t,\boldsymbol{\xi}_t) + \boldsymbol{y}\widetilde{V}(x_t)\Big]. \tag{6.40}$$

Up to a slight change of notation, we can write this problem as a minimax problem. In particular, a sufficient condition to obtain convergence and complexity bounds for risk averse MSP is that the set $\mathfrak{A}_\rho$ has a finite dimension and a finite diameter. For example, if Ω is finite, $\mathfrak{A}_\rho$ is contained in the space of random variables in Ω, isomorphic to a simplex of dimension $|\Omega|-1$ which has finite diameter. More generally, if $\mathfrak{A}_\rho$ is contained in the convex hull of n random variables $(\boldsymbol{y}_k)_{k\in[n]}$, then $\mathfrak{A}_\rho$ has a finite diameter smaller than $\max_{k,\ell\in[n]}(\|\boldsymbol{y}_k-\boldsymbol{y}_\ell\|_\infty)$ and a finite dimension smaller than $n-1$. Thus, we obtain complexity results similar to Corollary 6.17 and Theorem 6.21 with $K_\varepsilon := \left(\frac{2DL}{\varepsilon-\gamma_\Sigma}\right)^{d+n-1}(T-1)^{d+n}$.

We now comment the particular case of the average value at risk [RU$^+$00] with value $\alpha \in [0,1)$, denoted $AV@R_\alpha$ and defined as:

$$AV@R_\alpha(\boldsymbol{z}) := \inf_{s\in\mathbb{R}}\Big\{s + \frac{1}{1-\alpha}\mathbb{E}_\mathbb{P}\big[\max(\boldsymbol{z}-s,0)\big]\Big\} \tag{6.41}$$

We cannot use the dual representation Eq. (6.39) to derive complexity bounds as $\mathfrak{A}_{AV@R}$ has, in general, non-finite dimension. However note that in Eq. (6.41), since $AV@R_\alpha(\boldsymbol{z}) \leqslant \frac{\mathbb{E}_\mathbb{P}[\boldsymbol{z}]}{1-\alpha}$ the infimum on s over $\mathbb{R}$ can be replaced by a minimum on the compact interval $[0, \frac{1}{1-\alpha}\mathbb{E}_\mathbb{P}[\boldsymbol{z}]]$. To obtain an upper bound that does not depend on k and x_{t-1}, we set $\boldsymbol{z} = \min_{x_t\in\mathcal{X}(x_{t-1},\boldsymbol{\xi}_t)} \ell_t(x_t,\boldsymbol{\xi}_t) + \underline{V}_t^k(x_t)$ then $\mathbb{E}_\mathbb{P}[\boldsymbol{z}]$ is upper bounded by $\min_{x_t\in X_t^r}\mathbb{E}\big[\ell_t(x_t,\boldsymbol{\xi}_t) + \overline{V}_t^1(x_t)\big]$ which has a finite value by Assumption 6.3. Thus, MSP with nested average value at risk measure can be handled by this framework and we can obtain complexity results similar to Corollary 6.17 and Theorem 6.21 with $K_\varepsilon := \left(\frac{2DL}{\varepsilon-\gamma_\Sigma}\right)^{d+1}(T-1)^{d+2}$.

Acknowledgements We sincerely thank Mr Julien Weibel for interesting and useful discussions concerning Section 6.C.2. This research benefited from the support of the FMJH Program Gaspard Monge for optimization and operations research and their interactions with data science.

6.A Cut methodologies

In this section, for the sake of completeness, we give several cuts that are used in different algorithms to solve particular multistage problems.

Cut	Oracle needed	Setting and avantages
Benders	First order	Convex, simple to implement
Reverse norm	Zeroth order and Lipschitz constant	Lipschitz
Step	Zeroth order, ε and γ	Monotonic
Lagrangian	Solving a dual problem	Problem with small duality gap
Integer	Zeroth order	Binary variables
Adaptive partition	Adapted partition oracle	Linear, whole tangent cone
Generalized conjugacy	Conjugate computation	Regularisation
Saddle	First order and Lipschitz constant	Minimax problems
Fenchel conjugate	Fenchel Dual of Bellman equation	Linear, Exact upper bound

Table 6.2: Synthesis of different cuts and oracle required

6.A.1 Benders cuts for convex functions

The most commonly used cuts are the Benders cuts which are affine functions. This kind of cut only works if the expected cost-to-go functions are *convex*.

The word cut is actually used because the graph of a Benders cut is a hyperplane which is tangent to the epigraph of the approximated function.

Proposition 6.23. *Let F be a* convex *function and $g \in \partial F(x)$ a subgradient of F. We define the Benders cut f at $\hat{x}$ as*

$$f(x) := F(\hat{x}) + g^\top (x - \hat{x}) \tag{6.42}$$

Then, f is valid and tight at $\hat{x}$.

Proof. By definition of a subgradient, $f(x) := F(\hat{x}) + g^\top (x - \hat{x}) \leqslant F(x)$, thus f is valid. By definition of f, $f(\hat{x}) = F(\hat{x}) + g^\top (\hat{x} - \hat{x}) = F(\hat{x})$ thus f is tight at $\overline{x}$. □

We see that a first order oracle for the function F, *i.e.*, an oracle that returns the value $F(\hat{x})$ and a subgradient $g \in \partial F(\hat{x})$ for an input $\hat{x}$, provides a direct algorithm to compute Benders cut.

6.A.2 Reverse norm cuts for Lipschitz functions

Stochastic Lipschitz Dynamic Programming (SDLP) presented in [ACdC20] provides an algorithm to deal with non-convex Lipschitz multistage stochastic programs. In this setting, the cost functions ℓ_t are simply assumed to be Lipschitz continuous. Thus, the expected cost-to-go functions F_t is not necessarily convex and Benders cuts are not valid anymore. Ahmed, Cabral

and da Costa replaced these cuts by reverse norm cuts or augmented lagrangian cuts using only the Lipschitz property of expected cost-to-go functions.

Proposition 6.24. *Let F be a function with Lipschitz constant L (for norm $\|\cdot\|$). We define the reverse norm cut f of F at $\overline{x}$ as*

$$f(x) := F(\hat{x}) - L\|x - \hat{x}\| \tag{6.43}$$

Then, f is valid and 0-tight at $\hat{x}$.

Proof. For any given x and $\hat{x}$

$$\begin{aligned} f(x) &= F(\hat{x}) - L\|x - \hat{x}\| && \text{(6.44a)}\\ &= F(\hat{x}) - F(x) + F(x) - L\|x - \hat{x}\| && \text{(6.44b)}\\ &\leqslant L\|\hat{x} - x\| + F(x) - L\|x - \hat{x}\| && \text{(6.44c)}\\ &= F(x) && \text{(6.44d)} \end{aligned}$$

Thus, f is valid. By definition of f, $f(\hat{x}) = F(\hat{x}) - L\|\hat{x} - \hat{x}\| = F(\hat{x})$ thus f is tight at $\overline{x}$. □

We see that a zeroth order oracle for the function F, *i.e.*, an oracle that returns the value $F(\hat{x})$ for an input $\hat{x}$, together with a Lipschitz constant L provides a direct algorithm to compute reverse norm cuts. Thus, SDLP integrates our framework in Algorithm 6.1.

We can also define the norm cut $f(x) := F(\hat{x}) + L\|x - \hat{x}\|$. These norm cuts can be used to compute $\overline{V}_t^k$. The algorithm Tropical dynamic programming in [ACT20] uses this upper cuts together with Benders cuts for lower approximation $\underline{V}_t^k$ and thus integrates the framework of Algorithm 6.1.

6.A.3 Step cuts for monotonic functions

We now look at "almost monotonic" expected cost-to-go functions.

Proposition 6.25. *Let F be a function such that there exists $\delta > 0$ and $\gamma \geqslant 0$ with*

$$\forall x, y, \quad x \leqslant y + \delta\mathbf{1} \implies F(x) \leqslant F(y) + \gamma \tag{6.45}$$

where $\mathbf{1}$ is the vector whose coefficients are all equal to 1. We assume that F is upper bounded by $\overline{M}$.

For a point $\hat{x}$, we define the upper increasing step cut *f as*

$$f(x) := \begin{cases} F(\hat{x}) + \gamma & \text{if } x \leqslant \hat{x} + \delta\mathbf{1} \\ \overline{M} & \text{otherwise} \end{cases} \tag{6.46}$$

Then, the upper increasing step cut f satisfies

$$\begin{aligned} f(x) &\geqslant F(x), \quad \forall x && \text{(6.47)}\\ f(\check{x}) &\leqslant F(\check{x}) + \gamma && \text{(6.48)} \end{aligned}$$

Proof. Since $\hat{x} \leqslant \hat{x} + \delta$, $f(\hat{x}) = F(\hat{x}) + \gamma$. Moreover, if $x \leqslant \hat{x} + \delta\mathbf{1}$, by Eq. (6.45) $F(x) \leqslant F(\hat{x}) + \gamma = f(x)$ and otherwise $f(x) = \overline{M} \geqslant F(x)$. □

We could also define lower increasing step cuts for functions verifying Eq. (6.45). This cut methods also adapt to lower bounded decreasing functions, we will define in the same way upper and lower decreasing step cuts. However, these cuts are not Lipschitz. To integrate the

framework of Algorithm 6.1, we could adapt these step cuts by interpolating with affine functions between the constant regions of the cuts.

In [PWB20], Philpott, Wahid and Bonnans presented an algorithm called mixed integer dynamic approximation scheme (MIDAS). This method applies to multistage mixed integer problems, given as a maximization problems. After adapting the problem by taking the opposite of the expected cost-to-go function, adding the constant γ and choosing the right affinely interpolated step cut, the algorithm MIDAS integrates the framework of Algorithm 6.1 with step cuts.

6.A.4 Lagrangian cuts

Lagrangian cuts were introduced for TFDP by Zou, Ahmed and Sun in [ZAS19]. These cuts are based on the Lagrangian relaxation of an optimisation problems.

Proposition 6.26. *Let F be a function, H be a convex function and $x \mapsto Y(x)$ be a graph-convex set valued mapping see [RW09, p155] such that F is defined as*

$$F(x) := \inf_{y \in Y(x)} \ell(y) \tag{6.49}$$

We define the Lagrangian cut f at $\hat{x}$ as

$$f(x) := \hat{\lambda}^\top x + \hat{\beta} \tag{6.50}$$

where

$$\hat{\lambda} \in \operatorname*{argmax}_{\lambda} \lambda^\top \hat{x} + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \lambda^\top z \tag{6.51a}$$

$$\hat{\beta} = \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \hat{\lambda}^\top z \tag{6.51b}$$

Then, the Lagrangian cut is valid and tight at $\hat{x}$.

Proof. Consider $x \in \operatorname{dom}(Y)$. We rely on a strong duality result:

$$\begin{aligned} F(x) &= \inf_{y \in Y(x)} \ell(y) && \text{(6.52a)} \\ &= \inf_{y,z \,|\, y \in Y(z) \text{ and } z = x} \ell(y) && \text{(6.52b)} \\ &= \inf_{y,z \,|\, y \in Y(z)} \max_{\lambda} \ell(y) + \lambda^\top (x - z) && \text{(6.52c)} \\ &= \max_{\lambda} \lambda^\top x + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \lambda^\top z && \text{(6.52d)} \\ & && \text{(6.52e)} \end{aligned}$$

Indeed, as Y is graph-convex, we have that $\{(y, z) \,|\, y \in Y(z)\}$ is a non-empty convex set. Thus, we have

$$\begin{aligned} f(\hat{x}) &= \hat{\lambda}^\top \hat{x} + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \hat{\lambda}^\top z && \text{(6.53a)} \\ &= \max_{\lambda} \lambda^\top \hat{x} + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \lambda^\top z && \text{(6.53b)} \\ &= F(x) && \text{(6.53c)} \end{aligned}$$

and f is tight at $\hat{x}$. Moreover,

$$f(x) = \hat{\lambda}^\top x + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \hat{\lambda}^\top z \tag{6.54a}$$

$$\leqslant \max_{\lambda} \lambda^\top x + \inf_{y,z \,|\, y \in Y(z)} \ell(y) - \lambda^\top z \tag{6.54b}$$

$$= F(x) \tag{6.54c}$$

and then f is valid. □

Note that this result is still true without the convexity assumption if we replace the tightness result by a lower γ-tighness result where γ is the duality gap.

Secondly, in this simplified setting and when the variable x takes value in a continuous space, the Lagrangian cut can be seen as a Benders cut since $\hat{\lambda}$ is a subgradient of F at $\hat{x}$. Nevertheless, the point of view of Lagrangian allows new ideas to compute cuts. In particular, one can define the Lagrangian cut with a different relaxation to deal with more complex setting such as integer cases as presented in [ZAS19]. We can also combine the Lagrangian cut with the reverse norm cut, such ideas are presented under the name *augmented Lagrangian cuts* in [ACdC20]. Thus, the algorithm SLDP from [ACdC20] integrates the framework of Algorithm 6.1.

Zhang and Sun in [ZS19] generalized these Lagrangian cuts by introducing the point of view of generalized conjugacy (see [RW09, Chapter 11]). They called these new cuts *generalized conjugacy cuts*, and proved them to be tight and valid, thanks to the Fenchel-Young inequality.

6.A.5 Integer optimality cuts

The integer optimality cuts where first introduced by Laporte and Louveaux in [LL93] for 2-stage sochastic integer problems.

Proposition 6.27. *Let F be a function taking value in $\{0,1\}^d$ and $\hat{x} \in \{0,1\}^d$ be a binary vector with $\hat{S} := \{i \,|\, x_i = 1\}$. We assume that F is lower bounded by $\underline{M}$.*

We define the integer optimality cut *f as*

$$f(x) := (F(\hat{x}) - \underline{M})\Big(\sum_{i \in \hat{S}} x_i - \sum_{i \notin \hat{S}} x_i - |\hat{S}| + 1\Big) + \underline{M} \tag{6.55}$$

Then, f is a valid and tight at $\hat{x}$.

Proof. We have that $f(\hat{x}) = (F(\hat{x}) - \underline{M})(|\hat{S}| - 0 - |\hat{S}| + 1) + \underline{M} = F(\hat{x})$. Thus, f is tight at $\hat{x}$. Let $x \in \{0,1\}^d$ different from $\hat{x}$, we have $\sum_{i \in \hat{S}} x_i - \sum_{i \notin \hat{S}} x_i \leqslant |\hat{S}| - 1$. Then, $f(x) \leqslant \underline{M} \leqslant F(x)$ and thus f is valid. □

In [ZAS19], Zou, Ahmed and Sun presented an algorithm called Stochastic dual dynamic integer programming (SDDiP), suggesting to use integer optimality cuts or Lagrangian cuts instead of classical Benders cuts. By tightness and validity of these cuts, the algorithm SDDiP integrates the framework of Algorithm 6.1 (with a potentially large Lipschitz constant).

6.B Sufficient conditions for effective iterations

In this appendix, we want to prove Lemma 6.15. We start with a technical lemma linking the gap at $t-1$ with the expected gap for tentative points at t.

Lemma 6.28. *Let Assumptions 6.1 to 6.5 be satisfied and $t \in [T-1]$, assume that $\delta_t \in [0, D_t]$ and $\eta_t \in \mathbb{R}_+$ are given and ε_t defined by (6.20). Then, for all algorithms satisfying the framework of Algorithm 6.1, we have for $t \in [T-1]$*

$$0 \leqslant \overline{V}_{t-1}^k(x_{t-1}^k) - \underline{V}_{t-1}^k(x_{t-1}^k) \leqslant \mathbb{E}[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)] + \underline{\gamma}_{t-1} + \overline{\gamma}_{t-1} + \gamma_t^F \tag{6.56}$$

where we recall that $\boldsymbol{y}_t^k := \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \boldsymbol{\xi}_t)$.

Proof.

$$\begin{aligned}
\underline{V}_{t-1}^k(x_{t-1}^k) &\geqslant \mathcal{B}_{t-1}^k(\underline{V}_t^k)(x_{t-1}^k) - \underline{\gamma}_{t-1} && \text{(backward phase: } \underline{\gamma}_{t-1}\text{-tight cut)} && \text{(6.57a)}\\
&= \mathbb{E}\big[\min_{x \in \mathcal{X}_t(x_{t-1}^k, \boldsymbol{\xi}_t)} \ell_t(x, \boldsymbol{\xi}_t) + \underline{V}_t^k(x)\big] - \underline{\gamma}_{t-1} && \text{(definition of } \mathcal{B}_t) && \text{(6.57b)}\\
&\geqslant \mathbb{E}\big[\min_{x \in \mathcal{X}_t(x_{t-1}^k, \boldsymbol{\xi}_t)} \ell_t(x, \boldsymbol{\xi}_t) + \underline{V}_t^{k-1}(x)\big] - \underline{\gamma}_{t-1} && \text{(monotonicity of approx.)} && \text{(6.57c)}\\
&\geqslant \mathbb{E}\big[\ell_t(\boldsymbol{y}_t^k, \boldsymbol{\xi}_t) + \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] - \gamma_t^F - \underline{\gamma}_{t-1} && \text{(definition of } \mathcal{F}_{t-1}) && \text{(6.57d)}
\end{aligned}$$

$$\begin{aligned}
\overline{V}_{t-1}^k(x_{t-1}^k) &\leqslant \mathcal{B}_{t-1}^k(\overline{V}_t^k)(x_{t-1}^k) + \overline{\gamma}_{t-1} && \text{(backward phase)} && \text{(6.58a)}\\
&= \mathbb{E}\big[\min_{x \in \mathcal{X}_t(x_{t-1}^k, \boldsymbol{\xi}_t)} \ell_t(x, \boldsymbol{\xi}_t) + \overline{V}_t^k(x)\big] + \overline{\gamma}_{t-1} && \text{(definition of } \mathcal{B}_t) && \text{(6.58b)}\\
&\leqslant \mathbb{E}\big[\min_{x \in \mathcal{X}_t(x_{t-1}^k, \boldsymbol{\xi}_t)} \ell_t(x, \boldsymbol{\xi}_t) + \overline{V}_t^{k-1}(x)\big] + \overline{\gamma}_{t-1} && \text{(monotonicity of approx.)} && \text{(6.58c)}\\
&\leqslant \mathbb{E}\big[\ell_t(\boldsymbol{y}_t^k, \boldsymbol{\xi}_t) + \overline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] + \overline{\gamma}_{t-1} && \text{(as } \boldsymbol{y}_t^k \in \mathcal{X}_t(x_{t-1}^k, \boldsymbol{\xi}_t) \quad \mathbb{P}\text{-a.s.)} && \text{(6.58d)}
\end{aligned}$$

Combining these two results we get Eq. (6.56). □

Proof of Lemma 6.15. Let $t \in [T-1]$. We first prove that if one of the inequalities Eqs. (6.28a) and (6.28b) is satisfied then, x_{t-1}^k is ε_{t-1}-saturated as soon as x_t^k is not δ_t-distinguishable. Recall that $d_t^k(x) := \min_{\kappa < k | x_t^\kappa \text{ is } \varepsilon_t\text{-saturated}} \|x - x_t^\kappa\|$.

Assume now that x_{t+1}^k is not δ_{t+1}-distinguishable, then $d_t^k(x_t^k) \leqslant \delta_t$ and there exists $j < k$

such that x_t^j is ε_t-satured and $\|x_t^j - x_t^k\| \leqslant \delta_t$. If Eq. (6.28a) is satisfied, we have

$$\begin{aligned}
\mathbb{E}\big[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] &\leqslant \overline{V}_t^{k-1}(x_t^k) - \underline{V}_t^{k-1}(x_t^k) + (\overline{L}_t + \underline{L}_t)\eta_t && \text{Eq. (6.28a)} \\
&\qquad (6.59a)\\
&\leqslant \overline{V}_t^{k-1}(x_t^k) - \overline{V}_t^{k-1}(x_t^j) + \overline{V}_t^{k-1}(x_t^j) - \underline{V}_t^{k-1}(x_t^j) + \underline{V}_t^{k-1}(x_t^j) - \underline{V}_t^{k-1}(x_t^k) + (\overline{L}_t + \underline{L}_t)\eta_t \\
&\qquad (6.59b)\\
&\leqslant \overline{L}_t\|x_t^k - x_t^j\| + \overline{V}_t^{k-1}(x_t^j) - \underline{V}_t^{k-1}(x_t^j) + \underline{L}_t\|x_t^j - x_t^k\| + (\overline{L}_t + \underline{L}_t)\eta_t && \text{(Lipschitz)} \\
&\qquad (6.59c)\\
&\leqslant (\overline{L}_t + \underline{L}_t)\delta_t + \overline{V}_t^j(x_t^j) - \underline{V}_t^j(x_t^j) + (\overline{L}_t + \underline{L}_t)\eta_t && \text{(monotonicity)} \\
&\qquad (6.59d)\\
&\leqslant (\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \varepsilon_t && (\varepsilon_t\text{-saturation}) \\
&\qquad (6.59e)
\end{aligned}$$

Similarly, if Eq. (6.28b) is satisfied, we define $j(\xi)$ such that

$$j(\xi) \in \operatorname*{arg\,min}_{j \leqslant k-1, x_t^j \varepsilon_t\text{-saturated}} \|x_t^j - \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi)\| \tag{6.60}$$

In particular, $d_t^k(\mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi)) = \|x_t^{j(\xi)} - \mathcal{F}_{t-1}(\underline{V}_t^{k-1})(x_{t-1}^k, \xi)\|$ and thus $\mathbb{E}\big[d_t^k(\boldsymbol{y}_t^k)\big] = \mathbb{E}\big[\|x_t^{j(\boldsymbol{\xi})} - \boldsymbol{y}_t^k\|\big]$

$$\begin{aligned}
&\mathbb{E}\big[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] \\
&\quad\leqslant \mathbb{E}\big[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \overline{V}_t^{k-1}(x_t^{j(\boldsymbol{\xi}_t)}) + \overline{V}_t^{k-1}(x_t^{j(\boldsymbol{\xi}_t)}) - \underline{V}_t^{k-1}(x_t^{j(\boldsymbol{\xi}_t)}) + \underline{V}_t^{k-1}(x_t^{j(\boldsymbol{\xi}_t)}) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] \\
&\quad\leqslant \mathbb{E}\big[\overline{L}_t\|\boldsymbol{y}_t^k - x_t^{j(\boldsymbol{\xi}_t)}\| + \varepsilon_t + \underline{L}_t\|x_t^{j(\boldsymbol{\xi}_t)} - \boldsymbol{y}_t^k\|\big] \\
&\quad= (\overline{L}_t + \underline{L}_t)\mathbb{E}\big[d_t^k(\boldsymbol{y}_t^k)\big] + \varepsilon_t \\
&\quad\leqslant (\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \varepsilon_t
\end{aligned}$$

Then, in both cases, $\mathbb{E}\big[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] \leqslant (\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \varepsilon_t$. By Lemma 6.28, we have

$$\begin{aligned}
\overline{V}_{t-1}^k(x_{t-1}^k) - \underline{V}_{t-1}^k(x_{t-1}^k) &\leqslant \mathbb{E}\big[\overline{V}_t^{k-1}(\boldsymbol{y}_t^k) - \underline{V}_t^{k-1}(\boldsymbol{y}_t^k)\big] + \underline{\gamma}_{t-1} + \overline{\gamma}_{t-1} + \gamma_t^F && (6.61a)\\
&\leqslant (\overline{L}_t + \underline{L}_t)(\delta_t + \eta_t) + \varepsilon_t + \underline{\gamma}_{t-1} + \overline{\gamma}_{t-1} + \gamma_t^F = \varepsilon_{t-1} && (6.61b)
\end{aligned}$$

Thus, x_t^k is ε_t-saturated as soon as x_{t+1}^k is not δ_{t+1}-distinguishable.

We now prove by backward induction on t that iteration k is effective. We first prove that, for all $k \in \mathbb{N}^\star$, x_{T-1}^k is ε_{T-1}-saturated.

$$\begin{aligned}
\overline{V}_{T-1}(x_{T-1}^k) - \underline{V}_{T-1}(x_{T-1}^k) &\leqslant \mathcal{B}_{T-1}(\overline{V}_T^k)(x_{T-1}^k) + \overline{\gamma}_{T-1} - \mathcal{B}_{T-1}(\underline{V}_T^k)(x_{T-1}^k) + \underline{\gamma}_{T-1} && (6.62a)\\
&= \mathcal{B}_{T-1}(0)(x_{T-1}^k) + \overline{\gamma}_{T-1} - \mathcal{B}_{T-1}(0)(x_{T-1}^k) + \underline{\gamma}_{T-1} && (6.62b)\\
&= \overline{\gamma}_{T-1} + \underline{\gamma}_{T-1} = \varepsilon_{T-1} && (6.62c)
\end{aligned}$$

Let $t \geqslant 2$ such that, for every $\tau \geqslant t$, x_τ^k is ε_τ-saturated . If x_t^k is δ_t-distinguishable, then iteration k is effective. Otherwise, x_t^k is not δ_t-distinguishable and by the previous paragraph,

it implies that x_{t-1}^k is ε_{t-1}-saturated. Eventually, assume that x_1^k is ε_1-saturated. If x_1^k is δ_1-distinguishable, then iteration k is effective. Otherwise, there exists $j < k$ such that $\|x_1^j - x_1^k\| \leqslant \delta_1$ and x_1^j is ε_1 saturated. We get

$$\begin{aligned}
V_1(x_1^k) &\leqslant \overline{V}_1^j(x_1^k) && \text{(6.63a)}\\
&= \overline{V}_1^j(x_1^k) - \overline{V}_1^j(x_1^j) + \overline{V}_1^j(x_1^j) && \text{(6.63b)}\\
&\leqslant \overline{L}_1\|x_1^k - x_1^j\| + \overline{V}_1^j(x_1^j) && \text{(6.63c)}\\
&\leqslant \overline{L}_1\delta_1 + \varepsilon_1 + \underline{V}_1^j(x_1^j) && \text{(6.63d)}\\
&= \overline{L}_1\delta_1 + \varepsilon_1 + \underline{V}_1^j(x_1^j) - \underline{V}_1^j(x_1^k) + \underline{V}_1^j(x_1^k) && \text{(6.63e)}\\
&\leqslant \overline{L}_1\delta_1 + \varepsilon_1 + \underline{L}_1\|x_1^k - x_1^j\| + \underline{V}_1^{k-1}(x_1^k) && \text{(6.63f)}\\
&\leqslant (\overline{L}_1 + \underline{L}_1)\delta_1 + \varepsilon_1 + \underline{V}_1^{k-1}(x_1^k) && \text{(6.63g)}
\end{aligned}$$

Then,

$$\begin{aligned}
\ell_1(x_1^k, \xi_1) + V_1(x_1^k) &\leqslant (\overline{L}_1 + \underline{L}_1)\delta_1 + \varepsilon_1 + \ell_1(x_1^k, \xi_1) + \underline{V}_1^{k-1}(x_1^k) && \text{(6.64a)}\\
&\leqslant (\overline{L}_1 + \underline{L}_1)\delta_1 + \varepsilon_1 + \gamma_1^F + \min_{x_1 \in X_1(x_0)} \ell_1(x_1, \xi_1) + \underline{V}_1^{k-1}(x_1) && \text{(6.64b)}\\
&\leqslant \varepsilon_0 + \min_{x_1 \in X_1(x_0)} \ell_1(x_1, \xi_1) + V_1(x_1) && \text{(6.64c)}
\end{aligned}$$

Thus, in all the covered cases, iteration k is effective. □

6.C Probabilistic lemmas

In this appendix, we present useful probabilistic lemmas to prove the convergence of SDDP with randomized choice of $\boldsymbol{\xi}_t^k$.

6.C.1 A nested Hoeffding lemma

Lemma 6.29. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two independent random variables taking values respectively in the euclidean spaces $\mathbb{X}$ and $\mathbb{Y}$.*

Let $r > 0$ be a positive real and $f : \mathbb{X} \times \mathbb{Y} \mapsto \mathbb{R}$ be a measurable function such that $0 \leqslant f(\boldsymbol{X}, \boldsymbol{Y}) \leqslant r$ almost surely.

Then for every $\eta > 0$ and $A \in \sigma(\boldsymbol{X})$ such that $\mathbb{P}[A] > 0$, we have

$$\mathbb{P}\Big[f(\boldsymbol{X}, \boldsymbol{Y}) > \mathbb{E}\big[f(\boldsymbol{X}, \boldsymbol{Y})|\sigma(X)\big] - \eta \,\Big|\, A\Big] \geqslant 1 - e^{\frac{-2\eta^2}{r^2}} \tag{6.65}$$

Proof. Recall that the Hoeffding lemma states that if $\boldsymbol{Z}$ is a real random variable such that there exists $a, b \in \mathbb{R}$ with $a \leqslant \boldsymbol{Z} \leqslant b$ almost surely then for every $\eta > 0$ we have

$$\mathbb{P}\Big[\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}] \leqslant -\eta\Big] \leqslant e^{\frac{-2\eta^2}{(b-a)^2}} \tag{6.66}$$

By taking the complementary event, we have

$$\mathbb{P}\Big[\boldsymbol{Z} > \mathbb{E}[\boldsymbol{Z}] - \eta\Big] \geqslant 1 - e^{\frac{-2\eta^2}{(b-a)^2}} \tag{6.67}$$

Then for every $x \in \mathbb{X}$, by applying the Hoeffding lemma to the random variable $\boldsymbol{Z} = f(x, \boldsymbol{Y})$, $a = 0$ and $b = r$, we have

$$\mathbb{P}[f(x, \boldsymbol{Y}) > \mathbb{E}[f(x, \boldsymbol{Y})] - \eta] \geqslant 1 - e^{\frac{-2\eta^2}{r^2}} \tag{6.68}$$

Let $A \in \sigma(\boldsymbol{X})$ and $B \subset \mathbb{X}$ such that $A = \boldsymbol{X}^{-1}(B)$

$$\mathbb{P}\Big[\{f(\boldsymbol{X},\boldsymbol{Y}) > \mathbb{E}\big[f(\boldsymbol{X},\boldsymbol{Y})|\sigma(\boldsymbol{X})\big] - \eta\} \cap A\Big] = \int_\Omega \mathbb{1}_{\left\{f(\boldsymbol{X}(\omega),\boldsymbol{Y}(\omega)) > \mathbb{E}[f(\boldsymbol{X},\boldsymbol{Y})|\sigma(\boldsymbol{X})](\omega) - \eta\right\}} \mathbb{1}_{\omega\in A}\mathbb{P}(d\omega) \tag{6.69a}$$

$$= \int_\Omega \mathbb{1}_{\{f(\boldsymbol{X}(\omega),\boldsymbol{Y}(\omega)) > \mathbb{E}_{\boldsymbol{Y}}[f(\boldsymbol{X}(\omega),\boldsymbol{Y})] - \eta\}}(\omega)\mathbb{1}_{X(\omega)\in B}\mathbb{P}(d\omega) \tag{6.69b}$$

$$= \int_{\mathbb{X}}\int_{\mathbb{Y}} \mathbb{1}_{f(x,y) > \mathbb{E}[f(x,\boldsymbol{Y})] - \eta}\mathbb{1}_{x\in B}\mathbb{P}_{\boldsymbol{Y}}(dy)\mathbb{P}_{\boldsymbol{X}}(dx) \tag{6.69c}$$

$$= \int_{\mathbb{X}} \mathbb{1}_{x\in B}\Big(\int_{\mathbb{Y}} \mathbb{1}_{f(x,y) > \mathbb{E}[f(x,\boldsymbol{Y})] - \eta}\mathbb{P}_{\boldsymbol{Y}}(dy)\Big)\mathbb{P}_{\boldsymbol{X}}(dx) \tag{6.69d}$$

$$= \int_{\mathbb{X}} \mathbb{1}_{x\in B}\mathbb{P}\Big[f(x,\boldsymbol{Y}) > \mathbb{E}\big[f(x,\boldsymbol{Y})\big] - \eta\Big]\mathbb{P}_{\boldsymbol{X}}(dx) \tag{6.69e}$$

$$\geqslant \int_{\mathbb{X}} \mathbb{1}_{x\in B}(1 - e^{\frac{-2\eta^2}{r^2}})\mathbb{P}_{\boldsymbol{X}}(dx) \tag{6.69f}$$

$$= (1 - e^{-\frac{-2\eta^2}{r^2}})\mathbb{P}_{\boldsymbol{X}}\big[B\big] \tag{6.69g}$$

$$= (1 - e^{-\frac{-2\eta^2}{r^2}})\mathbb{P}\big[A\big] \tag{6.69h}$$

$$\tag{6.69i}$$

Thus, by dividing by $\mathbb{P}\big[A\big]$, we get $\mathbb{P}\Big[f(\boldsymbol{X},\boldsymbol{Y}) > \mathbb{E}\big[f(\boldsymbol{X},\boldsymbol{Y})|\sigma(X)\big] - \eta \,\Big|\, A\Big] \geqslant 1 - e^{\frac{-2\eta^2}{r^2}}$. □

Lemma 6.30. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables taking values in the euclidean space $\mathbb{X}$ and $\mathcal{A}_n = \sigma(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ be its adapted filtration.*

For every $n \in \mathbb{N}$, let r_n and η_n be two positive real and $f_n : \mathbb{X}^n \mapsto \mathbb{R}$ be a measurable function such that $0 \leqslant f_n(X_1, \dots, X_n) \leqslant r_n$ almost-surely.

We denote by E_n the event $\Big\{\omega \,\Big|\, f_n(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) > \mathbb{E}\big[f_n(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)|\mathcal{A}_{n-1}\big] - \eta_n\Big\} \in \mathcal{A}_n$. Then, for all $m \leqslant n \in \mathbb{N}$ and $A_{m-1} \in \mathcal{A}_{m-1}$ such that $\mathbb{P}\big[A_{m-1}\big] > 0$, we have

$$\mathbb{P}\left[\bigcap_{k=m}^{n} E_k \,\middle|\, A_{m-1}\right] \geqslant \prod_{k=m}^{n}\left(1 - e^{\frac{-2\eta_k^2}{r_k^2}}\right) \tag{6.70}$$

Proof. For every $n \in \mathbb{N}^\star$, $\eta > 0$ and $A_{n-1} \in \mathcal{A}_{n-1}$ such that $\mathbb{P}\big[A\big] > 0$, by the previous lemma applied to $X = (X_1, \dots, X_{n-1})$, $Y = X_n$, $f = f_n$, $\eta = \eta_n$ and $r = r_n$, we have

$$\mathbb{P}\Big[E_n \,\Big|\, A_{n-1}\Big] \geqslant 1 - e^{\frac{-2\eta_n^2}{r_n^2}} \tag{6.71}$$

Let $m \in \mathbb{N}$, we now prove our lemma by induction on n. If $n = m = 1$ the result is true by the Hoeffding lemma and for $n = m > 1$ the result is true by Eq. (6.71) with $A_{n-1} = A_{m-1}$.

Let $n \geqslant m$ and assume that $\mathbb{P}\Big[\bigcap_{k=m}^{n} E_k \cap A_{m-1}\Big] > 0$ and $\mathbb{P}\Big[\bigcap_{k=m}^{n} E_k \,\Big|\, A_{m-1}\Big] \geqslant \prod_{k=m}^{n}\left(1 - e^{\frac{-2\eta_k^2}{r_k^2}}\right)$.

$$\mathbb{P}\Big[\bigcap_{k=m}^{n+1} E_k \,\Big|\, A_{m-1}\Big] = \mathbb{P}\Big[E_{n+1} \,\Big|\, \bigcap_{k=m}^{n} E_k \cap A_{m-1}\Big]\mathbb{P}\Big[\bigcap_{k=m}^{n} E_k \,\Big|\, A_{m-1}\Big] \tag{6.72a}$$

$$\geqslant \left(1 - e^{\frac{-2\eta_{n+1}^2}{r_{n+1}^2}}\right)\prod_{k=m}^{n}\left(1 - e^{\frac{-2\eta_k^2}{r_k^2}}\right) \tag{6.72b}$$

where we underestimate the first factor thanks to Eq. (6.71) and $\bigcap_{k=m}^{n} E_k \in \mathcal{A}_n$ and the second factor thanks to the induction hypothesis. In particular, $\mathbb{P}\Big[\bigcap_{k=m}^{n+1} E_k \cap A_{m-1}\Big] > 0$ and induction ends the proof.

□

6.C.2 Stochastic dominance by geometric random variables

Recall that a real random variable $\boldsymbol{X}$ is (first-order) stochastically dominated by a real random variable $\boldsymbol{Y}$ if the cumulative density function of $\boldsymbol{X}$ is smaller than the cumulative density function of $\boldsymbol{Y}$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are integer random variables, $\boldsymbol{X}$ is stochastically dominated by $\boldsymbol{Y}$ is equivalent to $\mathbb{P}[\boldsymbol{X} \geqslant n] \leqslant \mathbb{P}[\boldsymbol{Y} \geqslant n]$, for all $n \in \mathbb{N}^\star$. We now present a lemma where we leverage this notion to bound the number of effective iteration in randomized algorithm in Theorem 6.21.

Lemma 6.31. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a space of probability, $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables and $\mathcal{A}_n = \sigma(\boldsymbol{X}_k)_{k\in\mathbb{N}^\star}$ be its adapted filtration. Let $(\boldsymbol{Y}_n)_{n\in\mathbb{N}}$ be a sequence of (non necessarily independent neither identically distributed) binary random variables, i.e. taking values in $\{0,1\}$, such that $\sigma(\boldsymbol{Y}_n) \subset \mathcal{A}_n$. Assume that there exists $p \in (0,1)$ such that for all $n \in \mathbb{N}^\star$ and all $A_n \in \mathcal{A}_n$ such that $\mathbb{P}[A_n] > 0$, we have*

$$\mathbb{P}[\boldsymbol{Y}_{n+1} = 1 \,|\, A_n] \geqslant p \tag{6.73}$$

For $m \in \mathbb{N}$, we define the stopping time $\boldsymbol{\tau}_m := \inf\{n \in \mathbb{N} \,|\, \sum_{k=1}^{n} \boldsymbol{Y}_i \geqslant m\}$. Let $B_{m,p}$ be a random variable with a negative binomial distribution representing the number of trials to obtain m successes with probability of success p, i.e. $\mathbb{P}[\boldsymbol{B}_{m,p} = n] = \binom{n-1}{m-1} p^m (1-p)^{n-m}$, for all $n \geqslant m$.

Then, $\boldsymbol{\tau}_m$ is stochastically dominated by $\boldsymbol{B}_{m,p}$ i.e.

$$\mathbb{P}[\boldsymbol{\tau}_m \geqslant n] \leqslant \mathbb{P}[\boldsymbol{B}_{m,p} \geqslant n], \qquad \forall n \in \mathbb{N}^\star \tag{6.74}$$

In particular,

$$\mathbb{E}[\boldsymbol{\tau}_m] \leqslant \mathbb{E}[\boldsymbol{B}_{m,p}] = \frac{m}{p} \tag{6.75}$$

Proof. Let $(\widetilde{\boldsymbol{Y}}_n)_{n\in\mathbb{N}^\star}$ a sequence of independent and identically distributed Bernoulli random variables with parameter p. For all $n \in \mathbb{N}^\star$, we define the random variables $\boldsymbol{S}_n := \sum_{k=1}^{n} \boldsymbol{Y}_n$ and $\widetilde{\boldsymbol{S}}_n := \sum_{k=1}^{n} \widetilde{\boldsymbol{Y}}_n$. We first show by induction on n that $\widetilde{\boldsymbol{S}}_n$ is stochastically dominated by $\boldsymbol{S}_n$, i.e. for all $a \in \mathbb{N}^\star$, we have $\mathbb{P}[\boldsymbol{S}_n \geqslant a] \geqslant \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a]$. Indeed, for $n = 1$ we have $\boldsymbol{S}_1 = \boldsymbol{Y}_1$ and $\boldsymbol{S}_2 = \boldsymbol{Y}_2$ then $\mathbb{P}[\boldsymbol{S}_n \geqslant 0] = \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant 0] = 1$ and $\mathbb{P}[\boldsymbol{S}_n \geqslant 1] = \mathbb{P}[\boldsymbol{Y}_n = 1] \geqslant p = \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant 1]$. Finally, for all $a \geqslant 2$, $\mathbb{P}[\boldsymbol{S}_n \geqslant 2] = \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant 2] = 0$.

We now assume that there exists $n \in \mathbb{N}^\star$ such that for all $a \in \mathbb{N}^\star$, $\mathbb{P}[\boldsymbol{S}_n \geqslant a] \geqslant \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a]$. We then have

$$\mathbb{P}[\boldsymbol{S}_{n+1} \geqslant a] = \mathbb{P}[\boldsymbol{S}_{n+1} \geqslant a, \boldsymbol{S}_n \leqslant a-2] + \mathbb{P}[\boldsymbol{S}_{n+1} \geqslant a, \boldsymbol{S}_n = a-1] + \mathbb{P}[\boldsymbol{S}_{n+1} \geqslant a, \boldsymbol{S}_n \geqslant a] \tag{6.76a}$$

$$= 0 + \mathbb{P}[\boldsymbol{S}_{n+1} \geqslant a \,|\, \boldsymbol{S}_n = a-1]\mathbb{P}[\boldsymbol{S}_n = a-1] + \mathbb{P}[\boldsymbol{S}_n \geqslant a] \tag{6.76b}$$

$$= \mathbb{P}[Y_{n+1} = 1 \,|\, \boldsymbol{S}_n = a-1]\mathbb{P}[\boldsymbol{S}_n = a-1] + \mathbb{P}[\boldsymbol{S}_n \geqslant a] \tag{6.76c}$$

$$\geqslant p\mathbb{P}[\boldsymbol{S}_n = a-1] + \mathbb{P}[\boldsymbol{S}_n \geqslant a] \qquad \text{by assumption (6.73)} \tag{6.76d}$$

$$= p(\mathbb{P}[\boldsymbol{S}_n \geqslant a-1] - \mathbb{P}[\boldsymbol{S}_n \geqslant a]) + \mathbb{P}[\boldsymbol{S}_n \geqslant a] \tag{6.76e}$$

$$= p\mathbb{P}[\boldsymbol{S}_n \geqslant a-1] + (1-p)\mathbb{P}[\boldsymbol{S}_n \geqslant a] \tag{6.76f}$$

$$\geqslant p\mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a-1] + (1-p)\mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a] \qquad \text{by induction assumption} \tag{6.76g}$$

$$= \mathbb{P}[\widetilde{\boldsymbol{Y}}_{n+1} = 1 \,|\, \widetilde{\boldsymbol{S}}_n \geqslant a-1]\mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a-1] + \mathbb{P}[\widetilde{\boldsymbol{Y}}_{n+1} = 0 \,|\, \widetilde{\boldsymbol{S}}_n \geqslant a]\mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a] \tag{6.76h}$$

$$= \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a-1, \widetilde{\boldsymbol{Y}}_{n+1} = 1] + \mathbb{P}[\widetilde{\boldsymbol{S}}_n \geqslant a, \widetilde{\boldsymbol{Y}}_{n+1} = 0] \tag{6.76i}$$

$$= \mathbb{P}[\widetilde{\boldsymbol{S}}_{n+1} \geqslant a, \widetilde{\boldsymbol{Y}}_{n+1} = 1] + \mathbb{P}[\widetilde{\boldsymbol{S}}_{n+1} \geqslant a, \widetilde{\boldsymbol{Y}}_{n+1} = 0] \tag{6.76j}$$

$$= \mathbb{P}[\widetilde{\boldsymbol{S}}_{n+1} \geqslant a] \tag{6.76k}$$

Then, by induction, $\widetilde{\boldsymbol{S}}_n$ is stochastically dominated by $\boldsymbol{S}_n$. For $m \in \mathbb{N}^\star$, we recall that we had $\boldsymbol{\tau}_m = \inf\{n \in \mathbb{N} \,|\, \boldsymbol{S}_n \geqslant m\}$, similarly we define $\widetilde{\boldsymbol{\tau}}_m := \inf\{n \in \mathbb{N} \,|\, \widetilde{\boldsymbol{S}}_n \geqslant m\}$. As $\widetilde{\boldsymbol{S}}_n$ is stochastically dominated by $\boldsymbol{S}_n$, it is easy to see that the stopping time $\boldsymbol{\tau}_m$ is stochastically dominated by the stopping time $\widetilde{\boldsymbol{\tau}}_m$. Indeed, $\mathbb{P}[\boldsymbol{\tau}_m \geqslant a] = \mathbb{P}[\boldsymbol{S}_a < m] = 1 - \mathbb{P}[\boldsymbol{S}_a \geqslant m+1] \leqslant 1 - \mathbb{P}[\widetilde{\boldsymbol{S}}_a \geqslant m+1] = \mathbb{P}[\widetilde{\boldsymbol{S}}_a < m] = \mathbb{P}[\widetilde{\boldsymbol{\tau}}_m \geqslant a]$. Finally, the random variable $\widetilde{\boldsymbol{\tau}}_1$ and the random variables $\widetilde{\boldsymbol{\tau}}_{k+1} - \widetilde{\boldsymbol{\tau}}_k$, for all $k \in \mathbb{N}^\star$, are independent and identically distributed geometric random variables with probability of success p. Thus, $\boldsymbol{B}_{m,p} := \widetilde{\boldsymbol{\tau}}_m = \widetilde{\boldsymbol{\tau}}_1 + \sum_{k=1}^{m-1}(\widetilde{\boldsymbol{\tau}}_{k+1} - \widetilde{\boldsymbol{\tau}}_k)$ is a random variable with negative binomial distribution representing the number of trials to obtain m successes with probability of success p and $\boldsymbol{\tau}_m$ is stochastically dominated by $\boldsymbol{B}_{m,p}$. □

7 Conclusion and perspectives

In this thesis, we showed how polyhedral geometry enlightens the structure of multistage stochastic problems. We understood how higher order polyhedral notions such as chamber complexes and fiber polyhedra characterize the expected cost-to-go functions. By doing so, we constructed explicit local or uniform, universal exact quantization results for MSLP with fixed recourse. These results allowed us to derive both complexity results and new algorithms such as GAPM and TFDP to deal with MSLP with general distributions. However, the size of the exact quantization can be exponential in the state dimension and thus difficult to use directly in practice.

We now present some perspectives for subsequent works leveraging this manuscript's results.

Higher order simplex method for 2SLP Simplex methods solve linear problems by following a path on the vertices of the polyhedron. Combinatorially, the simplex updates a basis by making pivots while reducing the cost. The tools of Chapter 3 are a first step to define a higher order simplex algorithm to solve 2SLP with general cost. In particular, a collection of active constraints sets not only represents the current visited cell of the chamber complex, see Proposition 3.33, but also the normal fan of the fibers, see Fig. 3.13, and the affine coefficients of the expected cost-to-go function, see Theorem 4.9. Geometrically, we could follow a path on the chamber complex while reducing the cost. Combinatorially, it would be equivalent to implement and update the current collection of active constraints sets while reducing the cost through local changes. Moreover, we could leverage the theory of regular subdivisions to understand the behavior of this collection as, by Proposition 3.19, the collection of active constraints sets is a regular subdivision. We leave the full development of this algorithm for further work.

2-time scale MSLP and nested fiber polyhedra A common application in stochastic optimization is the management of an electrical microgrid with storage capacity to maximize the return on the sales and purchases on the intraday and day-ahead markets, see [LF21, Chapter 3]. In this setting, the decision maker has to take a decision, before the first time step of the multistage problem, which impacts the dynamics of the multistage problem at each time, we speak of two-time-scale MSLP:

$$\min_{p\in\mathbb{R}^m,(\boldsymbol{x}_t)\in\mathbb{R}^{n_t}} \quad q^\top p + \mathbb{E}\big[\sum_{t=1}^{T} \boldsymbol{c}_t^\top \boldsymbol{x}_t\big] \tag{7.1a}$$

$$\text{s.t.} \quad Dp \leqslant d \tag{7.1b}$$

$$A_t\boldsymbol{x}_t + B_t\boldsymbol{x}_{t-1} + C_t p \leqslant h_t \quad \text{a.s.} \qquad \forall t \in [T] \tag{7.1c}$$

$$\boldsymbol{x}_t \in L_\infty(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_t}) \qquad \forall t \in [T] \tag{7.1d}$$

$$\sigma(\boldsymbol{x}_t) \subset \sigma(\boldsymbol{c}_1, \cdots, \boldsymbol{c}_t) \qquad \forall t \in [T] \tag{7.1e}$$

In Theorem 4.17, the nested fiber polyhedra only depends on A_t, B_t and $\boldsymbol{c}_t$ but not on the right handside h_t. Thus, by replacing h_t by $h_t - C_t p$ in Theorem 4.17, we have that Problem (7.1) is equivalent to the deterministic problem

$$\min_{p\in\mathbb{R}^m} q^\top p + \max_{(\lambda_t)_{t\in[T]}} \sum_{t=1}^{T} (C_t p - h_t)^\top \lambda_t \tag{7.2a}$$

$$\text{s.t. } Dp \leqslant d \tag{7.2b}$$

$$\lambda = (\lambda_1, \cdots, \lambda_T) \in E_1 \tag{7.2c}$$

By introducing, a lift variable Problem (7.1) is equivalent to the deterministic linear problem:

$$\min_{p\in\mathbb{R}^m} q^\top p + z \tag{7.3a}$$

$$\text{s.t. } Dp \leqslant d \tag{7.3b}$$

$$\sum_{t=1}^{T} (C_t p - h_t)^\top v_t \leqslant z \qquad \forall v = (v_1, \cdots, v_T) \in \mathrm{Vert}(E_1) \tag{7.3c}$$

This new formulation of this two-time-scale MSLP could open the door to new resolution methods, such as Benders decomposition to deal with the great number of constraints induced by the vertices of the weighted nested fiber polyhedron.

Reintroduce Approximation or Sampling In Section 5.4, we saw that the most time-consuming parts of GAPM algorithm are the exact computations of volumes. Indeed, it was proven to be $\sharp P$-complete [DF88]. Thus, to improve Algorithm 5.1, we could use precise rapid approximation volume algorithms, see e.g. [CV16]. More generally, we could develop approximate quantization for example by approximating the expectation as in Section 4.5.2 or use classical approximation ideas such as sampling, barycentric scenario trees or nested distance, see Section 2.1.3. In particular, the exact quantization results could provide new insights both to understand theoretically the convergence of such approximations or to improve them in practice. For example, we conjecture that the Hausdorff distance between nested weighted fiber polyhedra is connected to the nested distance of [PP12].

Exact quantization for stochastic integer linear problems The exact quantization result can be understood thanks to the basis decomposition theorem of Walkup and Wets [WW69], which shows, in a modern vocabulary, that the secondary fan is a key notion to understand the behavior of the value of a parametric linear problem. In [ST97], Sturmfels and Thomas extended this basis decomposition theorem to parametric integer linear problems showing that the key notions to understand how the value of an integer linear problem varies with its parameter are the Gröbner and Graver fans. We thus conjecture that there exist exact quantization results for stochastic integer linear problems by using these Gröbner and Graver fans.

Understanding the complexity of MSLP We saw that 2SLP with general cost distribution was in some way similar to the problem of computing the volume of polytope from a complexity point of view. Indeed, in [HKW16], Hanasusanto, Kuhn and Wiesemann proved that 2SLP was $\sharp$P-hard and we proved in Section 6.3 that 2SLP was polynomial when we fix the dimension of the recourse, like the computation of the volume, see [Bar08]. However, it was proven in [DF88] that computing the volume of a polytope is $\sharp$P-complete. This raises the question of showing that for a large enough class of cost distributions (including uniform measures on rational polytopes), 2SLP is not only $\sharp$P-hard but $\sharp$P-complete.

Moreover, we studied the complexity of TFDP algorithms in Chapter 6. If deterministic node selections, such as problem child and explorative, lead to a number of iterations polynomial in the horizon, the best bound yet proven for random node selection, is exponential in the horizon. We thus need further works to understand if TFDP algorithms with random node selection, such as SDDP, are polynomial in the horizon. Indeed, SDDP algorithm is not really affected by the horizon in practice. However, the proof of complexity of TFDP algorithms exploits very few the structure, aside from Lipschitz continuity and compactness, this raises the question of showing if more assumptions, such as convexity or linearity, improve the number of iterations.

A Proof of a counter example

We prove that Example 2.2 does not admit neither a local nor a uniform exact quantization.

Let $n = m = \ell = 1$, $\boldsymbol{q} = (1)$, $\boldsymbol{T} = (0)$, $\boldsymbol{W} = (-\boldsymbol{u})$, $\boldsymbol{h} = (-1)$ where $\boldsymbol{u}$ denotes a uniform random variable on $[1, 2]$. We then have for all $x \in \mathbb{R}^n$:

$$\hat{V}(x, \xi) = \begin{array}{ll} \min\limits_{y \in \mathbb{R}^m} & y \\ \text{s.t.} & uy \geqslant 1 \end{array} = \frac{1}{u} \tag{A.1}$$

In particular, by the law of total expectation and Jensen's inequality for all partition $\mathcal{P}$ of Ξ:

$$V(x) = \mathbb{E}[\hat{V}(x, \boldsymbol{\xi})|P] = \sum_{P \in \mathcal{P}} \mathbb{P}[P]\mathbb{E}[\hat{V}(x, \boldsymbol{\xi})|P] \geqslant \sum_{P \in \mathcal{P}} \mathbb{P}[P]\hat{V}(x, \mathbb{E}[\boldsymbol{\xi}|P]) \tag{A.2}$$

However, for all $1 \leqslant a < b \leqslant 2$ we have

$$\mathbb{E}\Big[\hat{V}(x, \boldsymbol{\xi}) \,|\, a \leqslant \boldsymbol{u} \leqslant b\Big] = \mathbb{E}\Big[\frac{1}{\boldsymbol{u}} \,|\, a \leqslant \boldsymbol{u} \leqslant b\Big] = \frac{1}{b-a} \ln\left(\frac{b}{a}\right) \tag{A.3a}$$

$$\hat{V}\Big(x, \mathbb{E}[\boldsymbol{\xi} \,|\, a \leqslant \boldsymbol{u} \leqslant b]\Big) = \frac{1}{\mathbb{E}[\boldsymbol{\xi} \,|\, a \leqslant \boldsymbol{u} \leqslant b]} = \frac{2}{b+a} \tag{A.3b}$$

Since, for all $t \in (1, +\infty)$, $2\frac{t-1}{t+1} \leqslant \ln(t)$, we have for all $1 \leqslant a < b \leqslant 2$, $\frac{2}{b+a} < \frac{1}{b-a} \ln\left(\frac{b}{a}\right)$ and then $\mathbb{E}\Big[\hat{V}(x, \boldsymbol{\xi}) \,|\, a \leqslant \boldsymbol{u} \leqslant b\Big] > \hat{V}\Big(x, \mathbb{E}[\boldsymbol{\xi} \,|\, a \leqslant \boldsymbol{u} \leqslant b]\Big)$.

Finally, for all partition $\mathcal{P}$ since the partition is finite there exists $1 \leqslant a < b \leqslant 2$ and $P \in \mathcal{P}$ such that $\mathbb{P}[a \leqslant \boldsymbol{u} \leqslant b \cap P] = \mathbb{P}[a \leqslant \boldsymbol{u} \leqslant b]$. We then have

$$\mathbb{P}[P]\mathbb{E}[\hat{V}(x, \boldsymbol{\xi})|P] = \mathbb{P}[a \leqslant \boldsymbol{u} \leqslant b]\mathbb{E}\Big[\hat{V}(x, \boldsymbol{\xi}) \,|\, a \leqslant \boldsymbol{u} \leqslant b\Big] \tag{A.4a}$$

$$+ \mathbb{P}[\{a < \boldsymbol{u} \text{ or } \boldsymbol{u} > b\} \cap P]\mathbb{E}\Big[\hat{V}(x, \boldsymbol{\xi}) \,|\, \{a < \boldsymbol{u} \text{ or } \boldsymbol{u} > b\} \cap P\Big] \tag{A.4b}$$

$$< \mathbb{P}[a \leqslant \boldsymbol{u} \leqslant b]\hat{V}\Big(x, \mathbb{E}[\boldsymbol{\xi} \,|\, a \leqslant \boldsymbol{u} \leqslant b]\Big) \tag{A.4c}$$

$$+ \mathbb{P}[\{a < \boldsymbol{u} \text{ or } \boldsymbol{u} > b\} \cap P]\hat{V}\Big(x, \mathbb{E}[\boldsymbol{\xi} \,|\, \{a < \boldsymbol{u} \text{ or } \boldsymbol{u} > b\} \cap P]\Big) \tag{A.4d}$$

$$= \mathbb{P}[P]\hat{V}\Big[x, \mathbb{E}[\boldsymbol{\xi}|P]\Big] \tag{A.4e}$$

Thus, for all $x \in \mathbb{R}^n$, there is no partition-based local, thus uniform, exact quantization result at x when $\boldsymbol{W}$ is non-finitely supported.

B Computing the quantized costs and probabilities

In this appendix, we show that, for three standard classes of distributions (uniform on a polytope, exponential, and Gaussian), the quantized costs $\check{c}_R$ and probabilities $\check{p}_R$ arising in the exact quantization (Theorems 4.2 and 4.3), can be effectively computed.

The formulas are summed up in Table B.1. They are detailed and established in Sections B.1–B.3. We provide these formulas for *simplices* or *simplicial cones* S with $\dim(S) = \dim(\operatorname{supp} \boldsymbol{c})$. This extends to any polyhedron R, through triangulation of $R \cap \operatorname{supp}(\boldsymbol{c})$ into simplices and simplicial cones $(S_k)_{k\in[l]}$. We then compute $\check{p}_R = \sum_{k=1}^{l} \check{p}_{S_k}$ and $\check{c}_R = \sum_{k=1}^{l} \check{p}_{S_k}\check{c}_{S_k}/\check{p}_R$ if $\check{p}_R \neq 0$ and $\check{c}_R = 0$ otherwise.

	Uniform	Exponential	Gaussian
$d\mathbb{P}(c)$	$\frac{\mathbb{1}_{c\in Q}}{\operatorname{Vol}_d(Q)} d\mathcal{L}_{\operatorname{Aff}(Q)}(c)$	$\frac{e^{\theta^\top c}\mathbb{1}_{c\in K}}{\Phi_K(\theta)} d\mathcal{L}_{\operatorname{Aff}(K)}(c)$	$\frac{e^{-\frac{1}{2}c^\top M^{-2}c}}{(2\pi)^{\frac{m}{2}}\det M} dc$
$\operatorname{supp} \boldsymbol{c}$	Polytope : Q	Cone : K	$\mathbb{R}^m$
$\check{p}_S$	$\frac{\operatorname{Vol}_d(S)}{\operatorname{Vol}_d(Q)}$	$\frac{\lvert\det(\operatorname{Ray}(S))\rvert}{\Phi_K(\theta)} \prod_{r\in\operatorname{Ray}(S)} \frac{1}{-r^\top\theta}$	$\operatorname{Ang}\left(M^{-1}S\right)$
$\check{c}_S$	$\frac{1}{d}\sum_{v\in\operatorname{Vert}(S)} v$	$\left(\sum_{r\in\operatorname{Ray}(S)} \frac{-r_i}{r^\top\theta}\right)_{i\in[m]}$	$\frac{\sqrt{2}\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} M \operatorname{SpCtr}\left(S \cap \mathbb{S}_{m-1}\right)$

Table B.1: Probabilities $\check{p}_S$ and expectations $\check{c}_S$ arising from different cost distributions over simplicial cones or simplices $S \subset \operatorname{supp}(\boldsymbol{c})$ with $\dim S = \dim(\operatorname{supp} \boldsymbol{c})$, where $\mathcal{L}_A$ is the Lebesgue measure on an affine space A.

B.1 Uniform distributions on polytopes

The *volume* of a polytope $Q \subset \mathbb{R}^m$ is the volume of Q seen as a subset of the smallest affine space $\operatorname{Aff}(Q)$ it lives in. The volume of a full dimensional simplex S in $\mathbb{R}^d$ with vertices $v_1, \ldots, v_{d+1}$ is given by $\operatorname{Vol}(S) = \frac{1}{n!}|\det(v_1 - v_{d+1}, \cdots, v_d - v_{d+1})|$, see for example [GK94] 3.1. The *centroid* of a non-empty polytope $Q \subset \mathbb{R}^m$ is $\operatorname{Ctr}(Q) := \frac{1}{\operatorname{Vol} Q}\int_Q y d\mathcal{L}_{\operatorname{Aff} Q}(y)$. For instance, the centroid of a simplex S of (non necessary full) dimension d is the equibarycenter of its vertices: $\operatorname{Ctr}(S) = \frac{1}{d+1}\sum_{v\in\operatorname{Vert}(S)} v$.

Assume now that Q is a polytope of dimension d, and that $\boldsymbol{c}$ is uniform on Q. Let $S \subset Q$ be a simplex with $\dim(S) = \dim(Q)$, then we have

$$\check{p}_S = \frac{\operatorname{Vol}_d S}{\operatorname{Vol}_d Q} \qquad \text{and} \qquad \check{c}_S = \frac{1}{d+1}\sum_{v\in\operatorname{Vert}(S)} v \ . \tag{B.1}$$

B.2 Exponential distributions on cones

Let $P \subset \mathbb{R}^m$ be a polyhedron and $\theta \in \operatorname{ri}\big((\operatorname{rc} P)^\circ\big)$. We denote by $\Phi_P(\theta) := \int_P e^{\theta^\top c} d\mathcal{L}_{\operatorname{Aff}(P)}(c)$ the *exponential valuation* of P with parameter θ.

Proposition B.1 (Brion's formula [Bri88])**.** *Let S be a full dimensional simplicial cone, and let by a sligth abuse of notation* $\operatorname{Ray}(S)$ *be a square matrix whose columns are obtained by selecting precisely one element in every extreme ray of S, so that $S = \operatorname{Cone}\big(\operatorname{Ray}(S)\big)$. Then for any $\theta \in \operatorname{ri} S^\circ$, the exponential valuation of S is given by*

$$\Phi_S(\theta) = |\det(\operatorname{Ray}(S))| \prod_{r \in \operatorname{Ray}(S)} \frac{1}{-\theta^\top r} \ . \tag{B.2}$$

Let K be a (non necessarily simplicial) full dimensional polyhedral cone of $\mathbb{R}^m$ and $\theta \in \operatorname{ri} K^\circ$ a vector. Assume that $\boldsymbol{c}$ has the following exponential density :

$$d\mathbb{P}(c) := e^{\theta^\top c} \mathbb{1}_{c \in K} \frac{1}{\Phi_K(\theta)} d\mathcal{L}_{\operatorname{Aff}(K)}(c) \tag{B.3}$$

Let $S \subset K$ be a simplicial cone with $\dim S = \dim K$, by Brion's formula (B.2),

$$\check{p}_S = \frac{\Phi_S(\theta)}{\Phi_K(\theta)} = \frac{1}{\Phi_K(\theta)} |\det(\operatorname{Ray}(S))| \prod_{r \in \operatorname{Ray}(S)} \frac{1}{-r^\top \theta} \tag{B.4}$$

Further,

$$\check{p}_S \check{c}_S = \mathbb{E}\big[\mathbb{1}_{\boldsymbol{c} \in S} \boldsymbol{c}^\top\big] = \frac{1}{\Phi_K(\theta)} \int_S c e^{\theta^\top c} dc = \frac{\nabla \Phi_S(\theta)}{\Phi_K(\theta)} \ . \tag{B.5}$$

By computing explicitly the latter gradient, dividing by $\check{p}_S$, and simplifying, we obtain:

$$\check{c}_S = \Big(\sum_{r \in \operatorname{Ray}(S)} \frac{-r_i}{r^\top \theta} \Big)_{i \in [m]} \ . \tag{B.6}$$

where r_i is the i-th coordinate of r.

B.3 Gaussian distributions

The *solid angle* of a pointed cone $K \subset \mathbb{R}^d$ is defined as the normalized volume of its intersection with the unit ball $\mathbb{B}_d$, i.e.: $\operatorname{Ang}(K) := \operatorname{Vol}_d(K \cap \mathbb{B}_d)/\operatorname{Vol}_d(\mathbb{B}_d)$. Recall that $\operatorname{Vol}_d(\mathbb{B}_d) = \pi^{\frac{d}{2}}/\Gamma(\frac{d}{2}+1)$ with Γ the Euler gamma function, and that ([Rib06]) for any function $f : \mathbb{R}^m \to \mathbb{R}$ invariant under rotations around the origin and any pointed cone $K \subset \mathbb{R}^m$, we have $\operatorname{Ang}(K) \int_{\mathbb{R}^m} f = \int_K f$.

Let $\boldsymbol{c}$ be a non-degenerate, centered, Gaussian random variable of variance M^2, where M is a symmetric positive definite matrix. Then, if K is a polyhedral cone, we have

$$\check{p}_K = \int_{M^{-1}K} \frac{e^{-\frac{1}{2}\|c\|_2^2}}{(2\pi)^{\frac{m}{2}}} dc = \operatorname{Ang}(M^{-1}K)$$

We shall use the notion of *spherical centroid* $\operatorname{SpCtr}(U)$ for a measurable subset U included in the unit sphere. It is defined as the barycenter of the elements of U with respect to the uniform measure on the sphere. Note that the spherical centroid does not belong to the sphere,

unless U is trivial. By denoting $\mathbb{S}_{m-1}$ the unit sphere in $\mathbb{R}^m$, we have

$$\begin{aligned}\check{p}_K \check{c}_K &= \int_{M^{-1}K} Mc \frac{e^{-\frac{1}{2}\|c\|_2^2}}{(2\pi)^{\frac{m}{2}}} dc = M \int_{\mathbb{R}^+} r^m \frac{e^{-\frac{r^2}{2}}}{(2\pi)^{\frac{m}{2}}} dr \int_{M^{-1}K \cap \mathbb{S}_{m-1}} \varphi d\varphi \\ &= M \frac{\Gamma(\frac{m+1}{2})}{\sqrt{2}\pi^{\frac{m}{2}}} \operatorname{Vol}_{m-1}(\mathbb{S}_{m-1}) \operatorname{Ang}\left(M^{-1}K\right) \operatorname{SpCtr}\left(M^{-1}K \cap \mathbb{S}_{m-1}\right) \\ &= M \frac{\sqrt{2}\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \operatorname{Ang}\left(M^{-1}K\right) \operatorname{SpCtr}\left(M^{-1}K \cap \mathbb{S}_{m-1}\right)\end{aligned}$$

Similarly, one can get explicit formulæ when $\boldsymbol{c}$ is distributed uniformly on an ellipsoid, or on the surface of an ellipsoid, or more generally, when the distribution of $\boldsymbol{c}$ is invariant under the action of an orthogonal group. Then, the quantized costs and probabilities $\check{c}_S$ and $\check{p}_S$ are still given by solid angles and spherical centroids, in a way similar to Table B.1.

Bibliography

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[ACdC20] Shabbir Ahmed, Filipe Goulart Cabral, and Bernardo Freitas Paulo da Costa. Stochastic lipschitz dynamic programming. *Mathematical Programming*, pages 1–39, 2020.

[ACT20] Marianne Akian, Jean-Philippe Chancelier, and Benoît Tran. Tropical dynamic programming for lipschitz multistage stochastic programming. *arXiv preprint arXiv:2010.10619*, 2020.

[ADE+07] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, David Heath, and Hyejin Ku. Coherent multiperiod risk adjusted values and bellman's principle. *Annals of Operations Research*, 152(1):5–22, 2007.

[ADEH99] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[AP18] Tsvetan Asamov and Warren B Powell. Regularized decomposition of high-dimensional multistage stochastic programs with markov uncertainty. *SIAM Journal on Optimization*, 28(1):575–595, 2018.

[Bar94] Alexander I Barvinok. A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed. *Mathematics of Operations Research*, 19(4):769–779, 1994.

[Bar08] Alexander Barvinok. *Integer points in polyhedra*, volume 452. European Mathematical Society, 2008.

[BB88] J. M. Borwein and P. B. Borwein. On the complexity of familiar functions and numbers. *SIAM Review*, 30(4):589–601, 1988.

[BBC11] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

[BCB+12] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[BDZ17] Regan Baucke, Anthony Downward, and Golbon Zakeri. A deterministic algorithm for solving multistage stochastic programming problems. *Optimization Online*, pages 1–25, 2017.

[BDZ18] Regan Baucke, Anthony Downward, and Golbon Zakeri. A deterministic algorithm for solving multistage stochastic minimax dynamic programmes. *Optimization Online*, 2018.

[Bel66] Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[Ber12] Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 1. Athena scientific, 2012.

[BFFdO20] Felipe Beltrán, Erlon C Finardi, Guilherme M Fredo, and Welington de Oliveira. Improving the performance of the stochastic dual dynamic programming algorithm using chebyshev centers. *Optimization and Engineering*, pages 1–22, 2020.

[BG21] Michelle Bandarra and Vincent Guigues. Single cut and multicut stochastic dual dynamic programming with cut selection for multistage stochastic linear programs: convergence proof and numerical experiments. *Computational Management Science*, 18(2):125–148, 2021.

[Bir85] John R Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations research*, 33(5):989–1007, 1985.

[BL11] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[Bri88] Michel Brion. Points entiers dans les polyedres convexes. In *Annales scientifiques de l'Ecole Normale Superieure*, volume 21, pages 653–663, 1988.

[BS92] Louis J Billera and Bernd Sturmfels. Fiber polytopes. *Annals of Mathematics*, pages 527–549, 1992.

[BTEGN09] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*, volume 28. Princeton university press, 2009.

[BW86] John R Birge and Roger J-B Wets. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. In *Stochastic Programming 84 Part I*, pages 54–102. Springer, 1986.

[CL98] Philippe Clauss and Vincent Loechner. Parametric analysis of polyhedral iteration spaces. *Journal of VLSI signal processing systems for signal, image and video technology*, 19(2):179–194, 1998.

[CS05] Michael S Casey and Suvrajeet Sen. The scenario generation algorithm for multistage stochastic linear programming. *Mathematics of Operations Research*, 30(3):615–631, 2005.

[CV16] Ben Cousins and Santosh Vempala. A practical volume algorithm. *Mathematical Programming Computation*, 8(2):133–160, 2016.

[dCL21] Bernardo Freitas Paulo da Costa and Vincent Leclère. Dual sddp for risk-averse multistage stochastic programs. *arXiv preprint arXiv:2107.10930*, 2021.

[DDB20] Anthony Downward, Oscar Dowson, and Regan Baucke. Stochastic dual dynamic programming with stagewise-dependent objective uncertainty. *Operations Research Letters*, 48(1):33–39, 2020.

[DF88] Martin E. Dyer and Alan M. Frieze. On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, 17(5):967–974, 1988.

[DGKR03] Jitka Dupačová, Nicole Gröwe-Kuska, and Werner Römisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3):493–511, 2003.

[DLRS10] Jesús A De Loera, Jörg Rambau, and Francisco Santos. *Triangulations Structures for algorithms and applications*. Springer, 2010.

[DM20] Daniel Duque and David P Morton. Distributionally robust stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(4):2841–2865, 2020.

[DMP20] Oscar Dowson, David P Morton, and Bernardo K Pagnoncelli. Partially observable multistage stochastic programming. *Operations Research Letters*, 48(4):505–512, 2020.

[Dow18] Oscar Dowson. *Applying stochastic optimisation to the new zealand dairy industry*. PhD thesis, University of Auckland, 2018.

[DR84] Ph J Davis and Ph Rabinowitz. *Methods of numerical integration. 1984*. Comput. Sci. Appl. Math. Acad. Press, 1984.

[DS06] Martin Dyer and Leen Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432, 2006.

[DT03] George Bernard Dantzig and Mukund N Thapa. *Linear programming: Theory and extensions*, volume 2. Springer, 2003.

[Dup90] Jitka Dupačová. Stability and sensitivity-analysis for stochastic programming. *Annals of operations research*, 27(1):115–142, 1990.

[Dup02] Jitka Dupačová. Applications of stochastic programming: Achievements and questions. *European Journal of Operational Research*, 140(2):281–290, 2002.

[EK08] Alexander Esterov and Askold Khovanskii. Elimination theory and newton polytopes. *Functional Analysis and Other Mathematics*, 2(1):45–71, 2008.

[EZ94a] NCP Edirisinghe and William T Ziemba. Bounding the expectation of a saddle function with application to stochastic programming. *Mathematics of Operations Research*, 19(2):314–340, 1994.

[EZ94b] NCP Edirisinghe and William T Ziemba. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research*, 19(2):292–313, 1994.

[FGL21] Maël Forcier, Stéphane Gaubert, and Vincent Leclère. Exact quantization of multistage stochastic linear problems. *arXiv preprint arXiv:2107.09566*, 2021.

[FL22a] Maël Forcier and Vincent Leclère. Convergence of trajectory following dynamic programming algorithms for multistage stochastic problems without finite support assumptions. 2022.

[FL22b] Maël Forcier and Vincent Leclère. Generalized adaptive partition-based method for two-stage stochastic linear programs: geometric oracle and analysis. *Operations Research Letters*, 2022.

[FP95] Komei Fukuda and Alain Prodon. Double description method revisited. In *Franco-Japanese and Franco-Chinese Conference on Combinatorics and Computer Science*, pages 91–111. Springer, 1995.

[FP96] K. Fukuda and A. Prodon. Double description method revisited. In *Selected papers from the 8th Franco-Japanese and 4th Franco-Chinese Conference on Combinatorics and Computer Science*, pages 91–111, London, UK, 1996. Springer-Verlag.

[Fra96] Karl Frauendorfer. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming*, 75(2):277–293, 1996.

[Fuk16] Komei Fukuda. Lecture: Polyhedral computation, spring 2016. 2016.

[GJ00] Ewgenij Gawrilow and Michael Joswig. `polymake`: a framework for analyzing convex polytopes. In *Polytopes—combinatorics and computation (Oberwolfach, 1997)*, volume 29 of *DMV Sem.*, pages 43–73. Birkhäuser, Basel, 2000.

[GK94] Peter Gritzmann and Victor Klee. On the complexity of some basic problems in computational convexity, ii. volume and mixed volumes. In *Polytopes: Abstract, Convex and Computational*, pages 373–466. Springer, 1994.

[GKZ90] IM Gelfand, MM Kapranov, and AV Zelevinsky. Newton polytopes of the classical resultant and discriminant. *Advances in Mathematics*, 84(2):237–254, 1990.

[GKZ94] Israel M Gelfand, Mikhail M Kapranov, and Andrei V Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Springer, 1994.

[GLP15] Pierre Girardeau, Vincent Leclere, and Andrew B Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Mathematics of Operations Research*, 40(1):130–145, 2015.

[GLS12] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[GLT20] Vincent Guigues, Migual A Lejeune, and Wajdi Tekaya. Regularized stochastic dual dynamic programming for convex nonlinear optimization problems. *Optimization and Engineering*, 21(3):1133–1165, 2020.

[Grü13] Branko Grünbaum. *Convex polytopes*, volume 221. Springer Science & Business Media, 2013.

[GTW19] Angelos Georghiou, Angelos Tsoukalas, and Wolfram Wiesemann. Robust dual dynamic programming. *Operations Research*, 67(3):813–830, 2019.

[Gui16] Vincent Guigues. Convergence analysis of sampling-based decomposition methods for risk-averse multistage stochastic convex programs. *SIAM Journal on Optimization*, 26(4):2468–2494, 2016.

[Gui20] Vincent Guigues. Inexact cuts in stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(1):407–438, 2020.

[GZ13] Horand Gassmann and William T Ziemba. *Stochastic programming: applications in finance, energy, planning and logistics*, volume 4. World Scientific, 2013.

[HdMDMF11] Tito Homem-de Mello, Vitor L De Matos, and Erlon C Finardi. Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling. *Energy Systems*, 2(1):1–31, 2011.

[HKW16] Grani A Hanasusanto, Daniel Kuhn, and Wolfram Wiesemann. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming*, 159(1-2):557–569, 2016.

[HR03] Holger Heitsch and Werner Römisch. Scenario reduction algorithms in stochastic programming. *Computational optimization and applications*, 24(2-3):187–206, 2003.

[HR07] Holger Heitsch and Werner Römisch. A note on scenario reduction for two-stage stochastic programs. *Operations Research Letters*, 35(6):731–738, 2007.

[IJCN15] Rasmus Ibsen-Jensen, Krishnendu Chatterjee, and Martin A Nowak. Computational complexity of ecological and evolutionary spatial dynamics. *Proceedings of the National Academy of Sciences*, 112(51):15636–15641, 2015.

[JKM08] Colin N Jones, Eric C Kerrigan, and Jan M Maciejowski. On polyhedral projection and parametric programming. *Journal of Optimization Theory and Applications*, 138(2):207–220, 2008.

[JT13] Michael Joswig and Thorsten Theobald. *Polyhedral and algebraic methods in computational geometry*. Springer Science & Business Media, 2013.

[JW92] Vaithilingam Jeyakumar and Henry Wolkowicz. Generalizations of slater's constraint qualification for infinite convex programs. *Mathematical Programming*, 57(1):85–101, 1992.

[Kha80] Leonid G Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.

[KLM96] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[KM+76] Peter Kall, Janos Mayer, et al. *Stochastic linear programming*, volume 7. Springer, 1976.

[Kuh06] Daniel Kuhn. *Generalized bounds for convex multistage stochastic programs*, volume 548. Springer Science & Business Media, 2006.

[KW07] Michal Kaut and Stein W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271, 2007.

[Lan20] Guanghui Lan. Complexity of stochastic dual dynamic programming. *Mathematical Programming*, pages 1–38, 2020.

[LCC+20] Vincent Leclere, Pierre Carpentier, Jean-Philippe Chancelier, Arnaud Lenoir, and François Pacaud. Exact converging bounds for stochastic dual dynamic programming via fenchel duality. *SIAM Journal on Optimization*, 30(2):1223–1250, 2020.

[LF21] Adrien Le Franc. *Subdifferentiability in convex and stochastic optimization applied to renewable power systems*. PhD thesis, École des Ponts ParisTech, 2021.

[LJ83] Hendrik W Lenstra Jr. Integer programming with a fixed number of variables. *Mathematics of operations research*, 8(4):538–548, 1983.

[LL93] Gilbert Laporte and François V Louveaux. The integer l-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.

[Löh16] Nils Löhndorf. An empirical analysis of scenario generation methods for stochastic optimization. *European Journal of Operational Research*, 255(1):121–132, 2016.

[Lou80] Francois V Louveaux. A solution method for multistage stochastic programs with recourse with application to an energy investment problem. *Operations Research*, 28(4):889–902, 1980.

[LR08] Shu Lu and Stephen M Robinson. Normal fans of polyhedral convex sets. *Set-Valued Analysis*, 16(2-3):281–305, 2008.

[LW97] Vincent Loechner and Doran K Wilde. Parameterized polyhedra and their vertices. *International Journal of Parallel Programming*, 25(6):525–549, 1997.

[MAB14] Francesca Maggioni, Elisabetta Allevi, and Marida Bertocchi. Bounds in multistage linear stochastic programming. *Journal of Optimization Theory and Applications*, 163(1):200–229, 2014.

[Mad59] Albert Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746, 1959.

[McM70] Peter McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17(2):179–184, 1970.

[MG07] Jiri Matousek and Bernd Gärtner. *Understanding and using linear programming*. Springer Science & Business Media, 2007.

[MP18] Francesca Maggioni and Georg Pflug. Guaranteed bounds for general non-discrete multistage risk-averse stochastic optimization programs. *SIAM Journal on Optimization*, 2052649:1–26, 2018.

[MTY19] Cara Monical, Neriman Tokcan, and Alexander Yong. Newton polytopes in algebraic combinatorics. *Selecta Mathematica*, 25(5):1–37, 2019.

[PdMF13] Andy Philpott, Vitor de Matos, and Erlon Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61(4):957–970, 2013.

[PG08] Andrew B Philpott and Ziming Guan. On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36(4):450–455, 2008.

[PP91] Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1-3):359–375, 1991.

[PP12] Georg Ch Pflug and Alois Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012.

[PWB20] Andrew B Philpott, Faisal Wahid, and J Frédéric Bonnans. Midas: A mixed integer dynamic approximation scheme. *Mathematical Programming*, 181(1):19–50, 2020.

[Ram96] Jörg Rambau. *Polyhedral subdivisions and projections of polytopes*. Phd thesis, TU Berlin, 1996.

[Rib06] Jason M Ribando. Measuring solid angles beyond dimension three. *Discrete & Computational Geometry*, 36(3):479–487, 2006.

[RM19] Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.

[Roc15] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[RPM21] Cristian Ramirez-Pico and Eduardo Moreno. Generalized adaptive partition-based method for two-stage stochastic linear programs with fixed recourse. *Mathematical Programming*, pages 1–20, 2021.

[RS91] Werner Römisch and Rüdiger Schultz. Distribution sensitivity in stochastic programming. *Mathematical Programming*, 50(1):197–226, 1991.

[RU+00] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[RW91] R Tyrrell Rockafellar and Roger J-B Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16(1):119–147, 1991.

[RW09] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[RW21] Johannes O Royset and Roger JB Wets. An optimization primer, 2021.

[RZ96] Jörg Rambau and Günter M Ziegler. Projections of polytopes and the generalized baues conjecture. *Discrete & Computational Geometry*, 16(3):215–237, 1996.

[San12] Francisco Santos. A counterexample to the hirsch conjecture. *Annals of mathematics*, pages 383–412, 2012.

[SB18] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[SD20] Alexander Shapiro and Lingquan Ding. Periodical multistage stochastic programs. *SIAM Journal on Optimization*, 30(3):2083–2102, 2020.

[SDR14] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

[Sei95] R Seidel. The upper bound theorem for polytopes: an easy proof of its asymptotic version. *Computational Geometry*, 5(2):115 – 116, 1995.

[Sha06] Alexander Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34(1):1–8, 2006.

[Sha11] Alexander Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, 2011.

[SL15] Yongjia Song and James Luedtke. An adaptive partition-based approach for solving two-stage stochastic programs with fixed recourse. *SIAM Journal on Optimization*, 25(3):1344–1367, 2015.

[Sma98] Steve Smale. Mathematical problems for the next century. *The mathematical intelligencer*, 20(2):7–15, 1998.

[SN05] Alexander Shapiro and Arkadi Nemirovski. On complexity of stochastic programming problems. In *Continuous optimization*, pages 111–146. Springer, 2005.

[SS22] Murwan Siddig and Yongjia Song. Adaptive partition-based sddp algorithms for multistage stochastic linear programming with fixed recourse. *Computational Optimization and Applications*, 81(1):201–250, 2022.

[ST97] Bernd Sturmfels and Rekha R Thomas. Variation of cost functions in integer programming. *Mathematical Programming*, 77(2):357–387, 1997.

[Sta75] Richard P Stanley. The upper bound conjecture and cohen-macaulay rings. *Studies in Applied Mathematics*, 54(2):135–142, 1975.

[STdCS13] Alexander Shapiro, Wajdi Tekaya, Joari Paulo da Costa, and Murilo Pereira Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European journal of operational research*, 224(2):375–391, 2013.

[Stu94] Bernd Sturmfels. On the newton polytope of the resultant. *Journal of Algebraic Combinatorics*, 3(2):207–236, 1994.

[vAdOS18] Wim van Ackooij, Welington de Oliveira, and Yongjia Song. Adaptive partition-based level decomposition methods for solving two-stage stochastic programs with fixed recourse. *Informs Journal on Computing*, 30(1):57–70, 2018.

[VAdOS19] Wim Van Ackooij, Welington de Oliveira, and Yongjia Song. On level regularization with normal solutions in decomposition methods for multistage stochastic programming problems. *Computational Optimization and Applications*, 74(1):1–42, 2019.

[VSW69] Richard M Van Slyke and Roger Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

[VWBC05] Sven Verdoolaege, Kevin M Woods, Maurice Bruynooghe, and Ronald Cools. Computation and manipulation of enumerators of integer projections of parametric polytopes. *CW Reports*, pages 104–104, 2005.

[vzGS78] Joachim von zur Gathen and Malte Sieveking. A bound on solutions of linear integer equalities and inequalities. *Proceedings of the American Mathematical Society*, 72(1):155–158, 1978.

[Wal14] Stephen G Walker. On a lower bound for the jensen inequality. *SIAM Journal on Mathematical Analysis*, 46(5):3151–3157, 2014.

[WKS14] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[WW69] David Walkup and Roger Wets. Lifting projections of convex polyhedra. *Pacific Journal of Mathematics*, 28(2):465–475, 1969.

[WZ05] Stein W Wallace and William T Ziemba. *Applications of stochastic programming*. SIAM, 2005.

[ZAS19] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502, 2019.

[Zie12] Günter M Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 2012.

[ZKW21] Jianzhe Zhen, Daniel Kuhn, and Wolfram Wiesemann. Mathematical foundations of robust and distributionally robust optimization. *arXiv preprint arXiv:2105.00760*, 2021.

[ZS19] Shixuan Zhang and Xu Andy Sun. Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization. *arXiv preprint arXiv:1912.13278*, 2019.

[ZS20] Shixuan Zhang and Xu Andy Sun. On distributionally robust multistage convex optimization: New algorithms and complexity analysis. *arXiv preprint arXiv:2010.06759*, 2020.