HAL
open science

Optimisation stochastique à grande échelle

Claire Tauvel

To cite this version:

Claire Tauvel. Optimisation stochastique à grande échelle. Mathématiques [math]. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT : . tel-00364777

HAL Id: tel-00364777

https://theses.hal.science/tel-00364777v1

Submitted on 27 Feb 2009

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

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

Université Joseph Fourier

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

Spécialité : "Mathématiques Appliquées"

préparée au Laboratoire Jean Kuntzmann (LJK)
dans le cadre de **l'Ecole Doctorale "Mathématiques, Sciences et Technologie de l'Information, Informatique"**

présentée et soutenue publiquement

par

Claire TAUVEL

le 9 décembre 2008

Titre :

Optimisation stochastique à grande échelle

Directeur de thèse : Anatoli IOUDITSKI

JURY

M.	Bernard DELYON	, Rapporteur
M.	Jean-Claude FORT	, Rapporteur
M.	Anatoli IOUDITSKI	, Directeur de thèse
M.	Frédéric BONNANS	, Examinateur
M.	Emmanuel GOBET	, Examinateur
M.	Claude LEMARECHAL	, Examinateur

Table des matières

1 Introduction **7**

2 Algorithme de descente en miroir **15**

2.1 Introduction . . . 17
2.2 Algorithme pour les problèmes d'optimisation sans contrainte fonctionnelle . . . 17
2.2.1 Problème étudié . . . 17
2.2.2 Algorithme . . . 17
2.2.3 Deux exemples . . . 18
2.2.4 Vitesse de convergence et précision des algorithmes . . . 21
2.3 Généralisation à d'autres classes de problèmes . . . 24
2.3.1 Résolution d'inégalités variationelles à opérateur monotone . . . 25
2.3.2 Problèmes de recherche de point selle . . . 28
2.4 Choix de gain adaptatif . . . 32
2.4.1 Introduction . . . 32
2.4.2 Construction . . . 34
2.4.3 Bornes en espérance . . . 35
2.4.4 Bornes en déviation - Intervalles de confiance . . . 39
2.4.5 Gain adaptatif pour les problèmes d'inégalités variationnelles . . . 40
2.4.6 Gain adaptatif pour les problèmes de point selle . . . 42
2.5 Exemples et simulations . . . 43
2.5.1 Introduction . . . 43
2.5.2 Jeux matriciels . . . 44
2.5.3 Problème de localisation minmax . . . 49

3 Algorithme d'extrapolation dual **59**

3.1 Introduction . . . 61
3.2 Algorithme pour les problèmes de minimisation sans contrainte . . . 61
3.2.1 Algorithme . . . 61
3.2.2 Bornes en espérance . . . 62
3.2.3 Bornes en déviation, intervalles de confiance . . . 67
3.3 Algorithme pour les inégalités variationnelles . . . 68
3.4 Algorithme pour les problèmes de point-selle . . . 72
3.5 Choix de gain adaptatif . . . 78
3.5.1 Introduction . . . 78
3.5.2 Construction et bornes pour l'algorithme . . . 79

4 Optimisation stochastique sous contraintes **85**
4.1 Introduction . . . 87
4.2 Réduction à un problème unidimensionnel . . . 88
4.2.1 Définition et propriétés de la fonction f^* . . . 88
4.2.2 Résolution à l'aide d'une variante de l'algorithme de Newton . . . 91
4.2.3 Résolution par dichotomie . . . 93
4.2.4 Remarques et comparaison des deux algorithmes . . . 94
4.2.5 Lien avec les problèmes de point-selle stochastiques . . . 96
4.3 Algorithme . . . 97
4.3.1 Introduction - notations . . . 97
4.3.2 Inégalités de déviations, compléments . . . 98
4.3.3 Complexité : préliminaires . . . 100
4.3.4 Algorithme : complexité et précision . . . 102
4.4 Exemple : problème de Markowitz . . . 104

5 Conclusion et perspectives **107**

A Eléments d'analyse convexe, transformée de Legendre-Fenchel **111**
A.1 Fonctions convexes régulières et fonctions fortement convexes . . . 113
A.1.1 Fonctions convexes régulières . . . 113
A.1.2 Fonctions fortement convexes . . . 114
A.2 Transformée de Legendre d'une fonction convexe . . . 114
A.2.1 Fonctions semi-continues inférieurement . . . 114
A.2.2 Sousgradient, sousdifférentiel . . . 115
A.2.3 Transformée de Legendre d'une fonction convexe . . . 115
A.2.4 Transformée de Legendre d'une fonction fortement convexe . . . 116

B Inégalités de déviation **119**
B.1 Introduction . . . 121
B.1.1 Hypothèses . . . 121
B.1.2 Lemmes calculatoires . . . 122
B.2 Borne avec un moment exponentiel d'ordre 1 . . . 123
B.3 Borne avec un moment exponentiel d'ordre 2 . . . 127

C Preuves diverses **131**
C.1 Preuve des bornes pour l'algorithme de descente en miroir . . . 133
C.1.1 Lemme technique préliminaire . . . 133
C.1.2 Cas des inégalités variationelles à opérateur monotone . . . 134
C.1.3 Cas des problèmes de point selle . . . 136
C.2 Preuve des bornes pour l'algorithme de descente en miroir à gain adaptatif . . . 139
C.2.1 Lemme préliminaire . . . 139
C.2.2 Cas des problèmes d'inégalités variationnelles . . . 140
C.2.3 Cas des problèmes de point selle . . . 142
C.3 Preuve des bornes pour l'algorithme d'extrapolation dual . . . 146
C.3.1 Cas des problèmes d'inégalités variationnelles . . . 146
C.3.2 Cas des problèmes de point selle . . . 148
C.4 Preuve des résultats du chapitre 4 . . . 152

D Index des hypothèses **155**

Bibliographie **160**

Chapitre 1

Introduction

Les problèmes d'optimisation émergent naturellement dans de nombreux domaines d'applications. Quoi que les gens fassent, à un certain moment, ils ont besoin ou envie de réaliser ces choses de la meilleure façon possible. Cette volonté, une fois modélisée sous forme mathématique, aboutit à un problème d'optimisation. Les difficultés apparaissent lors de l'étape suivante qui consiste à essayer de résoudre le problème obtenu. En effet, en règle générale, les problèmes d'optimisation ne sont pas solubles : en l'absence de structure adéquate pour le problème, "la meilleure" solution, lorsqu'elle existe, n'est pas nécessairement calculable explicitement en pratique, et même en se contentant d'une "bonne" solution, les temps de calculs requis par un ordinateur actuel pour résoudre des problèmes généraux, même simples en apparence, serait totalement prohibitifs. Par exemple, supposons que nous souhaitions résoudre le problème de minimisation d'une fonction lipschitzienne f de paramètre L sur le cube $[0,1]$ de $\mathbb{R}^n$ et que nous disposons d'un oracle d'ordre 0, c'est-à-dire d'une sorte de boîte noire qui nous fournit à la demande les valeurs de f aux points $x \in \mathbb{R}^n$ dont nous avons besoin. Il est possible de montrer (voir [28]) que quelle que soit notre stratégie pour résoudre ce problème à l'aide de l'oracle, et même en demandant une précision raisonnable pour la solution, le temps de calcul nécessaire à un ordinateur actuel pour trouver une solution se compte en milliers d'années dès que la dimension n du problème dépasse 10.
Ce constat étant fait, il est illusoire de vouloir résoudre tous les problèmes d'optimisation ou des problèmes d'optimisation trop généraux. Il convient par conséquent de se limiter, si nous souhaitons obtenir des algorithmes utilisables dans les cas pratiques, à des classes beaucoup plus réduites de problèmes. Nous ne nous intéresserons donc ici qu'à des problèmes du type

$$\min_{x \in Q} f(x), \tag{1.1}$$

où f est une fonction convexe continue sur laquelle porteront éventuellement des hypothèses de régularité supplémentaires et Q est un sous ensemble convexe compact de $\mathbb{R}^n$ qui possède une "bonne" géométrie. Ces problèmes ont le mérite d'être à la fois suffisamment généraux pour couvrir de larges domaines d'applications, et suffisamment structurés pour permettre la mise en oeuvre de méthodes de résolution qui puissent être efficaces.
Nous ajouterons dans cette thèse deux difficultés au problème précédent :

- Nous supposerons que n est "grand", ce qui correspond aux domaines d'applications actuels : l'avènement des ordinateurs a permis le recueil et le traitement de quantités toujours plus importantes de données et la plupart des systèmes étudiés dépendent de plus d'un grand nombre de paramètres. Ceci aura pour conséquence que nous nous limiterons à l'utilisation d'un oracle du premier ordre : la "boîte noire" à laquelle nous ferons appel pour résoudre les différents types de problèmes que nous allons abordés ne pourra nous donner que les valeurs de f et de son sous-gradient. En effet, les dérivées d'ordre supérieur nécessiteraient à la fois trop d'espace en mémoire et un temps de traitement prohibitif si nous envisagions de les utiliser dans des calculs.
- Nous nous placerons dans le cas bruité, c'est-à-dire que les valeurs fournies par l'oracle ne seront pas exactes, mais sujettes à des aléas (nous parlerons donc d'oracle stochastique). Par exemple, pour le problème (1.1), un oracle $\mathcal{O}$ stochastique du premier ordre nous fournira, pour un point x donné, non pas $(f(x), f'(x))$, mais $(F(x,\omega), F'(x,\omega))$ où $F(x,\omega) = f(x) + \eta(\omega)$ et $F'(x,\omega) = f'(x) + \eta'(\omega)$ où η et η' sont des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ (dont nous omettrons généralement de faire mention dans la suite), spécifiques à l'oracle utilisé et constituent un bruit qui perturbe les observations des valeurs de f et f'. Ce "bruit" ou ces "erreurs" peuvent être, dans le monde physique, imputables à des erreurs de mesure des phénomènes observés, ou liées à la nature aléatoire des phénomènes étudiés.

L'objet de cette thèse va donc être de résoudre certains problèmes d'optimisation convexe stochastique en grande dimension.
Commençons par préciser quelques notions que nous allons utiliser. Supposons que nous avons un problème $\mathcal{P}$ (par exemple le problème (1.1) pour une certaine fonction f et un certain ensemble Q donnés) que nous souhaitons résoudre en utilisant une certaine méthode $\mathcal{M}$. Nous allons, dans ce but, utiliser un algorithme itératif, c'est-à-dire un algorithme dont le schéma général est le suivant

Algorithme 1.1

- *Entrée : un point de recherche initial x_0 et une précision $\epsilon > 0$.*
- *Initialisation : $k = 0$, $I_{-1} = \varnothing$. Ici k est le compteur du nombre d'itérations et I_k est l'ensemble des informations accumulées après k itérations.*
- *Boucle principale*
  1. *Appeler l'oracle $\mathcal{O}$ au point x_k.*
  2. *mettre à jour l'ensemble des informations $I_k = I_{k-1} \bigcup (x_k, \mathcal{O}(x_k))$.*
  3. *Appliquer les règles de la méthode $\mathcal{M}$ à I_k pour déterminer un nouveau point de recherche x_{k+1}.*
  4. *Faire un test pour le critère d'arrêt. S'il est atteint, construire un point solution x^k, sinon incrémenter k et aller à 1).*

  Fin de la boucle

Il nous faut maintenant savoir comment nous allons mesurer les performances de notre méthode $\mathcal{M}$ pour le problème $\mathcal{P}$. Nous nous intéresserons pour cela à deux notions étroitement liés :
- sa précision,
- sa complexité.

Notons x^N la solution approchée fournie par notre algorithme après N itérations. La précision $\epsilon_f(x^N)$ de l'algorithme mesurera l'écart entre la valeur de f en x^N et la vraie valeur minimale de f qui est $f(x^*)$, autrement dit ce sera la quantité

$$\epsilon_f(x^N) = f(x^N) - f(x^*)$$

(qui est positive). Nous ne chercherons pas à comparer directement les valeurs de x^N et de x^*. La raison en est que dans les problèmes que nous étudions, la convergence de x^N vers x^* peut être très lente (et même aussi lente que l'on veut pourvu que la fonction f soit choisie de manière adéquate). Comme l'oracle que nous utilisons est stochastique, nous nous intéresserons naturellement à deux mesures de risque :
- la quantité
$$\mathcal{R}_N(\mathcal{M}, \mathcal{P}) = \mathbb{E}[\epsilon_{f_\mathcal{P}}(x^N)] = \mathbb{E}[f_\mathcal{P}(x^N) - f_\mathcal{P}(x^*)]$$
  (où l'aléa provient de la nature stochastique de l'oracle) qui sera la précision de l'algorithme étudié pour le problème (1.1) et la fonction $f_\mathcal{P}$ (c'est-à-dire le problème P) considérée,
- la quantité
$$\mathcal{R}_N^\alpha(\mathcal{M}, \mathcal{P}) = \inf_{\lambda > 0} \mathbb{P}(\epsilon_{f_\mathcal{P}}(x^N) \geq \lambda) \leq \alpha,$$

qui permet de borner les déviations de $\epsilon_f(x^N)$ et d'obtenir des intervalles de confiance pour $f_{\mathcal{P}}(x^*)$.

Bien évidemment, obtenir une borne pour la précision pour chaque fonction f considérée offre peu d'intérêt. En effet, connaître f est nécessaire pour calculer précisément $\mathbb{E}[\epsilon_f(x^N)]$ et le fait de nous placer dans un contexte stochastique implique que nous ne connaissons justement pas f.

Par conséquent nous allons nous intéresser à des bornes en pire cas pour la précision : plutôt que de considérer ϵ_f, nous allons étudier la quantité

$$\epsilon(x^N) = \sup_{\mathcal{P}\in\mathcal{F}} \epsilon_f(x^N),$$

où $\mathcal{F}$ sera une classe de problèmes (par exemple l'ensemble des problèmes de type (1.1) où Q est donné et f est une fonction convexe continue sur Q dont le sous-gradient en tout point est borné par une certaine constante M donnée). De même, nous allons chercher à minimiser les deux risques que nous avons donnés en pire cas (c'est-à-dire pour le "pire" des problèmes $\mathcal{P}$ de la classe $\mathcal{F}$). autrement dit nous allons étudier les deux quantités

$$\mathcal{R}_N(\mathcal{M},\mathcal{F}) = \sup_{\mathcal{P}\in\mathcal{F}} \mathcal{R}_N(\mathcal{M},\mathcal{P}) \text{ et } \mathcal{R}_N^\alpha(\mathcal{M},\mathcal{F}) = \sup_{\mathcal{P}\in\mathcal{F}} \mathcal{R}_N^\alpha(\mathcal{M},\mathcal{P}).$$

La quantité $\mathcal{R}_N(\mathcal{M},\mathcal{F})$ peut aussi être vue comme une fonction v du nombre de pas N effectué dans l'algorithme, nous appellerons $v(N)$ la vitesse de convergence de l'algorithme (ou de la méthode $\mathcal{M}$)sur la classe $\mathcal{F}$.

Venons-en maintenant au problème de la complexité de la méthode, c'est-à-dire à l'*effort* qu'il va falloir fournir pour calculer la solution approximative. Comme pour la précision, nous nous placerons dans le cas minimax et nous chercherons à borner la complexité du problème (1.1) non pas pour une fonction f donnée, mais la complexité en pire cas lorsque f appartient à un certain ensemble $\mathcal{F}$. Nous allons envisager deux mesures de la complexité :

- La complexité analytique qui est le nombre d'appels à l'oracle requis pour résoudre le problème $\mathcal{P}$ avec une précision ϵ donnée.
- La complexité arithmétique qui est le nombre total d'opérations arithmétiques (travail de l'oracle et de la méthode inclus) requis pour résoudre le problème $\mathcal{P}$ avec une précision ϵ.

Cette seconde notion est bien entendu la plus réaliste, mais dans la mesure où la complexité arithmétique peut souvent se déduire aisément de la complexité analytique, c'est cette dernière notion que nous utiliserons dans la suite.

De nombreux algorithmes ont été conçus pour résoudre ce type de problèmes stochastiques, le premier historiquement étant la méthode de Robbins et Monro en 1951 dans [31]. Cette procédure a été longuement étudiée par la suite, à commencer par les travaux [8] et [32], et ont montré que l'on pouvait atteindre asymptotiquement une vitesse de convergence optimale en $\mathcal{O}(1/N)$ pourvu que la fonction objectif soit deux fois différentiable, fortement convexe et que les tailles des pas à chaque itération soit choisies de manière convenable. Une amélioration conséquente de cette méthode fut apportée par Polyak (voir [29] et [30]) qui suggéra de faire des pas de taille plus importante et d'effectuer des moyennes des différents points de recherche successifs : la vitesse de convergence des algorithmes n'est pas changée, mais les procédures obtenues sont d'un emploi plus souple. Ces idées étaient déjà présentent dans [23] où les algorithmes décrits permettent d'atteindre des vitesses de convergence en $\mathcal{O}(1/\sqrt{N})$, vitesse non améliorable dans le cas convexe général.

D'autres ont suivi par la suite, le lecteur intéressé pourra trouver les principaux dans [9] , [5] et les références qui s'y trouvent. Malheureusement, la plupart de ces algorithmes sont inadaptés aux

problèmes en grande dimension qui nécessitent, pour être résolus de manière efficace, d'exploiter au mieux la structure du problème étudié. Les premières méthodes ayant une flexibilité suffisamment grande pour permettre ce traitement du problème ont été proposées par Nemirovski et Yudin dans [22]. L'idée principale de ces méthodes est d'exploiter la structure duale du problème, en séparant la trajectoire du gradient dans l'espace "dual" de celle du minimiseur dans l'espace "primal". D'autres méthodes reposant sur cette même idée ont vu le jour par la suite, notamment la méthode d'extrapolation duale de Nesterov (voir [24]), elle-même inspirée la méthode d'extrapolation du gradient de Korpelevich (voir [16]).

Plan de la thèse

Nous étudions dans un premier temps dans cette thèse des méthodes de ce type pour divers problèmes sans contraintes fonctionnelles dans le cas où l'oracle est stochastique et du premier ordre et nous montrons qu'elles sont efficaces au sens où elles atteignent les bornes minimax en $\mathcal{O}(1/\sqrt{N})$ données dans [22] pour ces problèmes. Nous envisageons ensuite des versions adaptatives de ces algorithmes dont nous montrons les principales propriétés de convergence. Enfin nous traitons le cas des problèmes sous contraintes. Plus précisément, le plan de ce mémoire est le suivant :

Le chapitre 2 traite d'un algorithme que nous nommerons dans la suite algorithme de descente en miroir en référence à l'algorithme introduit par Nemirovski et Yudin dans [22] qui porte le même nom et dont il est inspiré. Nous y présentons les différents types de problèmes que nous étudions, à savoir les problèmes de minimisation sans contrainte fonctionnelle du type du problème 1.1 (section 2.2), les problèmes d'inégalités variationnelles à opérateur monotone du type

$$\text{Trouver } x^* \in Q : \ \langle g(x), x - x^* \rangle \geq 0 \ \forall x \in Q,$$

où Q est compact et g vérifie $\langle g(x) - g(y), x - y \rangle \geq 0 \ \forall x, y \in Q$, et les problèmes de point selle du type

$$\min_{x \in X} \max_{y \in Y} L(x, y),$$

où X et Y sont compacts et L est convexe en x et concave en y (section 2.3). Le problème de résolution d'inégalité variationnelles est une généralisation des deux autres problèmes envisagés. Néanmoins le cas particulier des problèmes de point selle est un exemple intéressant à traiter en soi, d'une part parce qu'il nous sera utile pour le chapitre 4 et d'autre part parce qu'il se prête bien au "découpage" de l'opérateur monotone sous-jacent en deux "morceaux" (correspondant aux parties concaves et convexes de la fonction objectif considérée) auxquels il est possible d'accorder un traitement différent. Pour chacun de ces trois problèmes, nous proposons des algorithmes destinés à les résoudre. Ces algorithmes font intervenir deux suites de paramètres que nous notons $(\gamma_i)_i$ (taille des pas) et $(\beta_i)_i$ (gain). Nous donnons les valeurs pour ces paramètres qui permettent de minimiser l'espérance de l'erreur d'approximation en fonction de la géométrie du problème. Cette géométrie intervient sous deux formes dans l'algorithme : d'une part par le choix de la norme $\| \cdot \|$ considérée sur notre ensemble Q (ou X et Y pour les problèmes de point selle), et d'autre part au travers d'une fonction que nous noterons V et qui nous permettra de mettre en place le jeu de miroir entre l'espace dit *primal* auquel appartiennent les points de recherche considérés à chaque itération et l'espace dit *dual* auquel appartiennent les gradients de la fonction objectif (f ou L selon

le problème) ou les valeurs prises par g dans le cas des problèmes variationnels. Nous montrons que pour tous ces algorithmes, la vitesse de convergence est en $\mathcal{O}((M+\sigma)/\sqrt{N})$ où M est une borne pour l'opérateur considéré (le gradient de la fonction objectif pour les problèmes de minimisation et de point selle et l'opérateur g lui-même pour les problèmes variationnels) mesurée avec la norme duale $\| \cdot \|_*$ de la norme $\| \cdot \|$ précédemment choisie (par exemple dans le cas des inégalités variationnelles, $M = sup_{x \in Q} \| g(x) \|_*$, σ est l'écart type du bruit généré dans l'oracle et N est le nombre d'itérations effectuées. Nous donnons aussi des bornes en déviations pour l'erreur d'approximation commise par ces différents algorithmes. Nous donnons ensuite des versions adaptatives des algorithmes précédents où le choix du gain β_i se fait de manière récursive, sans connaissance prérequise des paramètres de l'oracle (en particulier de σ) et nous donnons les bornes en espérance et en déviations correspondantes (section 2.4). Le chapitre se termine par quelques applications simples (section 2.5) destinées à illustrer la mise en œuvre des différents algorithmes.

Le chapitre 3 traite la version stochastique de l'algorithme d'extrapolation dual introduit par Nesterov dans [24]. Nous étudions les différences de cet algorithme avec celui de descente en miroir, en particulier lorsque les fonctions objectif sont suffisamment régulières, c'est-à-dire dans le cas ou l'opérateur du problème variationnel considéré est lipschitzien et non plus simplement borné (ou à variations bornées). Nous montrons comment il est possible, à l'aide d'un choix de gain pertinent, d'améliorer la vitesse de convergence de l'algorithme en prenant en compte ces hypothèses de régularité : lorsque l'opérateur est à variations bornées par M, les vitesses obtenues sont en $\mathcal{O}((M+\sigma)\sqrt{N})$, lorsque l'opérateur est lipschitzien de constante M, elles deviennent $\mathcal{O}(\frac{\sigma}{\sqrt{N}} + \frac{M}{N})$. Les problèmes étudiés sont par ailleurs du même type que ceux du chapitre précédent, à savoir les problèmes de minimisation sans contrainte fonctionnelle (section 3.2), les inégalités variationnelles à opérateur monotone (section 3.3) et les problèmes de point selle (section 3.4). Nous donnons là aussi des bornes en espérance et en déviations pour chacun des algorithmes. Enfin nous envisageons, comme dans le chapitre précédent, une version adaptative de l'algorithme considéré (section 3.5).

Le chapitre 4 est quant à lui consacré à un problème spécifique, celui de l'optimisation sous contraintes. Ce type de problème s'écrit sous la forme

$$\min_{x \in X} f_0(x) \text{ sous contraintes } f_j(x) \leq 0 \ , \ j = 1, \ldots, m \, ;$$

où les $f_j(x)$, $j = 0, \ldots, m$ sont des fonctions convexes continues et X est un ensemble convexe compact inclus dans $\mathbb{R}^n$. Nous montrons dans un premier temps comment ramener se problème en un problème de recherche de zéro d'une fonction f^* convexe ne dépendant que d'une seule variable réelle. Nous proposons alors deux algorithmes pour résoudre ce nouveau problème, l'un inspiré de la méthode de Newton et l'autre de l'algorithme de dichotomie, dont nous effectuons une comparaison détaillée (section 4.2). Puis, nous expliquons comment utiliser l'algorithme de recherche de point-selle étudié dans la section 2.3 pour obtenir une méthode générale de résolution des problèmes d'optimisation stochastiques convexes sous contraintes (section 4.3), nous donnons des bornes pour la précision de cette méthode ainsi que sa complexité algorithmique. Enfin une application est donnée en fin de chapitre (section 4.4).

Ce document comporte en outre 4 annexes :

– la première regroupe les principaux résultats d'analyse convexe et les principales propriétés de la transformée de Legendre-Fenchel qui sont nécessaires à la lecture du document,

- la seconde donne les inégalités de déviation qui sont utilisées dans les chapitres 2, 3 et 4 pour construire des bornes en déviations et des intervalles de confiance pour l'erreur d'approxiamtion des différents algorithmes,
- la troisième annexe regroupe les preuves omises dans les chapitres 2, 3 et 4 afin d'alléger et clarifier la lecture du document,
- la quatrième annexe regroupe quant à elle les différentes hypothèses utilisées dans les chapitres 2 et 3.

Chapitre 2

Algorithme de descente en miroir

2.1 Introduction

Dans ce chapitre nous étudions une version stochastique de l'algorithme de descente en miroir inspiré de l'algorithme du même nom introduit par Nemirovski et Yudin dans [22].
Nous étudierons tout d'abord l'algorithme dans sa version destinée à résoudre des problèmes d'optimisation sans contrainte fonctionnelle. Nous donnerons des bornes pour l'espérance et les probabilités de déviations de l'erreur au N-ième pas de l'algorithme. Nous étendrons ensuite l'algorithme et les résultats de convergence obtenus à d'autres classes de problèmes, à savoir les problèmes d'inégalité variationnelle à opérateur monotone et les problèmes de point-selle. Nous donnerons ensuite une version de l'algorithme de descente en miroir où le choix du gain se fait de manière adaptative pour chacun des types de problèmes étudiés. Enfin nous terminerons ce chapitre en présentant quelques exemples d'utilisation de ces algorithmes qui seront illustrés numériquement.
Les résultats d'analyse convexe nécessaires à la compréhension des différents problèmes et résultats de ce chapitre sont consignés dans l'annexe A.

2.2 Algorithme pour les problèmes d'optimisation sans contrainte fonctionnelle

2.2.1 Problème étudié

Nous cherchons à résoudre le problème de minimisation sans contrainte suivant :

$$\min_{x \in Q} f(x),$$

où Q est un ensemble convexe compact et f une fonction convexe. Nous supposons Q inclus dans l'espace vectoriel de dimension finie E muni de la norme $\| \cdot \|$ (dite norme primale). Nous notons E^* l'espace dual de E et $\| \cdot \|_*$ la norme duale associée et $\langle \cdot, \cdot \rangle$ le produit scalaire sur $E \times E^*$. f est supposée continûment différentiable et nous noterons f' son gradient (cette hypothèse peut être affaiblie et le sera par la suite). Nous supposons que $\operatorname{argmin}_{x \in Q} f(x) = \{x^*\}$ et on note $f(x^*) = f^*$. Nous supposons que f' vérifie l'hypothèse suivante :

Hypothèse 2.1 *f' est borné sur Q :*

$$\sup_{x \in Q} \| f'(x) \|_* = M < +\infty.$$

Nous disposons en outre d'un oracle qui nous fournit une version aléatoire (bruitée) Ψ de f' et nous notons $\eta = \Psi - f'$. Les hypothèses à faire concernant η seront discutées ultérieurement.

2.2.2 Algorithme

L'algorithme que nous allons utiliser pour résoudre notre problème nécessite l'utilisation d'une fonction auxiliaire V qui est construite de la façon suivante : on se donne une fonction d fortement convexe sur Q de paramètre α_0 pour la norme $\| \cdot \|$. On suppose que d est differentiable en un point $x_0 \in Q$ et on considère la fonction V donnée par

$$V(x) = d(x) - d(x_0) - \langle d'(x_0), x - x_0 \rangle.$$

Notons que V est α_0-fortement convexe et que pour tout $x \in Q$

$$V(x) \geq \frac{\alpha_0}{2} \parallel x - x_0 \parallel^2 .$$

On note par W_β la β-conjuguée de V (où $\beta > 0$) c'est-à-dire que W_β est la transformée de Legendre-Fenchel de βV et est donc donnée par :

$$\forall z \in E^* \; W_\beta(z) = \max_{x \in Q} \{\langle z, x\rangle - \beta V(x)\} .$$

L'algorithme met aussi en jeu deux suites de paramètres strictement positifs $(\gamma_i)_i$ et $(\beta_i)_i$ tels que $\beta_{i+1} \geq \beta_i$, $i \geq 0$ (cette dernière condition est purement d'ordre technique : son intérêt apparaîtra dans la suite de l'étude). L'algorithme est donné par

Algorithme 2.1

- *Initialisation :* $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour* $i = 1$ *à* N *faire*
  - $\zeta_i = \zeta_{i-1} - \gamma_i \Psi(x_{i-1})$
  - $x_i = W'_{\beta_i}(\zeta_i)$
- *Sortie :*

$$x^N = \frac{\sum_{i=1}^N \gamma_i x_{i-1}}{\sum_{i=1}^N \gamma_i}$$

L'algorithme construit donc deux suites, à savoir la suite (x_i) dans l'espace primal E et la suite (ζ_i) qui elle appartient à l'espace dual E^*. Chaque itération de l'algorithme a lieu en deux étapes :

- La mise à jour du paramètre ζ_i qui constitue l'étape de descente à proprement parler, le paramètre γ_i permettant de régler la taille du pas à effectuer.
- La mise à jour du paramètre x_i qui consiste à "renvoyer" le paramètre dual ζ_i que l'on vient de calculer dans l'espace primal à l'aide de la fonction V (via sa transformée W_β). C'est cette étape qui a motivé l'appellation d'algorithme "en miroir" : la fonction V, joue le rôle du miroir et permet de "réfléchir" la trajectoire $(\zeta_i)_i$ calculée dans l'espace dual pour obtenir la trajectoire primale $(x_i)_i$ (le choix de la fonction V sera donc crucial dans la mise en œuvre de l'algorithme) .

Notons enfin que l'étape finale qui consiste à prendre la moyenne pondérée de la trajectoire permet dans le cas stochastique d'assurer sur le plan théorique de bonnes propriétés de convergence.

2.2.3 Deux exemples

Donnons maintenant des exemples où nous savons calculer les fonctions W_β et W'_β : le cas où E est un espace euclidien et le cas où Q est un simplexe. Ces deux exemples seront réutilisés dans les applications en fin de chapitre.

Le cas euclidien

On se place dans un espace euclidien E muni de son produit scalaire et de la norme $\| \cdot \|_2$ induite par celui-ci. Comme E est euclidien, on a $E^* = E$. On suppose que Q est un convexe d'intérieur non vide et x_0 un point de l'intérieur de Q. On choisit comme fonction d la fonction $x \to \frac{1}{2} \| x \|_2^2$, et alors $V(x) = \frac{1}{2} \| x - x_0 \|_2^2$. Les fonctions V et d sont fortement convexes de paramètre 1 pour la norme $\| \cdot \|_2$. On cherche comme précédemment le minimum d'une fonction convexe sur Q. On suppose que f est continûment différentiable et qu'elle satisfait l'Hypothèse 2.1. On dispose de plus d'un oracle similaire au précédent.
L'Algorithme 2.1 devient alors (Π_Q désigne la projection orthogonale sur Q)

Algorithme 2.2

- *Initialisation :* $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour* $i = 1$ *à* N *faire*
  - $\zeta_i = \zeta_{i-1} - \gamma_i \Psi(x_{i-1})$
  - $x_i = \Pi_Q(x_0 + \frac{\zeta_i}{\beta_i})$
- *Sortie :*

$$x^N = \frac{\sum_{i=1}^N \gamma_i x_{i-1}}{\sum_{i=1}^N \gamma_i}$$

Dans le cas où Q est une boule fermée de centre x_0 et de rayon $R > 0$, l'étape de mise à jour de l'algorithme précédent devient

Mise à jour : pour $i = 1$ à N faire

- $\zeta_i = \zeta_{i-1} - \gamma_i \Psi(x_{i-1})$
- $x_i = x_0 + \frac{\zeta_i}{\beta_i} \min(1, \frac{R\beta_i}{\|\zeta_i\|_2})$

Ce dernier algorithme est très proche de l'algorithme du gradient stochastique "classique", il n'en diffère que par la manière dont la projection est faite. L'algorithme de descente en miroir que nous avons décrit peut donc être considéré comme une généralisation de la méthode du gradient stochastique. A la différence de l'algorithme classique, l'algorithme de descente en miroir ne "casse" pas la trajectoire duale : la seule projection qui a lieu se produit dans l'espace primal. En conséquence, l'essentiel du travail effectué par l'algorithme se produit dans l'espace dual. La trajectoire primale n'en est qu'un reflet renvoyé par la fonction W_β qui, comme nous l'avons déjà dit, joue le rôle du miroir.

Le cas du simplexe

On se place dans $E = \mathbb{R}^m$ et on suppose que Q est le simplexe :

$$Q = \Delta = \{x \in \mathbb{R}^m | x \geq 0, \sum_{j=1}^m x^{(j)} = 1\},$$

où $x^{(j)}$ désigne la j-ème composante du vecteur x. Le choix naturel pour la norme primale est alors la norme L_1 : $\| x \|_1 = \sum_{j=1}^m |x^{(j)}|$ et la norme duale sur E^* est la norme L_∞ donnée par $\| z \|_\infty = \max_{j=1\cdots m} z^{(j)}$. Un choix possible pour la fonction proximale V est la fonction entropie $V(x) = \ln(m) + \sum_{j=1\cdots m} x^{(j)} \ln(x^{(j)})$ dont le minimum est atteint en l'unique point $x^* = (1/m, \cdots 1/m)^T$ avec $V(x^* = 0)$. Cette fonction est α_0-fortement convexe de paramètre $\alpha_0 = 1$. De plus les fonctions W_β et ∇W_β peuvent être calculées explicitement et sont données par :

$$\begin{aligned} \forall z \in E^*,\, W_\beta(z) &= \beta \ln\left(\frac{1}{m}\sum_{i=1}^m e^{z^{(i)}/\beta}\right), \\ \frac{\partial W_\beta(z)}{\partial z^{(j)}} &= e^{z^{(j)}/\beta}\left(\sum_{i=1}^m e^{z^{(i)}/\beta}\right)^{-1}, \; j = 1, \cdots, m. \end{aligned}$$

On se donne comme précédemment deux suites de paramètres $(\gamma_i)_i$ et $(\beta_i)_i$, la seconde étant supposée croissante (au sens large). Le choix 'naturel' pour x_0 est le point où la fonction proximale (c'est-à-dire V) atteint son minimum (aussi appelé centre proximal), dans notre cas il s'agit du point $x_0 = (\frac{1}{m}, \cdots, \frac{1}{m})^T) \in \Delta$. L'algorithme 2.1 devient alors

Algorithme 2.3

- *Initialisation :* $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour* $i = 1$ *à* N *faire*
  - $\zeta_i = \zeta_{i-1} - \gamma_i \Psi(x_{i-1})$
  - $x_i^{(j)} = (W'_{\beta_i}(\zeta_i))^{(j)} = e^{\zeta_i^{(j)}/\beta_i}\left(\sum_{k=1}^m e^{\zeta^{(k)}/\beta_i}\right)^{-1}, \; j = 1, \cdots, m$
- *Sortie :*

$$x^N = \frac{\sum_{i=1}^N \gamma_i x_{i-1}}{\sum_{i=1}^N \gamma_i}$$

Remarques

- Le simplexe Δ peut être vu comme l'ensemble des lois de probabilités sur un ensemble fini à m éléments : si $x \in \Delta$, pour tout $1 \leq j \leq m$ $x^{(j)}$ est la probabilité attachée au j-ème élément. La fonction entropie $V(x)$ que nous avons utilisée comme fonction proximale correspond alors à la divergence de Kullback entre la loi uniforme et la loi donnée par x sur ce même ensemble.
- Au cours de la mise à jour dans l'algorithme, la nouvelle valeur du point de recherche x est calculée à l'aide du paramètre dual ζ. Par conséquent nous pouvons considérer x comme une fonction de la variable $z = -\zeta$. Plus précisément, x est donné par

$$x^{(j)} = e^{-z^{(j)}/\beta}\left(\sum_{j=1}^m e^{-z^{(j)}/\beta}\right)^{-1}, \; j = 1, \cdots, m.$$

  x définit donc une mesure de Gibbs sur les coordonnées de z, β étant interprété comme le paramètre de température.

2.2.4 Vitesse de convergence et précision des algorithmes

Nous allons maintenant chercher des bornes pour l'erreur commise par l'algorithme, à savoir la quantité $\varepsilon_N = f(x^N) - f^*$. Les bornes que nous obtiendrons dans la suite pour les autres types de problèmes seront semblables. Les preuves sont données ici à titre illustratif afin d'éclairer le rôle des différents paramètres de l'algorithme et la portée des hypothèses effectuées.

Majoration de l'erreur

Nous commençons par donner une borne pour ε_N qui nous servira à établir ensuite des bornes en espérance et pour les probabilités de déviation de l'erreur d'approximation commise par l'algorithme que nous avons donné.

Proposition 2.1

$$\left(\sum_{i=1}^{N}\gamma_i\right)\varepsilon_N \leq \sum_{i=1}^{N}\gamma_i(f(x_{i-1})-f(x^*)) \leq \beta_N V(x^*) - \sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1}-x^*\rangle + \sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2 . \tag{2.1}$$

Preuve
Puisque f est convexe on a

$$\varepsilon_N \leq \left[\sum_{i=1}^{N}\gamma_i(f(x_{i-1}) - f(x^*))\right]\left(\sum_{i=1}^{N}\gamma_i\right)^{-1} .$$

Alors

$$\begin{aligned}\left(\sum_{i=1}^{N}\gamma_i\right)\varepsilon_N &\leq \sum_{i=1}^{N}\gamma_i(f(x_{i-1}) - f(x^*)),\\ &\leq \sum_{i=1}^{N}\gamma_i\langle f'(x_{i-1}), x_{i-1} - x^*\rangle,\\ &= \sum_{i=1}^{N}\gamma_i\langle\Psi(x_{i-1}), x_{i-1} - x^*\rangle - \sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1} - x^*\rangle. \end{aligned} \tag{2.2}$$

Comme V est fortement convexe de paramètre α_0, $W_{\beta_{i-1}}$ est continûment différentiable et son gradient est lipschitzien de constante $(2\alpha_0\beta_{i-1})^{-1}$. C'est ici que l'intérêt de la condition $(\beta_i)_i$ croissante va intervenir : $\beta \to W_\beta(z)$ étant décroissante, on a pour tout i

$$\begin{aligned}W_{\beta_i}(\zeta_i) - W_{\beta_{i-1}}(\zeta_{i-1}) &\leq W_{\beta_{i-1}}(\zeta_i) - W_{\beta_{i-1}}(\zeta_{i-1}),\\ &\leq -\gamma_i\langle\Psi(x_{i-1}), x_{i-1}\rangle + \frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2 . \end{aligned} \tag{2.3}$$

A l'aide de (2.2) et (2.3), on obtient

$$\sum_{i=1}^{N}\gamma_i(f(x_{i-1}) - f(x^*)) \leq W_{\beta_0}(\zeta_0) - W_{\beta_N}(\zeta_N) + \langle x^*, \zeta_N\rangle - \sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1} - x^*\rangle + \sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2 .$$

D'autre part $W_{\beta_0}(\zeta_0) = 0$ et $\langle x^*, \zeta_N\rangle - W_{\beta_N}(\zeta_N) \leq \beta_N V(x^*)$, si bien que

$$\left(\sum_{i=1}^{N}\gamma_i\right)\varepsilon_N \leq \sum_{i=1}^{N}\gamma_i(f(x_{i-1})-f(x^*)) \leq \beta_N V(x^*)-\sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1}-x^*\rangle+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2 . \tag{2.4}$$

□

Borne en espérance

Nous allons maintenant utiliser l'équation 2.1 pour obtenir une borne en espérance pour ε_N. Comme notre algorithme est stochastique, il nous faut nous donner quelques hypothèses portant sur les variables aléatoires $\eta(x_i)$ pour étudier son comportement.
Notons $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$ et pour tout $i \geq 0$, $\mathcal{F}_i = \sigma\left(\mathcal{F}_{i-1}, \eta(x_{i-1})\right)$. Nous supposons que l'hypothèse suivante est vérifiée

Hypothèse 2.2 *Pour tout i, $\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0$ p.s., et $\mathbb{E}[\parallel \eta(x_i) \parallel_*^2 |\mathcal{F}_i] \leq \sigma^2 < +\infty$ p.s.*

Supposons de plus que l'hypothèse 2.1 est satisfaite. L'idée principale consiste à prendre l'espérance dans la borne que nous avons obtenue précédemment :

$$\varepsilon_N \leq \left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\left(\beta_N V(x^*) - \sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1}-x^*\rangle + \sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2\right).$$

et de choisir $(\gamma_i)_i$ et $(\beta_i)_i$ de façon à la minimiser. Il vient

$$\begin{aligned}\mathbb{E}\left[\varepsilon_N\right] &\leq \left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\mathbb{E}\left[\beta_N V(x^*) - \sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1}-x^*\rangle + \sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2\right],\\ &\leq \left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\left(\beta_N V(x^*) + \sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}(M+\sigma)^2\right).\end{aligned}$$

En effet, grâce à l'hypothèse 2.2 le terme central est d'espérance nulle, et l'utilisation conjointe des hypothèses 2.1 et 2.2 permet de majorer le terme de droite.
Le membre de droite de cette dernière inégalité est minimal quand les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ sont constantes : $\gamma_i \equiv \gamma$ et $\beta_i \equiv \beta$ avec la relation

$$\frac{\beta}{\gamma} = \frac{(M+\sigma)\sqrt{N}}{\sqrt{2\alpha_0 V(x^*)}}.$$

Alors on a

$$\mathbb{E}[\varepsilon_N] \leq 2\sqrt{\frac{V(x^*)}{2\alpha_0}}(M+\sigma)\frac{1}{\sqrt{N}}.$$

Si le nombre de pas qu'il faudra effectuer dans l'algorithme n'est pas connu à l'avance (par exemple dans le cas de traitement de données "on line"), les suites $(\gamma_i)_i$ et $(\beta_i)_i$ peuvent être choisies

de façon à ce qu'elles dépendent de i mais pas de N. Par exemple avec $\gamma_i \equiv 1$ et $\beta_i = (M + \sigma)(\alpha_0 V(x^*))^{-\frac{1}{2}}\sqrt{i+1}$, on obtient

$$\mathbb{E}[\varepsilon_N] \leq 2\sqrt{\frac{V(x^*)}{\alpha_0}}(M+\sigma)\frac{\sqrt{N+1}}{N}.$$

Remarques

- A priori, nous ne connaissons pas $V(x^*)$, par conséquent dans les cas de mise en œuvre pratique, ce paramètre sera majoré par $\overline{V} = \sup_{x\in Q} V(x)$ que nous saurons calculer en général.
- Toutes les preuves et algorithmes donnés ici peuvent s'étendre au cas où f est supposée seulement continue et convexe et où $f'(x)$ désigne un élément quelconque du sous-différentiel $\partial f(x)$ de f au point x. Nous nous placerons donc dans ce cadre par la suite.

Bornes en déviation - Intervalles de confiance pour ε_N

Afin de compléter notre étude de la convergence de l'algorithme, nous allons chercher des bornes pour les probabilités de déviation de ε_N. Ce type de résultats nous sera aussi utile dans le chapitre 4 consacré aux problèmes d'optimisation sous contraintes fonctionnelles.
Commençons par préciser les hypothèses sous lesquelles nous nous plaçons. Nous supposons encore que les hypothèses 2.1 et 2.2 sont vérifiées, nous utiliserons en plus l'une des deux hypothèses suivantes :

Hypothèse 2.3 *Il existe une constante* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa\frac{\| \eta(x_i) \|_*}{\sigma}\right)\middle| \mathcal{F}_i\right] \leq e \ p.s.$$

Hypothèse 2.4 *Il existe une constante* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa\frac{\| \eta(x_i) \|_*^2}{\sigma^2}\right)\middle| \mathcal{F}_i\right] \leq e \ p.s.$$

Les valeurs choisies ici pour les paramètres $(\beta_i)_i$ et $(\gamma_i)_i$ sont celles qui minimisaient la borne obtenue pour l'espérance de ε_N, à savoir $\gamma_i = 1$ et $\beta_i = \frac{(M+\sigma)\sqrt{N}}{\sqrt{2\alpha_0 V(x^*)}}$ pour tout i. Nous noterons D le diamètre de l'ensemble Q relativement à la norme considérée, autrement dit

$$D = \max_{x,u\in Q} \| x - u \| .$$

Les inégalités de déviations données dans l'annexe B permettent alors d'obtenir les deux propositions suivantes

Proposition 2.2 *On suppose que l'hypothèse 2.3 est vérifiée. Pour tout* $\alpha > 2\exp\left(-\frac{N}{4}\left(\frac{\varkappa}{2} + \frac{e}{\varkappa\sqrt{6\pi}}\right)\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{3\sqrt{2}}{2}\frac{(M+\sigma)}{\sqrt{N}}\sqrt{\frac{V(x^*)}{\alpha_0}} + \frac{D\sigma\lambda_1^*(N,\alpha)}{\sqrt{N}} + \sqrt{\frac{2V(x^*)}{\alpha_0}}\frac{\sigma^2}{(M+\sigma)}\frac{\lambda_2^*(N,\alpha)}{\sqrt{N}}$$

avec

$$\lambda_1^*(N,\alpha) = 2\sqrt{\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}}\sqrt{\ln(2\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \max\left\{\frac{37}{\varkappa^2 N^{\frac{1}{3}}}, \frac{4}{\varkappa^2 N}\left[\ln\left(2(Ne+1)\alpha^{-1}\right)\right]^2\right\},$$

on a

$$\mathbb{P}(\varepsilon_N > \lambda) \leq \alpha.$$

Proposition 2.3 *On suppose que l'hypothèse 2.4 est vérifiée. Pour tout* $\alpha > 2\exp\left(-\frac{\varkappa N}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{3\sqrt{2}}{2}\frac{(M+\sigma)}{\sqrt{N}}\sqrt{\frac{V(x^*)}{\alpha_0}} + \frac{D\sigma\lambda_1^*(N,\alpha)}{\sqrt{N}} + \sqrt{\frac{2V(x^*)}{\alpha_0}}\frac{\sigma^2}{(M+\sigma)}\frac{\lambda_2^*(N,\alpha)}{\sqrt{N}}$$

avec

$$\lambda_1^*(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(2\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(2\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N > \lambda) \leq \alpha.$$

Ces bornes en déviation nous permettent de mieux appréhender la précision de l'algorithme et peuvent être utiles, par exemple pour construire des intervalles de confiance pour ε_N. Plus précisément, $\lambda^*(N,\alpha)$ est la somme de trois termes :

- Un terme qui ne dépend pas de α, qui est fonction des choix que nous avons fait pour les paramètres (les deux suites $(\beta_i)_i$ et $(\gamma_i)_i$ mais aussi la fonction proximale V utilisée).
- Le terme en λ_1, qui donne les déviations de la différence de martingale

$$-\sum_{i=1}^{N}\gamma_i\langle\eta(x_{i-1}), x_{i-1} - x^*\rangle$$

  qui apparaît dans la borne (2.1). Ce terme est le terme dominant à α fixé lorsque N devient grand.
- le terme en λ_2 qui est issu du terme quadratique dans (2.1). Ce terme est négligeable devant les deux précédents lorsque, à α fixé, N est suffisamment grand.

Remarque : Cette décomposition en trois termes est typique des algorithmes que nous étudions ici et nous la retrouverons dans la plupart des bornes en déviations que nous donnerons par la suite (les exceptions étant les algorithmes à choix de gain adaptatif où apparaîtrons des termes supplémentaires).

2.3 Généralisation à d'autres classes de problèmes

Nous allons maintenant généraliser les résultats précédents au problème plus général de résolution d'inégalité variationnelles. Nous verrons ensuite une autre génération possible grâce au cas particulier des problèmes de point-selle stochastiques.

Les preuves des résultats qui sont énoncés dans cette partie sont sensiblement les mêmes que celles de la partie précédente. Afin de clarifier la lecture du document, elles n'apparaissent pas dans le corps du texte et sont consignées dans l'annexe C.

2.3.1 Résolution d'inégalités variationelles à opérateur monotone

Le problème que nous abordons ici est le problème le plus général que nous chercherons à résoudre. Comme nous le verrons, il recouvre les autres problèmes considérés.
Comme pour les algorithmes précédents le domaine d'intérêt est un ensemble convexe fermé Q inclus dans un espace vectoriel de dimension finie $(E, \| \cdot \|)$ et nous notons $(E^*, \| \cdot \|_*)$ son espace dual. Soit $g : Q \to E^*$ un opérateur continu et monotone sur Q :

$$\langle g(x) - g(y), x - y \rangle \geq 0 \ \forall x, y \in Q.$$

Nous souhaitons résoudre l'inégalité variationelle suivante :

$$\text{Trouver } x^* \in Q : \ \langle g(x), x - x^* \rangle \geq 0 \ \forall x \in Q. \tag{2.5}$$

Comme g est supposé continu et monotone, cette version dite *faible* du problème est équivalente à la version dite *forte* suivante :

$$\text{Trouver } x^* \in Q : \ \langle g(x^*), x - x^* \rangle \geq 0 \ \forall x \in Q.$$

par analogie avec ce qui a été fait pour le problème de minimisation, nous serons amené à considérer les hypothèses suivantes pour g :

Hypothèse 2.5 *L'opérateur g est borné sur Q, il existe une constante $M > 0$ telle que $\forall x \in Q \ \| g(x) \|_* \leq M$.*

Nous disposons d'un oracle qui nous fournit une version aléatoire (bruitée) Ψ de g et on note $\eta = \Psi - g$. Les hypothèses à faire concernant η seront discutées ultérieurement.

La mise en oeuvre de l'algorithme nécessite les mêmes ingrédients que celui de la partie précédente : nous avons besoin d'une fonction d fortement convexe sur Q de paramètre α_0 pour la norme $\| \cdot \|$. d est supposée différentiable en un point $x_0 \in Q$ et nous considérons la fonction V donnée par

$$V(x) = d(x) - d(x_0) - \langle d'(x_0), x - x_0 \rangle.$$

V est alors α_0-fortement convexe et pour tout $x \in Q$

$$V(x) \geq \frac{\alpha_0}{2} \| x - x_0 \|^2 .$$

Notons W_β la β-conjuguée de V (où $\beta > 0$). Nous avons aussi besoin de deux suites de paramètres strictement positifs $(\gamma_i)_i$ et $(\beta_i)_i$ tels que $\beta_{i+1} \geq \beta_i$, $i \geq 0$. Nous considérons alors l'algorithme suivant :

Algorithme 2.4

- *Initialisation :* $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour $i = 1$ à N faire*
  - $\zeta_i = \zeta_{i-1} + \gamma_i \Psi(x_{i-1})$
  - $x_i = W'_{\beta_i}(\zeta_i)$
- *Sortie :*

$$x^N = \frac{\sum_{i=1}^N \gamma_i x_{i-1}}{\sum_{i=1}^N \gamma_i}$$

Remarquons que le problème de minimisation d'une fonction convexe abordé dans la première partie de ce chapitre est un cas particulier de problème de résolution d'inégalité variationelle à opérateur monotone, il suffit de considérer l'opérateur $g = -f'$. Il n'est donc pas surprenant que les deux algorithmes mis en oeuvre soient presque les mêmes.

Nous donnons maintenant les principaux résultats de convergence pour cet algorithme, les preuves se trouvant dans l'annexe C.

On s'intéresse donc à la quantité $\epsilon(x^N) = \max_{x\in Q}\{\langle g(x), x^N - x\rangle\}$.

Proposition 2.4 *Pour tout $N > 0$, on a*

$$\epsilon(x^N) \leq \left(\sum_{i=1}^N \gamma_i\right)^{-1} \left(\max_{x\in Q} \beta_N V(x) + \max_{x\in Q} \sum_{i=1}^N \gamma_i \langle \eta(x_{i-1}), x - x_{i-1}\rangle + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2\right).$$

Nous voulons maintenant une borne en espérance pour $\epsilon(x^N)$. Commençons par donner un contexte probabiliste et quelques hypothèses sur les variables aléatoires $\eta(x_i)$.
On définit une famille de tribu $\mathcal{F}_i$ par $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$, et pour $i \geq 0$, $\mathcal{F}_i$ est la tribu engendrée par $\mathcal{F}_{i-1}$ et $\eta(x_i)$ de sorte que x_i est $\mathcal{F}_{i-1}$-mesurable.
On suppose que l'Hypothèse 2.5 est satisfaite et que

Hypothèse 2.6 *Pour tout i, $\mathbb{E}[\eta(x_i)|\mathcal{F}_{i-1}] = 0$ et $\mathbb{E}[\parallel \eta(x_i) \parallel_*^2 |\mathcal{F}_{i-1}] \leq \sigma^2$ p.s.*

On note $\overline{V} = \max_{x\in Q} V(x)$

Proposition 2.5 *On a*

$$\mathbb{E}[\epsilon(x^N)] \leq \left(\sum_{i=1}^N \gamma_i\right)^{-1} \left[\beta_N \overline{V} + 2\sigma\sqrt{\frac{\overline{V}\sum_{i=1}^N \gamma_i^2}{2\alpha_0}} + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}(M+\sigma)^2\right].$$

Quand $(\beta_i)_i$ et $(\gamma_i)_i$ sont constantes et vérifient

$$\frac{\beta_N}{\gamma_N} = \frac{(M+\sigma)\sqrt{N}}{\sqrt{2\alpha_0\overline{V}}},$$

l'expression précédente est minimale et devient

$$\mathbb{E}[\epsilon(x^N)] \leq \frac{2\sqrt{\overline{V}}(M+2\sigma)}{\sqrt{2\alpha_0 N}}.$$

Si on ne connaît pas N à l'avance, le choix $\gamma_i \equiv 1$ et $\beta_i = \frac{(M+\sigma)\sqrt{i+1}}{\alpha_0\overline{V}}$, permet d'obtenir :

$$\mathbb{E}[\epsilon(x^N)] \leq \left(\sqrt{2}M + (1+\sqrt{2})\sigma\right)\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{\sqrt{N+1}}{N}.$$

Nous voulons maintenant donner des bornes pour les déviations de $\epsilon(x^N)$. Nous supposerons, comme pour le problème précédent, l'existence d'un moment exponentiel, c'est-à-dire que l'une des deux hypothèses suivantes est vérifiée

Hypothèse 2.7 *Il existe* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$ $\mathbb{E}\left[\exp\left(\varkappa \frac{\|\eta(x_i)\|_*}{\sigma_i}\right) \middle| \mathcal{F}_i\right] \leq e$ *p.s.*

Hypothèse 2.8 *Il existe* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$ $\mathbb{E}\left[\exp\left(\varkappa \frac{\|\eta(x_i)\|_*^2}{\sigma_i^2}\right) \middle| \mathcal{F}_i\right] \leq e$ *p.s.*

Nous supposerons de plus que $\gamma_i = 1$ et $\beta_i = \frac{(M+\sigma)\sqrt{N}}{\sqrt{2\alpha_0 \overline{V}}}$ pour tout i.

Proposition 2.6 *On suppose que l'hypothèse 2.7 est vérifiée. Pour tout* $\alpha > 2\exp\left(-\frac{N}{4}\left(\frac{\varkappa}{2} + \frac{e}{\varkappa\sqrt{6\pi}}\right)\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{1}{\sqrt{N}}\left((3M+2\sigma)\sqrt{\frac{\overline{V}}{2\alpha_0}} + D\sigma\lambda_1^*(N,\alpha) + 3\sigma\sqrt{\frac{\overline{V}}{2\alpha_0}}\lambda_2^*(N,\alpha)\right),$$

où D *désigne le diamètre de l'ensemble* Q *et*

$$\lambda_1^*(N,\alpha) = 2\sqrt{\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}}\sqrt{\ln(2\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \max\left\{\frac{37}{\varkappa^2 N^{\frac{1}{3}}}, \frac{4}{\varkappa^2 N}\left[\ln\left(2(Ne+1)\alpha^{-1}\right)\right]^2\right\},$$

on a

$$\mathbb{P}(\epsilon(x^N) > \lambda) \leq \alpha.$$

Proposition 2.7 *On suppose que l'hypothèse 2.8 est vérifiée. Pour tout* $\alpha > 2\exp\left(-\frac{\varkappa N}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{1}{\sqrt{N}}\left((3M+2\sigma)\sqrt{\frac{\overline{V}}{2\alpha_0}} + D\sigma\lambda_1^*(N,\alpha) + 3\sigma\sqrt{\frac{\overline{V}}{2\alpha_0}}\lambda_2^*(N,\alpha)\right),$$

où D *désigne le diamètre de l'ensemble* Q *et*

$$\lambda_1^*(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(2\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(2\alpha^{-1})},$$

on a

$$\mathbb{P}(\epsilon(x^N) > \lambda) \leq \alpha.$$

Nous avons réussi à généraliser les résultats obtenus pour les problèmes de minimisation au cas des problèmes variationnels stochastiques à opérateurs monotones. Nous pouvons remarquer que les bornes obtenues sont pratiquement les mêmes que précédemment, en particulier nous retrouvons la même décomposition de λ^* en somme de trois termes. Nous allons maintenant aborder le traitement des problèmes de point selle par l'algorithme de descente en miroir.

2.3.2 Problèmes de recherche de point selle

Nous cherchons à trouver le point selle d'une fonction convexe-concave $L(x,y)$: L est supposée convexe en x et concave en y. Les variables x et y appartiennent respectivement à des ensembles convexes compacts X et Y. Le problème à résoudre est donc le suivant :

$$\min_{x\in X}\max_{y\in Y} L(x,y).$$

X et Y sont supposés inclus dans des espaces vectoriels de dimension finie E_X et E_Y munis respectivement des normes $\|\cdot\|_X$ et $\|\cdot\|_Y$. On note $(E_X^*, \|\cdot\|_{X,*})$ et $(E_Y^*, \|\cdot\|_{Y,*})$ les espaces duaux associés.
Notons $\partial_x L(x,y)$ le sous-différentiel de la fonction $x \to L(x,y)$ et $\partial_y L(x,y)$ le super-différentiel de la fonction $y \to L(x,y)$; $L'_X(x,y)$ (respectivement $L'_Y(x,y)$) désignera dans la suite un élément quelconque de $\partial_x L(x,y)$ (respectivement $\partial_y L(x,y)$). Nous noterons (x^*, y^*) un point-selle de L.
Les problèmes recherche de point selle rentrent, tout comme les problèmes d'optimisation sans contrainte fonctionnelle, dans le cadre plus général des problèmes de résolution d'inégalité variationelles. Il suffit pour le voir de choisir :

$$Q = X \times Y,\ g = \begin{pmatrix} L'_X \\ -L'_Y \end{pmatrix},\ \Psi = \begin{pmatrix} \Psi_X \\ -\Psi_Y \end{pmatrix},\ \text{et } \eta = \begin{pmatrix} \eta_X \\ -\eta_Y \end{pmatrix}.$$

Cependant, l'algorithme que nous allons donner ne sera pas tout à fait la traduction de l'algorithme 2.4 au cas particulier du point selle. En effet, afin de mieux prendre en compte la géométrie de notre problème, et en particulier le fait que X et Y peuvent avoir des structures très différentes, nous allons remplacer la suite $(\beta_i)_i$de l'algorithme 2.4 par deux suites $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ la première étant utilisée pour effectuer la descente en gradient dans E_X^*, et la seconde d'effectuer la remontée en gradient dans l'espace E_Y^*. Pour des raisons techniques qui apparaîtront dans la suite, il n'et pas possible d'effectuer de 'dédoublement' de ce genre pour la suite $(\gamma_i)_i$.

Par analogie avec les cas précédents, nous serons amenés à considérer l'hypothèse suivante :

Hypothèse 2.9 *Les sous-/super-différentiels de L sont bornés : pour tout $x \in X$ et tout $y \in Y$, on a*

$$\begin{aligned} \| L'_X(x,y) \|_{X,*} &\leq M_X, \\ \| L'_Y(x,y) \|_{Y,*} &\leq M_Y. \end{aligned}$$

Les réalisations des différentes fonctions qui sont mises en jeu dans l'algorithme (c'est-à-dire les fonctions L'_X et L'_Y) nous sont fournies par un oracle stochastique. Nous noterons $\eta_X(x,y) = \Phi(x,y) - L'_X(x,y)$ et $\eta_Y(x,y) = \Psi(x,y) - L'_X(x,y)$. Les hypothèses portant sur Φ et Ψ seront données lors de l'étude de la précision de l'algorithme.

Fixons un point initial $(x_0, y_0) \in X \times Y$ pour notre algorithme. Pour mettre à jour le point de recherche, nous aurons besoin de deux fonctions fortement convexes d_X et d_Y de paramètres de convexité égaux à $\alpha_{X,0}$ et $\alpha_{Y,0}$, l'une définie sur X et l'autre sur Y. Les fonctions proximales associées $V_X(x,x_0)$ et $V_Y(y,y_0)$ sont alors définies à l'aide de d_X et d_Y comme précédemment et on note $W_{X,\beta}(\cdot)$ et $W_{Y,\beta}(\cdot)$ leurs β-conjuguées respectives.
Pour la mise en oeuvre de l'algorithme, nous aurons besoin de trois suites de paramètres strictement positifs $(\gamma_i)_{i\geq 1}$, $(\beta_{X,i})_{i\geq 0}$ et $(\beta_{Y,i})_{i\geq 0}$, avec les conditions de monotonie $\beta_{X,i+1} \geq \beta_{X,i}$ and $\beta_{Y,i+1} \geq$

$\beta_{Y,i} \forall i \geq 0$. L'algorithme est donné par :

Algorithme 2.5

– *Initialisation* $(x_0, y_0) \in X \times Y$ *et* $(\zeta_0, \xi_0) = 0 \in E_X^* \times E_Y^*$.
– *Pour* $i = 1, \cdots, N$ *faire la mise à jour*

$$\begin{aligned} \zeta_i &= \zeta_{i-1} - \gamma_i \Phi(x_{i-1}, y_{i-1}), \\ \xi_i &= \xi_{i-1} + \gamma_i \Psi(x_{i-1}, y_{i-1}), \\ x_i &= \nabla W_{X,\beta_{X,i}}(\zeta_i), \\ y_i &= \nabla W_{Y,\beta_{Y,i}}(\xi_i). \end{aligned}$$

– *Sortie*

$$x^N = \frac{\sum_{i=1}^N \gamma_i \, x_{i-1}}{\sum_{i=1}^N \gamma_i}, \; y^N = \frac{\sum_{i=1}^N \gamma_i \, y_{i-1}}{\sum_{i=1}^N \gamma_i}.$$

Nous voulons pouvoir mesurer la qualité des solutions fournies par l'algorithme. Comme les problèmes de recherche de point-selle sont un cas particulier de problèmes de résolution d'inégalités variationelles, la fonction d'erreur considérée est la même et nous cherchons donc à évaluer la quantité $\Delta_N = \max_{y \in Y} L(x^N, y) - \min_{x \in X} L(x, y^N)$. Puisque $L(x, y)$ est une fonction convexe-concave, on a

$$\begin{aligned} \forall x \in X \; \sum_{i=1}^N \gamma_i L(x, y_{i-1}) &\leq \left(\sum_{i=1}^N \gamma_i \right) L(x, y^N), \\ \forall y \in Y \; \sum_{i=1}^N \gamma_i L(x_{i-1}, y) &\geq \left(\sum_{i=1}^N \gamma_i \right) L(x^N, y). \end{aligned}$$

Notons $\check{x} = \operatorname{argmin}_{x \in X} L(x, y^N)$ et $\check{y} = \operatorname{argmax}_{x \in X} L(x^N, y)$. Alors

$$0 \leq \left(\sum_{i=1}^N \gamma_i \right) \Delta_N \leq \sum_{i=1}^N \gamma_i (L(x_{i-1}, \check{y}) - L(\check{x}, y_{i-1})). \tag{2.6}$$

La proposition suivante permet de borner ce majorant de Δ_N

Proposition 2.8 *Pour tout* $x \in X$ *et* $y \in Y$ *:*

$$\begin{aligned} \sum_{i=1}^N [\gamma_i L(x_{i-1}, y) - \gamma_i L(x, y_{i-1})] \leq \; & \beta_{X,N} V_X(x, x_0) - \sum_{i=1}^N \gamma_i \langle (x_{i-1} - x), \eta_X(x_{i-1}, y_{i-1}) \rangle \\ & + \beta_{Y,N} V_Y(y, y_0) + \sum_{i=1}^N \gamma_i \langle (y_{i-1} - y), \eta_Y(x_{i-1}, y_{i-1}) \rangle \\ & + \sum_{i=1}^N \left(\frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}} + \frac{\gamma_i^2 \parallel \Psi(x_{i-1}, y_{i-1}) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}} \right). \end{aligned} \tag{2.7}$$

Notons alors $\overline{x} = \text{argmax}_{x\in X} \sum_{i=1}^N \gamma_i \langle x, \eta_X(x_{i-1}, y_{i-1})\rangle$ et $\overline{y} = \text{argmin}_{y\in Y} \sum_{i=1}^N \gamma_i \langle y, \eta_Y(x_{i-1}, y_{i-1})\rangle$. On déduit de la Proposition 2.8

$$\begin{aligned}\left(\sum_{i=1}^N \gamma_i\right)\Delta_N \quad &\leq \quad \beta_{X,N} V_X(\check{x}, x_0) - \sum_{i=1}^N \gamma_i \langle (x_{i-1} - \overline{x}), \eta_X(x_{i-1}, y_{i-1})\rangle \\ &\quad + \beta_{Y,N} V_Y(\check{y}, y_0) + \sum_{i=1}^N \gamma_i \langle (y_{i-1} - \overline{y}), \eta_Y(x_{i-1}, y_{i-1})\rangle \\ &\quad + \sum_{i=1}^N \left(\frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}} + \frac{\gamma_i^2 \parallel \Psi(x_{i-1}, y_{i-1}) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}} \right).\end{aligned} \tag{2.8}$$

Donnons-nous maintenant un contexte probabiliste et des hypothèses sur les variables aléatoires $\eta_X(x_i, y_i)$ et $\eta_Y(x_i, y_i)$.
Définissons la famille de tribu $(\mathcal{F}_i)_{i\geq -1}$ par $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$, et pour $i \geq 0$, $\mathcal{F}_i$ est la tribu engendrée par $\mathcal{F}_{i-1}$ et les variables aléatoires $\eta_X(x_i, y_i)$ et $\eta_Y(x_i, y_i)$, de sorte que x_i et y_i sont $\mathcal{F}_{i-1}$-mesurables. Nous supposerons que l'hypothèse suivante est satisfaite

Hypothèse 2.10 *Pour tout i, $\mathbb{E}[\eta_X(x_{i-1}, y_{i-1})|\mathcal{F}_{i-1}] = 0$, $\mathbb{E}[\eta_Y(x_{i-1}, y_{i-1})|\mathcal{F}_{i-1}] = 0$, $\mathbb{E}[\parallel \eta_X(x_{i-1}, y_{i-1}) \parallel_*^2 |\mathcal{F}_{i-1}] \leq \sigma_X^2$ et $\mathbb{E}[\parallel \eta_Y(x_{i-1}, y_{i-1}) \parallel_*^2 |\mathcal{F}_{i-1}] \leq \sigma_Y^2$ p.s.*

Notons $\overline{V}_X = \max_{x\in X} V(x, x_0)$ et $\overline{V}_Y = \max_{y\in Y} V(y, y_0)$. La proposition suivante borne l'espérance de Δ_N.

Proposition 2.9 *On a*

$$\begin{aligned}\mathbb{E}[\Delta_N] \leq \frac{\beta_{X,N}\overline{V}_X + \beta_{Y,N}\overline{V}_Y}{\sum_{i=1}^N \gamma_i} + \frac{\left(\sum_{i=1}^N \gamma_i^2\right)^{\frac{1}{2}}}{\sum_{i=1}^N \gamma_i} \left(\sigma_X \sqrt{\frac{2\overline{V}_X}{\alpha_{X,0}}} + \sigma_Y \sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}} \right) \\ + \frac{1}{\sum_{i=1}^N \gamma_i} \sum_{i=1}^N \frac{\gamma_i^2}{2} \left[\frac{(M_X + \sigma_X)^2}{\alpha_{X,0}\beta_{X,i-1}} + \frac{(M_Y + \sigma_Y)^2}{\alpha_{Y,0}\beta_{Y,i-1}} \right].\end{aligned}$$

Cette borne est minimale lorsque $\gamma_i \equiv 1$, $\beta_{X,i} \equiv \frac{(M_X+\sigma_X)\sqrt{N}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} \equiv \frac{(M_Y+\sigma_Y)\sqrt{N}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}$. On obtient alors

$$\mathbb{E}[\Delta_N] \leq \frac{2\sqrt{\overline{V}_X}(M_X + 2\sigma_X)}{\sqrt{2\alpha_{X,0}N}} + \frac{2\sqrt{\overline{V}_Y}(M_Y + 2\sigma_Y)}{\sqrt{2\alpha_{Y,0}N}}.$$

Si on ne connaît pas à l'avance le nombre de pas N à effectuer, le choix $\gamma_i \equiv 1$, $\beta_{X,i} = \frac{(M_X+\sigma_X)\sqrt{i+1}}{\sqrt{\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} \equiv \frac{(M_Y+\sigma_Y)\sqrt{i+1}}{\sqrt{\alpha_{Y,0}\overline{V}_Y}}$ permet d'obtenir

$$\mathbb{E}[\Delta_N] \leq \frac{2\sqrt{\overline{V}_X}(\sqrt{2}M_X + (1+\sqrt{2})\sigma_X)\sqrt{N+1}}{\sqrt{2\alpha_{X,0}N}} + \frac{2\overline{V}_Y(\sqrt{2}M_Y + (1+\sqrt{2})\sigma_Y)\sqrt{N+1}}{\sqrt{2\alpha_{Y,0}}N}.$$

$$\mathbb{E}[\Delta_N] \leq \left(\sqrt{2}M_X + (1+\sqrt{2})\sigma_X\right)\sqrt{\frac{2\overline{V}_X}{\alpha_{X,0}}}\frac{\sqrt{N+1}}{N} + \left(\sqrt{2}M_Y + (1+\sqrt{2})\sigma_Y\right)\sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}}\frac{\sqrt{N+1}}{N}.$$

Nous voulons maintenant donner des bornes pour les déviations de Δ_N. Nous nous placerons dans le cas où l'une des deux hypothèses suivantes est vérifiée

Hypothèse 2.11 *Il existe une constante* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$ $\mathbb{E}\left[\exp\left(\varkappa \frac{\|\eta_X(x_i,y_i)\|_{X,*}}{\sigma_X}\right)\middle| \mathcal{F}_i\right] \leq e$ *p.s. et*

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_Y(x_i,y_i) \|_{Y,*}}{\sigma_Y}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.12 *Il existe une constante* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$ $\mathbb{E}\left[\exp\left(\varkappa \frac{\|\eta_X(x_i,y_i)\|^2_{X,*}}{\sigma_X^2}\right)\middle| \mathcal{F}_i\right] \leq e$ *p.s. et*

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_Y(x_i,y_i) \|^2_{Y,*}}{\sigma_Y^2}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Nous supposerons de plus que $\gamma_i = 1$, $\beta_{X,i} = \frac{(M_X+\sigma_X)\sqrt{N}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} = \frac{(M_Y+\sigma_Y)\sqrt{N}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}$ pour tout i.

Proposition 2.10 *On suppose que l'hypothèse 2.11 est vérifiée. Pour tout* $\alpha > 4\exp\left(-\frac{N}{4}\left(\frac{\varkappa}{2} + \frac{e}{\varkappa\sqrt{6\pi}}\right)\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{1}{\sqrt{N}}\left((3M_X + 2\sigma_X)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + (3M_Y + 2\sigma_Y)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + (D_X\sigma_X + D_Y\sigma_Y)\lambda_1(N,\alpha)\right.$$
$$\left. + 3\sigma_X\lambda_2(N,\alpha)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + 3\sigma_Y\lambda_2(N,\alpha)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\right),$$

où D_X *et* D_Y *désignent les diamètres respectifs des ensembles* X *et* Y *et*

$$\lambda_1(N,\alpha) = 2\sqrt{\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}}\sqrt{\ln(4\alpha^{-1})} \text{ et } \lambda_2(N,\alpha) = \max\left\{\frac{37}{\varkappa^2 N^{\frac{1}{3}}}, \frac{4}{\varkappa^2 N}\left[\ln\left(4(Ne+1)\alpha^{-1}\right)\right]^2\right\},$$

on a

$$\mathbb{P}(\Delta_N > \lambda) \leq \alpha.$$

Proposition 2.11 *On suppose que l'hypothèse 2.12 est vérifiée. Pour tout* $\alpha > 4\exp\left(-\frac{\varkappa N}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{1}{\sqrt{N}}\left((3M_X + 2\sigma_X)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + (3M_Y + 2\sigma_Y)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + (D_X\sigma_X + D_Y\sigma_Y)\lambda_1(N,\alpha)\right.$$
$$\left. + 3\sigma_X\lambda_2(N,\alpha)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + 3\sigma_Y\lambda_2(N,\alpha)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\right),$$

où D_X *et* D_Y *désignent les diamètres respectifs des ensembles* X *et* Y *et*

$$\lambda_1(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(4\alpha^{-1})} \text{ et } \lambda_2(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(4\alpha^{-1})},$$

on a

$$\mathbb{P}(\Delta_N > \lambda) \leq \alpha.$$

Comme nous l'avons déjà dit, les problèmes de point selle sont un cas particulier des problèmes d'inégalités variationnelles, il n'est donc pas surprenant de retrouver dans les bornes obtenues dans les deux propositions précédentes le même type de décomposition pour λ^* que dans le cas des inégalités variationnelles, à savoir un terme constant qui dépend des choix de paramètres, un terme issu de la différence de martingale présente dans la borne pour l'erreur et un troisième terme provenant des termes quadratiques présents dans les différentes bornes obtenues pour Δ_N. Il est par contre intéressant de noter qu'au vu des bornes obtenues ici, tout ce passe comme si nous avions deux problèmes distincts : le premier ne dépendant que de la variable $x \in X$ et le second ne dépendant que de la variable $y \in Y$. Il n'y a en effet aucun terme qui couple les paramètres de la composante sur X et de la composante sur Y du problème.Dés lors, l'intérêt du dédoublement de la suite $(\beta_i)_i$ est clair : il permet de s'adapter à la géométrie et au bruit (c'est-à-dire aux variables aléatoires η_X et η_Y) de chacune des composantes (en X et en Y) du problème de point selle.

2.4 Choix de gain adaptatif

2.4.1 Introduction

Les différentes bornes que nous avons données pour l'espérance de l'erreur commise par les algorithmes que nous avons étudiés reposent sur un choix pertinent pour les valeurs prises par la suite $(\beta_i)_i$. Ce choix pertinent dépend de plusieurs constantes, certaines 'facilement' accessibles comme α_0 et $\overline{V}$ ou D (puisqu'elles sont fonction du choix de la fonction proximale à utiliser et du domaine de recherche), et d'autres que nous ne connaissons pas a priori comme σ et M, puisqu'elles requièrent des informations sur la fonction objectif (en particulier sa régularité) ou sur le bruit ambiant (l'oracle) auxquelles nous n'avons pas nécessairement accès.
Dans ce qui suit, on va expliciter comment construire de manière récursive la suite $(\beta_i)_i$ en prenant en compte les observations (i.e. les données fournies par l'oracle) de manière à ce que cette suite dépende explicitement d'aussi peu de paramètres non connu du problème que possible. L'objectif est bien entendu d'obtenir une version on line performante des différents algorithmes étudiés. Comme nous l'avons vu au travers du cas particulier de l'algorithme de descente en miroir pour la recherche de point-selle, la suite $(\beta_i)_i$ est celle que nous pouvons manipuler le plus facilement, tout en conservant les différentes bornes obtenues. Nous choisirons donc ici de garder la suite $(\gamma_i)_i$ constante et, pour alleger les notations, nous fixerons $\gamma_i \equiv 1$.
Avant cela et afin de donner une raison d'être supplémentaire au choix adaptatif du gain, revenons un instant en arrière et essayons d'améliorer la borne pour l'erreur dans l'algorithme 2.1 dans le cas où l'hypothèse suivante est vérifiée

Hypothèse 2.13 *f' est lipschitzien sur Q : il existe une constante positive $L < \infty$ telle que pour tous $x, x' \in Q$,*

$$\| f'(x) - f'(x') \|_* \leq L \| x - x' \| .$$

Remarque : Comme Q est supposé compact, cette hypothèse entraîne l'hypothèse 2.1.

Commençons par rappeler l'inégalité (2.4) :

$$N\varepsilon_N \leq \sum_{i=1}^{N}(f(x_{i-1}) - f(x^*)) \leq \beta_N V(x^*) - \sum_{i=1}^{N}\langle \eta(x_{i-1}), x_{i-1} - x^* \rangle + \sum_{i=1}^{N} \frac{1}{2\alpha_0\beta_{i-1}} \| \Psi(x_{i-1}) \|_*^2 .$$

D'après la proposition A.3, pour tout i

$$\begin{aligned} \| \Psi(x_{i-1}) \|_*^2 &\leq 2 \| \eta(x_{i-1}) \|_*^2 + 2 \| f'(x_{i-1}) \|_*^2, \\ &\leq 2 \| \eta(x_{i-1}) \|_*^2 + 4L \left(f(x_{i-1}) - f(x^*) \right), \end{aligned}$$

il vient

$$\sum_{i=1}^{N} \left(1 - \frac{2L}{\alpha_0 \beta_{i-1}} \right) (f(x_{i-1}) - f(x^*)) \leq \beta_N V(x^*) - \sum_{i=1}^{N} \langle \eta(x_{i-1}), x_{i-1} - x^* \rangle + \sum_{i=1}^{N} \frac{1}{\alpha_0 \beta_{i-1}} \| \eta(x_{i-1}) \|_*^2 .$$

Supposons que pour tout i $\beta_i \geq 4L\alpha_0^{-1}$, alors

$$N\varepsilon_N \leq \sum_{i=1}^{N} (f(x_{i-1}) - f(x^*)) \leq 2\beta_N V(x^*) - 2\sum_{i=1}^{N} \langle \eta(x_{i-1}), x_{i-1} - x^* \rangle + \sum_{i=1}^{N} \frac{2}{\alpha_0 \beta_{i-1}} \| \eta(x_{i-1}) \|_*^2 . \tag{2.9}$$

En particulier si $\sqrt{N} \geq \frac{4L}{\sigma} \sqrt{\frac{V(x^*)}{\alpha_0}}$, en choisissant la suite $(\beta_i)_i$ constante et égale à $\sigma \sqrt{\frac{N}{\alpha_0 V(x^*)}}$, il vient

$$\mathbb{E}[\varepsilon_N] \leq \frac{4\sigma}{\sqrt{N}} \sqrt{\frac{V(x^*)}{\alpha_0}}. \tag{2.10}$$

Contrairement à ce que nous avions dans le cas où le sous-gradient était borné, cette borne ne dépend plus du paramètre de régularité de la fonction. Remarquons que si la condition $\sqrt{N} \geq \frac{4L}{\sigma} \sqrt{\frac{V(x^*)}{\alpha_0}}$ n'est pas satisfaite, nous pouvons choisir $\beta_i \geq 4L\alpha_0^{-1}$ pour tout i et alors

$$\mathbb{E}[\varepsilon_N] \leq \frac{8L\, V(x^*)}{\alpha_0 N} + \frac{2\sigma}{\sqrt{N}} \sqrt{\frac{V(x^*)}{\alpha_0}}.$$

Nous avons donc obtenu

Proposition 2.12 *Supposons que l'hypothèse 2.13 est vérifiée. Alors*

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\sigma}{\sqrt{N}} \sqrt{\frac{V(x^*)}{\alpha_0}} + \max \left(\frac{8L\, V(x^*)}{\alpha_0 N}, \frac{2\sigma}{\sqrt{N}} \sqrt{\frac{V(x^*)}{\alpha_0}} \right).$$

Cette borne est obtenue lorsque l'algorithme 2.4 est mis en œuvre en choisissant les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes et respectivement égales à 1 et $\max \left(\frac{4L}{\alpha_0}, \sigma \sqrt{\frac{N}{\alpha_0 V(x^*)}} \right)$.

Remarque : Notons que dans le cas où l'oracle est déterministe (c'est-à-dire si $\eta(x_i) = 0$ pour tout i), le même phénomène à lieu. Pour des choix adéquats des paramètres $(\gamma_i)_i$ et $(\beta_i)_i$:

- si l'hypothèse 2.1 est vérifiée $\varepsilon_N \leq \sqrt{\frac{2V(x^*)}{\alpha_0}} \frac{M}{\sqrt{N}}$;
- si l'hypothèse 2.13 est vérifiée, $\varepsilon_N \leq \frac{8L\, V(x^*)}{\alpha_0 N}$.

Nous avons donc une meilleure vitesse de convergence lorsque la fonction f est plus régulière. Une des principales motivations du choix de gain adaptatif que nous allons proposer est l'obtention d'une propriété similaire pour la vitesse de convergence de l'algorithme dans le cas où l'oracle est stochastique mais sans que la régularité de la fonction f soit nécessairement connue. Autrement dit, nous souhaitons obtenir un algorithme qui soit capable, dans une certaine mesure de s'adapter à la régularité de la fonction objectif f. Pour cela, nous allons jouer sur la suite $(\beta_i)_i$ dont le choix, comme nous l'avons vu, est déterminant pour exploiter au mieux la régularité de la fonction objectif.

2.4.2 Construction

Il nous faut maintenant choisir les valeurs prises par les différents termes de la suite $(\beta_i)_i$ à l'aide des observations $\Psi(x_i)$. Pour cela revenons à la borne que nous avions obtenue pour l'algorithme 2.4 :

$$\varepsilon_N = f(x^{(N)}) - f(x^*) \leq \frac{\beta_N V(x^*)}{N} - \frac{1}{N}\sum_{i=1}^{N}\langle \eta(x_{i-1}), x_{i-1} - x^*\rangle + \frac{1}{N}\sum_{i=1}^{N}\frac{\| \Psi(x_{i-1}) \|_*^2}{2\alpha_0\beta_{i-1}},$$

où x^* désigne un point solution du problème. Si nous cherchons une borne en espérance pour ε_N, et puisque le second terme du majorant ci-dessus est d'espérance nulle, il faut que nous sachions contrôler le premier et le troisième terme. De ces deux termes, le second, à savoir

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\| \Psi(x_{i-1}) \|_*^2}{2\alpha_0\beta_{i-1}}$$

est le plus “complexe" puisqu'il dépend explicitement de tous les éléments de la suite $(\beta_i)_i$ et de toutes les observations $\Psi(x_i)$. Afin d'obtenir une expression plus simple pour ε_N, nous allons choisir les valeurs de la suite $(\beta_i)_i$ de façon à faire disparaître ce terme. Plus précisément, nous allons construire la suite $(\beta_i)_i$ par

$$\beta_{i+1} = \beta_i + \frac{u_{i+1}}{\beta_i} \text{ où } u_i = \frac{\| \Psi(x_{i-1}) \|_*^2}{\alpha_0 \overline{V}}. \tag{2.11}$$

où la valeur initiale β_0 est fixée (nous verrons comment plus loin). Notre nouvel algorithme s'écrit alors

Algorithme 2.6

- *Initialisation :* $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour* $i = 1$ *à* N *faire*
  - $\zeta_i = \zeta_{i-1} - \Psi(x_{i-1})$
  - $\beta_i = \beta_{i-1} + \frac{\|\Psi(x_{i-1})\|_*^2}{\alpha_0 \overline{V}\beta_{i-1}}$
  - $x_i = W'_{\beta_i}(\zeta_i)$
- *Sortie :*

$$x^N = \frac{1}{N}\sum_{i=1}^{N} x_{i-1}$$

En supposant que l'hypothèse 2.2 est toujours satisfaite, il vient

$$\varepsilon_N \leq \frac{2}{N}\beta_N\overline{V} - \frac{1}{N}\sum_{i=1}^{N}\langle \eta(x_{i-1}), x_{i-1} - x^*\rangle \text{ et } \mathbb{E}[\varepsilon_N] \leq \frac{2\mathbb{E}[\beta_N]\overline{V}}{N}.$$

L'étude de la convergence de l'algorithme se ramène donc pour l'essentiel à celle de β_N.

2.4.3 Bornes en espérance

Nous voulons maintenant donner des bornes en espérance pour β_N et ε_N lorsque l'une des hypothèses 2.1 et 2.13 est vérifiée.
Afin de fixer les idées, commençons par estimer l'ordre de grandeur de β_N. Il découle de (2.11) que

$$\begin{aligned}\beta_N^2 &= \beta_0^2 + 2\sum_{i=1}^N u_i + \sum_{i=1}^N \frac{u_i^2}{\beta_{i-1}^2},\\ &\leq \beta_0^2 + 2\sum_{i=1}^N u_i + \sum_{i=1}^N \frac{u_i^2}{\beta_0^2}.\end{aligned}$$

Par conséquent si nous supposons que les variables aléatoires $\eta(x_i)$ ont un moment d'ordre 4 fini, nous obtiendrons une borne pour $\mathbb{E}[\varepsilon_N]$. Plus précisément

Proposition 2.13 *Supposons qu'il existe une constante $\mu > 0$ finie telle que pour tout $i \geq 0$,*

$$\mathbb{E}[\| \eta(x_i) \|_*^4 \,|\mathcal{F}_i] \leq \mu^4.$$

Alors pour tout $\beta_0 > 0$,

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\overline{V}}{N}\sqrt{\beta_0^2 + \frac{2N}{\alpha_0\overline{V}}(M+\sigma)^2 + \frac{N}{\alpha_0^2\overline{V}^2\beta_0^2}(M+\mu)^4}.$$

Remarque : L'intérêt de cette proposition n'est pas de donner une borne précise pour $\mathbb{E}[\varepsilon_N]$: la majoration utilisée pour le terme $\sum_{i=1}^N \frac{u_i^2}{\beta_{i-1}^2}$ est beaucoup trop grossière. En revanche, elle nous incite à penser que β_i est de l'ordre de $\sqrt{i}$ pour tout i (une minoration de la valeur de β_i nécessiterait des informations plus précises sur la loi de répartition de $\eta(x_i)$ que celles données par l'hypothèse 2.2). Si c'est effectivement le cas et si le terme $\sum_{i=1}^N \frac{u_i^2}{\beta_{i-1}^2}$ est négligeable devant $\sum_{i=1}^N u_i$, alors le terme dominant dans la borne donnée dans la proposition 2.13 pour $\mathbb{E}[\varepsilon_N]$ est $2\sqrt{\frac{2\overline{V}}{N\alpha_0}}(M+\sigma)$ qui est comparable aux bornes obtenues en début de chapitre.

Compte tenu de la remarque précédente, nous allons nous donner des hypothèses supplémentaires sur la loi des variables aléatoires $\eta(x_i)$ afin d'améliorer la borne donnée dans la proposition 2.13. Plus précisément, nous supposerons que l'hypothèse 2.4 est vérifiée. Afin de rendre cette hypothèse plus facilement exploitable, nous utiliserons la majoration de β_N donnée dans la première inégalité du lemme C.3

$$\beta_N^2 \leq \beta_0^2 + 2\sum_{i=1}^N u_i + \frac{u_N^{*2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0}\sum_{i=1}^N u_i\right),$$

où $u_N^* = \max_{i=1,\cdots,N} u_i$. Nous allons maintenant majorer l'espérance des différents termes du membre de droite de cette dernière inégalité.

Borne en espérance lorsque l'hypothèse 2.1 est vérifiée

Lemme 2.1 *Supposons que l'hypothèse 2.1 est vérifiée. Pour tout $N \geq 3$*

$$\begin{aligned}
\mathbb{E}[u_N^{*2}] &\leq \frac{4\sigma^4}{\alpha_0^2 \overline{V}^2 \varkappa^2} \left(\varkappa \frac{M^2}{\sigma^2} + 1 + \ln(N) \right)^2, \\
\mathbb{E}\left[u_N^* \ln\left(1 + \frac{2}{\beta_0^2} \sum_{i=1}^N u_i \right) \right] &\leq \frac{2\sigma^2}{\alpha_0 \overline{V} \varkappa} \left(\ln(N) + \ln\left(1 + \frac{4\sigma^2}{\alpha_0 \overline{V} \varkappa \beta_0^2} \right) + \varkappa \frac{M^2}{\sigma^2} + 1 \right)^2.
\end{aligned}$$

Preuve
Remarquons que

$$\begin{aligned}
u_N^* &= \frac{2\sigma^2}{\alpha_0 \overline{V} \varkappa} \max_{i=1\ldots N} \frac{\varkappa \parallel \Psi(x_{i-1}) \parallel_*^2}{2\sigma^2}, \\
&\leq \frac{2\sigma^2}{\alpha_0 \overline{V} \varkappa} \ln\left(\sum_{i=1}^N \exp\left(\frac{\varkappa M^2}{\sigma^2} + \frac{\varkappa \parallel \eta(x_{i-1}) \parallel_*^2}{\sigma^2} \right) \right). \qquad (2.12)
\end{aligned}$$

La première inégalité du lemme se déduit alors du fait que la fonction $t \to (\ln(t))^2$ est concave sur $[e, +\infty[$ et de l'inégalité de Jensen.
Par ailleurs :

$$\begin{aligned}
\ln\left(1 + \frac{2}{\beta_0^2} \sum_{i=1}^N u_i \right) &= \ln\left(1 + \frac{4\sigma^2}{\alpha_0 \overline{V} \varkappa \beta_0^2} \sum_{i=1}^N \frac{\varkappa}{2\sigma^2} \parallel \Psi(x_{i-1}) \parallel_*^2 \right), \\
&\leq \ln\left(\left(1 + \frac{4\sigma^2}{\alpha_0 \overline{V} \varkappa \beta_0^2} \right) \sum_{i=1}^N \exp\left(\frac{\varkappa M^2}{\sigma^2} + \frac{\varkappa}{\sigma^2} \parallel \eta(x_{i-1}) \parallel_*^2 \right) \right).
\end{aligned}$$

Si bien que

$$u_N^* \ln\left(1 + \frac{2}{\beta_0^2} \sum_{i=1}^N u_i \right) \leq \frac{2\sigma^2}{\alpha_0 \overline{V} \varkappa} \left\{ \ln\left(\left(1 + \frac{4\sigma^2}{\alpha_0 \overline{V} \varkappa \beta_0^2} \right) \sum_{i=1}^N \exp\left(\frac{\varkappa M^2}{\sigma^2} + \frac{\varkappa}{\sigma^2} \parallel \eta(x_{i-1}) \parallel_*^2 \right) \right) \right\}^2. \qquad (2.13)$$

La seconde inégalité du lemme se déduit alors de cette dernière majoration à l'aide de l'inégalité de Jensen. □

Remarque : Comme l'hypothèse 2.13 implique l'hypothèse 2.1, le lemme 2.1 reste vrai si on suppose que l'hypothèse 2.13 est satisfaite.

Corollaire 2.1 *Supposons que l'hypothèse 2.1 soit vérifiée. Alors pour tout $N \geq 3$ et pour tout $\beta_0 > 0$*

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\beta_0 \overline{V}}{N} + \frac{2(M+\sigma)}{\sqrt{N}} \sqrt{\frac{2\overline{V}}{\alpha_0}} + \frac{\ln(N)^2}{N^{\frac{3}{2}}} \sqrt{\frac{2\overline{V}}{\alpha_0}} \frac{\sigma}{\varkappa} \left(\frac{4\sigma^2}{\beta_0^2 \alpha_0 \varkappa \overline{V}} + 1 \right) + \frac{C}{N^{\frac{3}{2}}},$$

où C est une constante.

Preuve
On déduit du lemme C.3 et de l'inégalité de Jensen que

$$\mathbb{E}[\beta_N] \leq \sqrt{\beta_0^2 + \frac{2N(M+\sigma)^2}{\alpha_0 \overline{V}} + \frac{\mathbb{E}[u_N^{*2}]}{\beta_0^2} + \frac{1}{2}\mathbb{E}\left[u_N^* \ln\left(1 + \frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)\right]}.$$

Soit

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\overline{V}}{N}\left(\beta_0^2 + 2\frac{2N(M+\sigma)^2}{\alpha_0\overline{V}} + \frac{4\sigma^4}{\beta_0^2\alpha_0^2\overline{V}^2\varkappa^2}\left(\varkappa\frac{M^2}{\sigma^2} + 1 + \ln(N)\right)^2 \right.$$
$$\left. + \frac{2\sigma^2}{\alpha_0\overline{V}\varkappa}\left(\varkappa\frac{M^2}{\sigma^2} + 1 + \ln(N) + \ln\left(1 + \frac{4\sigma^2}{\beta_0^2\alpha_0\overline{V}\varkappa}\right)\right)^2\right)^{\frac{1}{2}},$$

d'après le lemme 2.1

Puis en utilisant le fait que pour tous x, $y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$, il vient après simplification

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\beta_0\overline{V}}{N} + \frac{2(M+\sigma)}{\sqrt{N}}\sqrt{\frac{2\overline{V}}{\alpha_0}} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{\sigma}{\varkappa}\left(\frac{4\sigma^2}{\beta_0^2\alpha_0\varkappa\overline{V}} + 1\right) + \frac{C}{N^{\frac{3}{2}}},$$

où

$$C = \frac{\sigma}{\varkappa}\sqrt{\frac{\overline{V}}{2\alpha_0}}\left[\frac{4\sigma^2}{\beta_0^2\alpha_0\overline{V}\varkappa}\left(\varkappa\frac{M^2}{\sigma^2} + 1\right)^2 + \varkappa\frac{M^2}{\sigma^2} + 1 + \ln\left(1 + \frac{4\sigma^2}{\beta_0^2\alpha_0\overline{V}\varkappa}\right)^2\right].$$

□

Borne en espérance lorsque l'hypothèse 2.13 est vérifiée

Nous voulons améliorer la borne précédente dans le cas où l'hypothèse 2.13 est vérifiée.

Théorème 2.1 *Supposons que l'hypothèse 2.13 est vérifiée et que* $\beta_0 = 16L\alpha_0^{-1}$. *Alors pour tout* $N > 3$,

$$E[\varepsilon_N] \leq \frac{8\sigma}{\sqrt{N}}\sqrt{\frac{\overline{V}}{\alpha_0}} + \frac{64\overline{V}L}{\alpha_0 N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\frac{2\sigma}{\varkappa}\sqrt{\frac{\overline{V}}{\alpha_0}}\left(\frac{\alpha_0\sigma^2}{64\overline{V}\varkappa L^2} + 1\right) + \frac{C}{N^{\frac{3}{2}}},$$

où C *est une constante.*

Preuve
Puisque l'hypothèse 2.13 est vérifiée, on a

$$\begin{aligned}
2\sum_{i=1}^N u_i &= \frac{2}{\alpha_0\overline{V}}\sum_{i=1}^N \| \Psi(x_{i-1}) \|_*^2, \\
&\leq \frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N (\| f'(x_{i-1}) \|_*^2 + \| \eta(x_{i-1}) \|_*^2), \\
&\leq \frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{8L}{\alpha_0\overline{V}}\sum_{i=1}^N (f(x_i) - f(x^*)).
\end{aligned}$$

Isolons maintenant le dernier terme de l'inégalité précédente dans le majorant de β_N donné par le lemme C.3. Il vient

$$\begin{aligned}
\beta_N &\leq \sqrt{\beta_0^2 + \frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{8L}{\alpha_0\overline{V}}\sum_{i=1}^N (f(x_i)-f(x^*)) + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)}, \\
&\leq \sqrt{\beta_0^2 + \frac{8L}{\alpha_0\overline{V}}\sum_{i=1}^N (f(x_i)-f(x^*))} + \sqrt{\frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)}, \\
&\leq \beta_0 + \frac{4L}{\alpha_0\overline{V}\beta_0}\sum_{i=1}^N (f(x_i)-f(x^*)) + \sqrt{\frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)},
\end{aligned}$$

où la dernière inégalité provient du fait que $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$ pour tous x, $y > 0$.
Puisque $\varepsilon_N \leq \sum_{i=1}^N (f(x_i) - f(x^*))$, nous avons obtenu

$$\begin{aligned}
\sum_{i=1}^N (f(x_i)-f(x^*))(1-\frac{8L}{\alpha_0\beta_0}) &\leq 2\overline{V}\beta_0 - \sum_{i=1}^N \langle \eta(x_{i-1}), x_{i-1}-x^*\rangle \\
&\quad + 2\overline{V}\sqrt{\frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)}.
\end{aligned}$$

Si nous choisissons β_0 de sorte que $\frac{8L}{\alpha_0\beta_0} \leq \frac{1}{2}$, il vient

$$N\varepsilon_N \leq 4\overline{V}\beta_0 - 2\sum_{i=1}^N \langle \eta(x_{i-1}), x_{i-1}-x^*\rangle + 4\overline{V}\sqrt{\frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N \| \eta(x_{i-1}) \|_*^2 + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)}. \tag{2.14}$$

Le lemme 2.1 nous permet alors d'obtenir

$$\begin{aligned}
\mathbb{E}[\varepsilon_N] \leq \frac{64\overline{V}L}{\alpha_0 N} + \frac{4\overline{V}}{N}\Bigg(&\frac{4N\sigma^2}{\alpha_0\overline{V}} + \frac{\sigma^4}{32\overline{V}^2\varkappa^2L^2}\left[\left(\varkappa\frac{M^2}{\sigma^2}+1\right)^2 + \ln(N)^2\right] \\
&+ \frac{2\sigma^2}{\alpha_0\overline{V}\varkappa}\left[\ln(N)^2 + \left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{\alpha_0\sigma^2}{64\overline{V}\varkappa\beta_0^2}\right)\right)^2\right]\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

On utilise alors le fait que pour tous $x, y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$ pour obtenir

$$E[\varepsilon_N] \leq \frac{8\sigma}{\sqrt{N}}\sqrt{\frac{\overline{V}}{\alpha_0}} + \frac{64\overline{V}L}{\alpha_0 N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\frac{2\sigma}{\varkappa}\sqrt{\frac{\overline{V}}{\alpha_0}}\left(\frac{\alpha_0\sigma^2}{64\overline{V}\varkappa L^2}+1\right) + \frac{C}{N^{\frac{3}{2}}},$$

où

$$C = \frac{4\sigma}{\varkappa}\sqrt{\frac{\overline{V}}{\alpha_0}}\left[\frac{\sigma^2}{64\overline{V}\varkappa L^2}\left(\varkappa\frac{M^2}{\sigma^2}+1\right)^2 + \left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{\alpha_0\sigma^2}{64\overline{V}\varkappa\beta_0^2}\right)\right)^2\right].$$

□

Le théorème 2.1 appelle plusieurs remarques. Il est clair que les intérêts de cet algorithme ne se limitent pas aux possibilités d'exécution *on line*, qui comme nous l'avons vu, existaient déjà pour l'algorithme présenté au début de ce chapitre. Nous avons en plus apporté les améliorations suivantes :

- β_0 (et par conséquent, toute la suite $(\beta)_i$) ne dépend plus de σ, il n'est donc plus nécessaire de connaître ce paramètre pour mettre en oeuvre l'algorithme. Cette amélioration était nécessaire à la réalisation d'une version utilisable dans les cas pratiques des algorithmes que nous avons donnés. En effet, le paramètre σ est fonction de l'oracle et, à moins d'opter volontairement pour une stratégie randomisée pour la résolution d'un problème (comme c'est le cas par exemple pour les problèmes de jeux matriciels que nous aborderons plus loin dans ce chapitre), il est souvent difficilement accessible.
- L'algorithme obtenu est, dans une certaine mesure capable (ou tout du moins les bornes que nous avons obtenues le sont) de s'adapter à la régularité de la fonction objectif, il suffit pour cela de choisir β_0 de manière convenable. Plus précisément, dans le cas où la fonction considérée est à gradient borné, β_0 doit simplement être choisi strictement positif et dans le cas où le gradient de la fonction objectif est lipschitzien il suffit de le choisir suffisamment grand (le "suffisamment grand" dépendant de la constante de Lipschitz du gradient).

De plus, nous avons obtenu, des vitesses de convergences comparables à celles obtenues dans les cas où le gain est déterministe, ce qui signifie que le coût, en termes de temps de calcul, de ces améliorations est asymptotiquement négligeable. Plus précisément, nous avons toujours deux termes principaux : un terme qui ne dépend que du paramètre de régularité de la fonction (L ou M) et un terme qui ne dépend que de l'oracle et plus précisément du paramètre σ. Dans le cas où le choix du gain est adaptatif, il s'y rajoute deux termes qui décroissent comme $\frac{\ln(N)^2}{N^{\frac{3}{2}}}$ et $\frac{1}{N^{\frac{3}{2}}}$ respectivement et sont par conséquent négligeables, quand N est suffisamment grand, devant les deux premiers.

2.4.4 Bornes en déviation - Intervalles de confiance

Nous voulons maintenant obtenir des bornes en déviation lorsque l'hypothèse 2.4 et l'hypothèse 2.1 ou l'hypothèse 2.13 sont vérifiées.
Nous aurons besoin du lemme suivant

Lemme 2.2 *On suppose que les hypothèses 2.1 et 2.4 sont vérifiées. Alors pour tout N et tout $\alpha > 0$, on a*

$$\mathbb{P}\left(u_N^{*2} \geq \left\{\frac{2\sigma^2}{\varkappa\overline{V}\alpha_0}\ln\left(\exp(\varkappa M^2\sigma^{-2})\frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha,$$

$$\mathbb{P}\left(u_N^*\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right) \geq \frac{2\sigma^2}{\varkappa\overline{V}\alpha_0}\left\{\ln\left(\left(1+\frac{4\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\exp(\varkappa M^2\sigma^{-2})\frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha.$$

Preuve
Appliquons l'inégalité de Markov aux variables aléatoires $\exp\left(\frac{\varkappa}{\sigma^2}\parallel\eta(x_i)\parallel_*^2\right)$. Il vient pour tout $\lambda > 0$

$$\mathbb{P}\left(\sum_{i=1}^N \exp\left(\frac{\varkappa}{\sigma^2}\parallel\eta(x_i)\parallel_*^2\right) \geq \lambda\right) \leq \frac{Ne}{\lambda}.$$

Les inégalités recherchées découlent alors des inégalités (2.12), (2.13) et de la formule précédente avec $\lambda = \frac{Ne}{\alpha}$. □

Nous allons maintenant utiliser les propositions B.5, B.6 et le lemme précédent pour obtenir des bornes pour les déviations de ε_N.

Proposition 2.14 *Supposons que l'hypothèse 2.1 est vérifiée. Pour tout* $\alpha > 4\exp\left(-\frac{\varkappa N}{4}\right)$, *tout* $\beta_0 > 0$, *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{D\sigma}{N}\lambda_1(N,\alpha) + \frac{2\overline{V}}{N}\sqrt{\beta_0^2 + \frac{4N}{\alpha_0\overline{V}}(M^2+\sigma^2+\lambda_2(N,\alpha)) + \left(\frac{2\sigma^2}{\varkappa\overline{V}\alpha_0\beta_0}\right)^2\lambda_3(N,\alpha) + \frac{\sigma^2}{\varkappa\overline{V}\alpha_0\beta_0}\lambda_4(N,\alpha)},$$

avec

$$\lambda_1(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(4\alpha^{-1})},\ \lambda_2(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(4\alpha^{-1})},\ \lambda_3(N,\alpha) = \left\{\ln\left(\exp(\varkappa M^2\sigma^{-2})\frac{4N}{e\alpha}\right)\right\}^2$$

et

$$\lambda_4(N,\alpha) = \left\{\ln\left(\left(1+\frac{4\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\exp(\varkappa M^2\sigma^{-2})\frac{4N}{e\alpha}\right)\right\}^2,$$

on a

$$\mathbb{P}(\epsilon_N > \lambda) \leq \alpha.$$

Proposition 2.15 *Supposons que l'hypothèse 2.13 est vérifiée.Pour tout* $\alpha > 4\exp\left(-\frac{\varkappa N}{4}\right)$, *tout* β_0 *tel que* $\frac{8L}{\alpha_0\beta_0} \leq \frac{1}{2}$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{4\overline{V}\beta_0}{N} + \frac{2D\sigma}{\sqrt{N}}\lambda_1(N,\alpha) + \frac{4\overline{V}}{N}\sqrt{\frac{4\sigma^2 N}{\alpha_0\overline{V}}(1+\lambda_2(N,\alpha)) + \left(\frac{2\sigma^2}{\varkappa\overline{V}\alpha_0\beta_0}\right)^2\lambda_3(N,\alpha) + \frac{\sigma^2}{\varkappa\overline{V}\alpha_0\beta_0}\lambda_4(N,\alpha)},$$

avec

$$\lambda_1(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(4\alpha^{-1})},\ \lambda_2(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(4\alpha^{-1})},\ \lambda_3(N,\alpha) = \left\{\ln\left(\exp(\varkappa M^2\sigma^{-2})\frac{4N}{e\alpha}\right)\right\}^2$$

et

$$\lambda_4(N,\alpha) = \left\{\ln\left(\left(1+\frac{4\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\exp(\varkappa M^2\sigma^{-2})\frac{4N}{e\alpha}\right)\right\}^2,$$

on a

$$\mathbb{P}(\epsilon_N > \lambda) \leq \alpha.$$

2.4.5 Gain adaptatif pour les problèmes d'inégalités variationnelles

Nous objectif est de mettre en œuvre la méthode de construction adaptative du gain que nous venons de décrire pour résoudre les problèmes d'inégalités variationnelles. Nous avons déjà abordés ces problèmes dans la partie 2.3.1 de ce chapitre, nous conserverons ici les mêmes notations. Afin de clarifier la lecture, la plupart des preuves de cette partie ont été omises et reportées dans l'annexe

C.
Comme nous l'avons déjà vu, l'algorithme de descente en miroir nécessite deux suites de paramètres $(\gamma_i)_i$ et $(\beta_i)_i$. Nous allons fixons arbitrairement les termes de la suite $(\gamma_i)_i$ à 1 et une valeur $\beta_0 > 0$. Les autres termes de la suite $(\beta_i)_i$ sont alors définis récursivement par

$$\beta_{i+1} = \beta_i + \frac{u_{i+1}}{\beta_i} \text{ où } u_i = \frac{\| \Psi(x_{i-1}) \|_*^2}{\alpha_0 \overline{V}}.$$

La version adaptative de l'algorithme 2.4 est alors donnée par

Algorithme 2.7

- *Initialisation :* $x_0 \in Q$, $\zeta_0 = 0 \in E^*$
- *Mise à jour : pour* $i = 1$ *à* N *faire*
  - $\zeta_i = \zeta_{i-1} - \Psi(x_{i-1})$
  - $\beta_i = \beta_{i-1} + \frac{\|\Psi(x_{i-1})\|_*^2}{\alpha_0 \overline{V} \beta_{i-1}}$
  - $x_i = W'_{\beta_i}(\zeta_i)$
- *Sortie :*

$$x^N = \frac{1}{N} \sum_{i=1}^{N} x_{i-1}$$

Nous voulons maintenant des bornes en espérance et en déviations pour $\epsilon(x^N) = \max_{x \in Q} \langle g(x), x^N - x \rangle$. Nous supposerons dans la suite que les hypothèses 2.5, 2.6 et 2.8 sont vérifiées

Proposition 2.16 *Pour tout* $\beta_0 > 0$ *et tout* $N \geq 3$,

$$\mathbb{E}[\epsilon(x^N)] \leq \frac{2(3\sigma + 2M)}{\sqrt{N}} \sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{2\overline{V}\beta_0}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}} \sqrt{\frac{\overline{V}}{2\alpha_0}} \frac{2\sigma^2}{\varkappa(M+\sigma)} \left(1 + \frac{4\sigma^2}{\alpha_0 \beta_0^2 \overline{V} \varkappa}\right) + \frac{C}{N^{\frac{3}{2}}}$$

où C *est une constante.*

Afin de compléter cette étude, donnons maintenant une borne pour les déviations de $\epsilon(x^N)$.

Proposition 2.17 *On suppose que l'hypothèse 2.8 est vérifiée. Alors* $\forall \alpha \geq 4 \exp(-\frac{\varkappa N}{4})$, *tout* $\beta_0 > 0$ *et tout* $\lambda \geq \lambda^*(N, \alpha)$ *où*

$$\begin{aligned} \lambda^*(N,\alpha) &= \frac{\sigma}{\sqrt{N}} \sqrt{\frac{2\overline{V}}{\alpha_0}} + \frac{D\sigma}{\sqrt{N}} \lambda_1(N,\alpha) + \frac{\sigma}{\sqrt{N}} \sqrt{\frac{\overline{V}}{2\alpha_0}} \lambda_2(N,\alpha) \\ &\quad + \frac{2\overline{V}}{N} \sqrt{\beta_0^2 + \frac{4N}{\alpha_0 \overline{V}} (M^2 + \sigma^2 + \sigma^2 \lambda_2(N,\alpha)) + \left(\frac{2\sigma^2}{\varkappa \overline{V} \alpha_0 \beta_0}\right)^2 \lambda_3(N,\alpha) + \frac{\sigma^2}{\varkappa \overline{V} \alpha_0 \beta_0} \lambda_4(N,\alpha)}, \end{aligned}$$

avec

$$\lambda_1(N,\alpha) = \sqrt{\frac{4}{\varkappa} \ln(4\alpha^{-1})},\ \lambda_2(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N} \ln(4\alpha^{-1})},\ \lambda_3(N,\alpha) = \left\{ \ln\left(\exp(\varkappa M^2 \sigma^{-2}) \frac{4N}{e\alpha} \right) \right\}^2,$$

et

$$\lambda_4(N,\alpha) = \left\{ \left(\left(1 + \frac{4\sigma^2}{\varkappa \overline{V} \alpha_0 \beta_0^2} \right) \exp(\varkappa M^2 \sigma^{-2}) \frac{4N}{e\alpha} \right) \right\},$$

on a

$$\mathbb{P}(\epsilon(x^N) > \lambda) \leq \alpha.$$

2.4.6 Gain adaptatif pour les problèmes de point selle

Afin de compléter l'étude de la version adaptative de l'algorithme de descente en miroir, il nous reste à traiter le cas des problèmes de point selle. Les hypothèses et notations utilisés ici seront les mêmes que dans la partie 2.3.2 de ce chapitre.
La version de l'algorithme de descente en miroir que nous avons mis en œuvre pour les problèmes de point selle nécessite trois suites de paramètres $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$. Nous allons fixer arbitrairement les termes de la suite $(\gamma_i)_i$ à 1 et deux valeur $\beta_{X,0} > 0$ et $\beta_{Y,0} > 0$. Les autres termes des suites $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ sont alors définis récursivement par

$$\beta_{X,i+1} = \beta_{X,i} + \frac{u_{i+1}}{\beta_{X,i}} \text{ où } u_i = \frac{\| \Phi(x_{i-1}, y_{i-1}) \|_{X,*}^2}{\alpha_{X,0} \overline{V}_X},$$

et

$$\beta_{Y,i+1} = \beta_{Y,i} + \frac{v_{i+1}}{\beta_{Y,i}} \text{ où } v_i = \frac{\| \Psi(x_{i-1}, y_{i-1}) \|_{Y,*}^2}{\alpha_{Y,0} \overline{V}_Y},$$

La version adaptative de l'algorithme 2.5 est alors donnée par

Algorithme 2.8

- *Initialisation* $(x_0, y_0) \in X \times Y$ *et* $(\zeta_0, \xi_0) = 0 \in E_X^* \times E_Y^*$.
- *Pour* $i = 1, \cdots, N$ *faire la mise à jour*

$$\begin{aligned}
\zeta_i &= \zeta_{i-1} - \Phi(x_{i-1}, y_{i-1}), \\
\xi_i &= \xi_{i-1} + \Psi(x_{i-1}, y_{i-1}), \\
\beta_{X,i} &= \beta_{X,i-1} + \frac{\| \Phi(x_{i-1}, y_{i-1}) \|_{X,*}^2}{\alpha_{X,0} \overline{V}_X \beta_{X,i-1}}, \\
\beta_{Y,i} &= \beta_{Y,i-1} + \frac{\| \Psi(x_{i-1}, y_{i-1}) \|_{Y,*}^2}{\alpha_{Y,0} \overline{V}_Y \beta_{Y,i-1}}, \\
x_i &= W'_{X,\beta_{X,i}}(\zeta_i), \\
y_i &= W'_{Y,\beta_{Y,i}}(\xi_i).
\end{aligned}$$

- *Sortie*

$$x^N = \frac{1}{N} \sum_{i=1}^{N} x_{i-1}, \; y^N = \frac{1}{N} \sum_{i=1}^{N} y_{i-1}.$$

En vue d'étudier la convergence de cet algorithme, nous supposons dans la suite que les hypothèses 2.9, 2.10 et 2.12 sont satisfaites et nous nous intéressons à la quantité

$$\Delta_N = \max_{y\in Y} L(x^N, y) - \min_{x\in X} L(x, y^N).$$

Proposition 2.18 *Pour tout $\beta_{X,0} > 0$, tout $\beta_{Y,0} > 0$ et tout $N \geq 3$,*

$$\mathbb{E}[\Delta_N] \leq \frac{2(3\sigma_X + 2M_X)}{\sqrt{N}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + \frac{2\overline{V}_X\beta_{X,0}}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{2\sigma_X^2}{\varkappa(M_X+\sigma_X)}\left(1 + \frac{4\sigma_X^2}{\alpha_{X,0}\beta_{X,0}^2\overline{V}_X\varkappa}\right)$$
$$+ \frac{2(3\sigma_Y + 2M_Y)}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + \frac{2\overline{V}_Y\beta_{Y,0}}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{2\sigma_Y^2}{\varkappa(M_Y+\sigma_Y)}\left(1 + \frac{4\sigma_Y^2}{\alpha_{Y,0}\beta_{Y,0}^2\overline{V}_Y\varkappa}\right) + \frac{C}{N^{\frac{3}{2}}},$$

où C est une constante.

La proposition suivante donne des bornes pour les déviations de Δ_N

Proposition 2.19 *On suppose que l'hypothèse 2.12 est satisfaite. Alors $\forall\alpha \geq 8\exp\left(-\frac{\varkappa N}{4}\right)$, tout $\beta_{X,0} > 0$, tout $\beta_{Y,0} > 0$ et tout $\lambda \geq \lambda^*(N,\alpha)$ où*

$$\lambda^*(N,\alpha) = (D_X\sigma_X + D_Y\sigma_Y)\frac{\lambda_1(N,\alpha)}{\sqrt{N}} + \left(\frac{\sigma_X}{\sqrt{N}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\right)(2 + \lambda_2(N,\alpha))$$
$$+ \frac{2\overline{V}_X}{N}\sqrt{\beta_{X,0}^2 + \frac{4N}{\alpha_{X,0}\overline{V}_X}\left(M_X^2 + \sigma_X^2 + \sigma_X^2\lambda_2(N,\alpha)\right) + \left(\frac{2\sigma_X^2}{\varkappa\overline{V}_X\alpha_{X,0}\beta_{X,0}}\right)^2\lambda_{3,X}(N,\alpha) + \frac{\sigma_X^2\lambda_{4,X}(N,\alpha)}{\varkappa\overline{V}_X\alpha_{X,0}\beta_{X,0}}}$$
$$+ \frac{2\overline{V}_Y}{N}\sqrt{\beta_{Y,0}^2 + \frac{4N}{\alpha_{Y,0}\overline{V}_Y}\left(M_Y^2 + \sigma_Y^2 + \sigma_Y^2\lambda_2(N,\alpha)\right) + \left(\frac{2\sigma_Y^2}{\varkappa\overline{V}_Y\alpha_{Y,0}\beta_{Y,0}}\right)^2\lambda_{3,Y}(N,\alpha) + \frac{\sigma_Y^2\lambda_{4,Y}(N,\alpha)}{\varkappa\overline{V}_Y\alpha_{Y,0}\beta_{Y,0}}},$$

avec

$$\lambda_1(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(8\alpha^{-1})},\ \lambda_2(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})},$$

$$\lambda_{3,X}(N,\alpha) = \left\{\ln\left(\exp(\varkappa M_X^2\sigma_X^{-2})\frac{8N}{e\alpha}\right)\right\}^2,\ \lambda_{4,X}(N,\alpha) = \left\{\ln\left(\left(1 + \frac{4\sigma_X^2}{\varkappa\overline{V}_X\alpha_{X,0}\beta_{X,0}^2}\right)\exp(\varkappa M_X^2\sigma_X^{-2})\frac{8N}{e\alpha}\right)\right\}^2,$$

$$\lambda_{3,Y}(N,\alpha) = \left\{\ln\left(\exp(\varkappa M_Y^2\sigma_Y^{-2})\frac{8N}{e\alpha}\right)\right\}^2 \text{ et } \lambda_{4,Y}(N,\alpha) = \left\{\ln\left(\left(1 + \frac{4\sigma_Y^2}{\varkappa\overline{V}_Y\alpha_{Y,0}\beta_{Y,0}^2}\right)\exp(\varkappa M_Y^2\sigma_Y^{-2})\frac{8N}{e\alpha}\right)\right\}^2,$$

on a

$$\mathbb{P}(\Delta_N \geq \lambda) \leq \alpha.$$

2.5 Exemples et simulations

2.5.1 Introduction

Nous donnons dans cette partie différents exemples d'utilisation des algorithmes étudiés dans la première partie de ce chapitre.
Un des principaux points clé de leur mise en oeuvre réside dans le choix de la fonction V :

– pour obtenir un algorithme efficace, il faut que le calcul de la trajectoire dans l'espace primal x_i à partir de la trajectoire duale ζ_i se fasse rapidement (autrement dit, on doit savoir expliciter la fonction W'_β et/ou la calculer en un temps non prohibitif),
– nous avons vu dans la partie théorique qui précède qu'une des constantes caractéristique du problème est $\sqrt{\frac{\overline{V}}{\alpha_0}}$, nous essayerons donc de faire en sorte que cette constante soit aussi faible que possible.

Nous allons nous intéresser dans ce qui suit à trois exemples : les jeux matriciels, le problème des moindres carrés en norme L_1 et un problème de localisation.

2.5.2 Jeux matriciels

Nous considérons le problème de point selle suivant :

$$\min_{x\in X}\max_{y\in Y} L(x,y) \text{ avec } L(x,y) = \langle p,y\rangle + \langle y, Ax\rangle + \langle q,x\rangle, \tag{2.15}$$

où $X = \{x \in \mathbb{R}^r | x \geq 0, \sum_{j=1}^r x^{(j)} = 1\}$, $Y = \{y \in \mathbb{R}^s | y \geq 0, \sum_{k=1}^s y^{(k)} = 1\}$, $p \in \mathbb{R}^s$, $q \in \mathbb{R}^s$ et $A \in \mathcal{M}_{s,r}(\mathbb{R})$. Les choix 'naturels' pour les normes sont les normes ℓ_1 sur $\mathbb{R}^r$ et $\mathbb{R}^s$ respectivement. Nous allons chercher une solution de ce problème en utilisant l'algorithme 2.5. Les fonctions V_X et V_Y que nous allons utiliser sont les fonctions entropie

$$V_X(x) = \ln(r) + \sum_{k=1}^r x^{(k)}\ln(x^{(k)}),\ V_Y(y) = \ln(s) + \sum_{k=1}^s y^{(k)}\ln(y^{(k)}).$$

Les différentes constantes du problème sont alors données par $M_X = \| q \|_\infty + \| A \|_\infty$, $M_Y = \| p \|_\infty + \| A \|_\infty$, $\overline{V}_X \leq \ln(r)$ et $\overline{V}_Y \leq \ln(s)$ où $\| A \|_\infty = \max_{1\leq j\leq r, 1\leq k\leq s} |a_{j,k}|$.
Afin de résoudre le problème (2.15), qui est de nature déterministe, nous allons utiliser une stratégie randomisée. Une approche "déterministe" du problème consisterait à "regarder" tous les coefficients de la matrice A afin de résoudre le problème. Nous allons au contraire essayer de trouver une solution approchée, proche de la vraie solution tout en ne regardant qu'une faible partie des coefficients : au cours du i-ème pas de l'algorithme, nous ne regarderons qu'une seule ligne et qu'une seule colonne de la matrice A, le choix de cette ligne et de cette colonne étant déterministe et ne dépendant que de la valeur courante (x_i, y_i) du point de recherche. Plus précisément, considérons les coordonnées des vecteurs x_i et y_i comme des lois de probabilité sur les ensembles $\{1,\cdots,r\}$ et $\{1,\cdots,s\}$ respectivement et choisissons $\hat{j} \in \{1,\cdots,r\}$ et $\hat{k} \in \{1,\cdots,s\}$ selon ces deux lois. Nous définissons alors les versions stochastiques des dérivées partielles de L par

$$\begin{aligned}\Phi(x_i,y_i) &= q^T + A_{\hat{k},\cdot},\\ \Psi(x_i,y_i) &= p^T + A_{\cdot,\hat{j}},\end{aligned}$$

où $A_{k,\cdot}$ et $A_{\cdot,j}$ désignent respectivement la k-ième ligne et la j-ème colonne de la matrice A. Nous voulons maintenant vérifier que l'hypothèse 2.10 est satisfaite et trouver une borne pour σ_X et σ_Y. Remarquons que

$$\begin{aligned}\mathbb{E}[\Phi(x_i,y_i)|\mathcal{F}_{i-1}] &= \mathbb{E}[q^T + A_{\hat{k},\cdot}|\mathcal{F}_{i-1}],\\ &= q^T + \sum_{k=1}^s y_i^{(k)} A_{k,\cdot},\\ &= q^T + y_i^T A = \frac{\partial L}{\partial x}(x_i,y_i),\end{aligned}$$

et de manière analogue

$$\mathbb{E}[\Psi(x_i, y_i)|\mathcal{F}_{i-1}] = \frac{\partial L}{\partial y}(x_i, y_i).$$

De plus, nous avons

$$\begin{aligned}
\mathbb{E}\left[\| \Phi(x_i, y_i) - \frac{\partial L}{\partial x}(x_i, y_i) \|_\infty^2 |\mathcal{F}_{i-1}\right] &= \mathbb{E}\left[\| (q^T + A_{\hat{k},.}) - (q^T + y_i^T A) \|_\infty^2 |\mathcal{F}_{i-1}\right], \\
&= \mathbb{E}\left[\| A_{\hat{k},.} - y_i^T A \|_\infty^2 |\mathcal{F}_{i-1}\right], \\
&\leq \mathbb{E}\left[(2 \| A \|_\infty)^2\right], \\
&\leq 4 \| A \|_\infty^2 .
\end{aligned}$$

On montre de manière analogue que

$$\mathbb{E}\left[\| \Psi(x_i, y_i) - \frac{\partial L}{\partial y}(x_i, y_i) \|_\infty^2 |\mathcal{F}_{i-1}\right] \leq 4 \| A \|_\infty^2,$$

par conséquent l'hypothèse 2.10 est vérifiée avec $\sigma_X = \sigma_Y = 2 \| A \|_\infty$. On peut aussi montrer que l'hypothèse 2.12 est vérifiée avec $\varkappa = 1$.
Il nous reste à préciser les valeurs des paramètres qui règlent la taille des pas de la phase de descente de l'algorithme. Conformément à la proposition 2.9, nous choisirons

$$\gamma_i = 1,\ \beta_{X,i} = \frac{(M_X + \sigma_X)\sqrt{N}}{\sqrt{2\alpha_{X,0}\overline{V}_X}} = \frac{(\| q \|_\infty + 3 \| A \|_\infty)\sqrt{N}}{\sqrt{2\ln(r)}},\ \beta_{Y,i}\frac{(\| p \|_\infty + 3 \| A \|_\infty)\sqrt{N}}{\sqrt{2\ln(s)}}.$$

Dans le cas particulier qui nous intéresse, l'algorithme 2.5 s'écrit

Algorithme 2.9

- *Initialisation* $x_0 = (\frac{1}{r}, \cdots, \frac{1}{r})^T$, $y_0 = (\frac{1}{s}, \cdots, \frac{1}{s})^T$ *et* $(\zeta_0, \xi_0) = 0 \in \mathbb{R}^r \times \mathbb{R}^s$.
- *Pour* $i = 1, \cdots, N$ *choisir* $\hat{j} \in \{1, \cdots, r\}$ *et* $\hat{k} \in \{1, \cdots, s\}$ *suivant les lois données respectivement par* x_{i-1} *et* y_{i-1} *et faire la mise à jour*

$$\begin{aligned}
\zeta_i &= \zeta_{i-1} - (q^T + A_{\hat{k},.}), \\
\xi_i &= \xi_{i-1} + (p^T + A_{.,\hat{j}}), \\
x_i^{(j)} &= \exp(\frac{\zeta_i^{(j)}}{\beta_{X,i}}) \left(\sum_{\ell=1}^{r} \exp(\frac{\zeta_i^{(\ell)}}{\beta_{X,i}})\right)^{-1} \quad \textit{pour } j = 1, \cdots, r, \\
y_i^{(k)} &= \exp(\frac{\xi_i^{(k)}}{\beta_{Y,i}}) \left(\sum_{m=1}^{s} \exp(\frac{\xi_i^{(m)}}{\beta_{Y,i}})\right)^{-1} \quad \textit{pour } k = 1, \cdots, s.
\end{aligned}$$

- *Sortie*

$$x^N = \frac{1}{N}\sum_{i=1}^{N} x_{i-1},\ y^N = \frac{1}{N}\sum_{i=1}^{N} y_{i-1}.$$

Notons comme précédemment $\Delta_N = \max_{y\in Y} L(x^N, y) - \min_{x\in X} L(x, y^N)$. En appliquant les propositions 2.9 et 2.11 à notre problème, nous obtenons les résultats suivants pour Δ_N (qui, rappelons-le, est toujours positif) :

Théorème 2.2 *Bornes pour le problème de jeu matriciel 2.15 :*

$$0 \leq \mathbb{E}[\Delta_N] \leq \sqrt{\frac{2\ln(r)}{N}}(\| q \|_\infty + 5 \| A \|_\infty) + \sqrt{\frac{2\ln(s)}{N}}(\| p \|_\infty + 5 \| A \|_\infty).$$

Pour tout $\alpha > 4\exp(-\frac{N}{4})$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{(3 \| q \|_\infty + 7 \| A \|_\infty)\sqrt{\ln(r)} + 3 \| p \|_\infty + 7 \| A \|_\infty)\sqrt{\ln(s)}}{\sqrt{2N}} + \frac{16 \| A \|_\infty \sqrt{\ln(4\alpha^{-1})}}{\sqrt{N}} + \sqrt{\frac{14}{5}}\frac{6 \| A \|_\infty}{N}\sqrt{\ln(4\alpha^{-1})}(\sqrt{\ln(r)} + \sqrt{\ln(s)}),$$

on a

$$\mathbb{P}(\Delta_N > \lambda) \leq \alpha.$$

Remarque : Il est intéressant de noter que la vitesse de convergence et la précision de l'algorithme de descente en miroir ne dépendent dans cet exemple que très faiblement des dimensions r et s du problème étudié, en effet ils n'apparaissent que sous la forme $\sqrt{\ln(r)}$ et $\sqrt{\ln(s)}$. Cette faible dépendance et le fait que l'oracle utilisé soit du premier ordre justifient l'emploi de cet algorithme pour les problèmes de grandes dimensions.

Dans l'exemples qui suit, nous avons $\| p \|_\infty = \| q \|_\infty = 0$, $\| A \|_\infty = 1$ et $r = s = n$ si bien que les bornes données dans le théorème 2.2 deviennent

$$\mathbb{E}[\Delta_N] \leq 10\sqrt{\frac{2\ln(n)}{N}} \text{ et } \mathbb{P}(\Delta_N > \lambda) \leq \alpha,$$

pour tout $\alpha > 4\exp(-\frac{N}{4})$ et tout

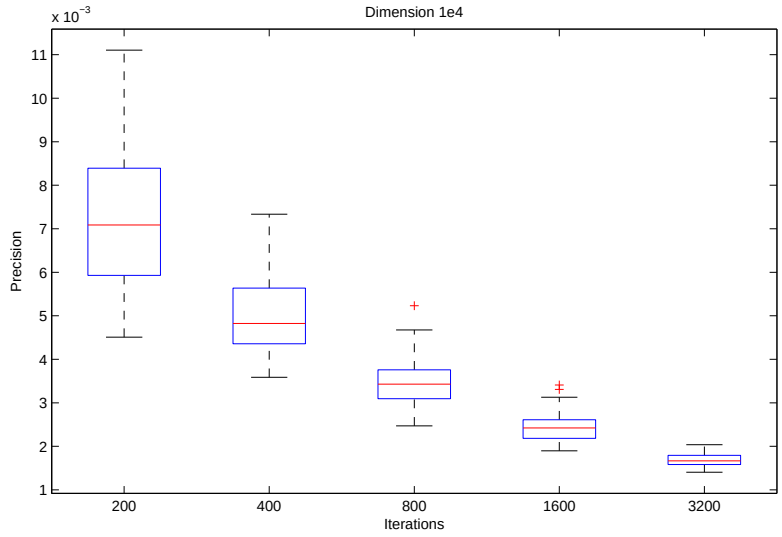
$$\lambda \geq 14\sqrt{\frac{\ln(n)}{2N}} + 16\sqrt{\frac{\ln(4\alpha^{-1})}{N}} + 12\sqrt{\frac{14\ \ln(n)\ \ln(4\alpha^{-1})}{5}}.$$

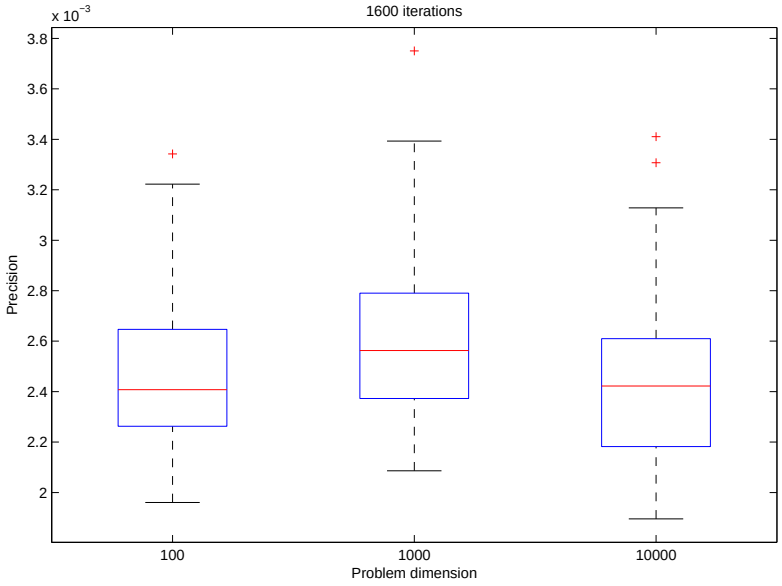
La matrice A considérée est donnée par

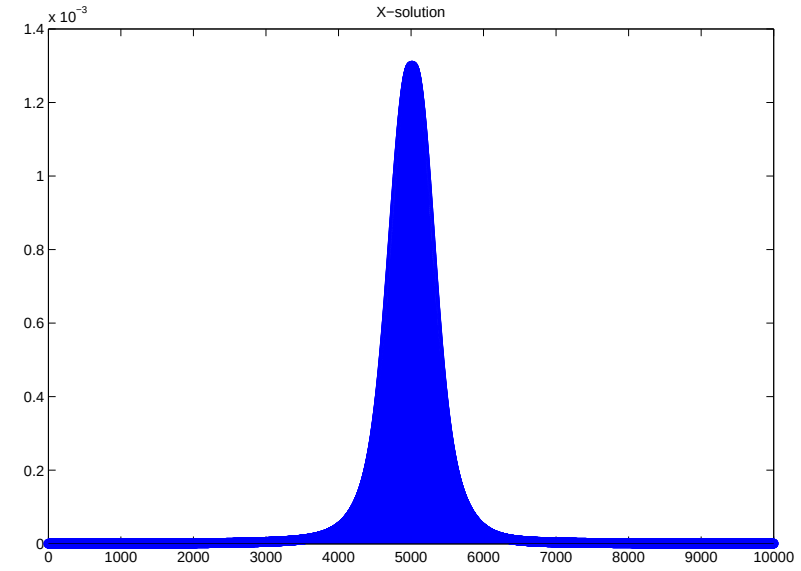
$$A = \frac{1}{n}\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 1 & \ddots & & n-1 \\ \vdots & \ddots & \ddots & \ddots & \cdots \\ n-1 & & \ddots & 1 & 2 \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix}.$$

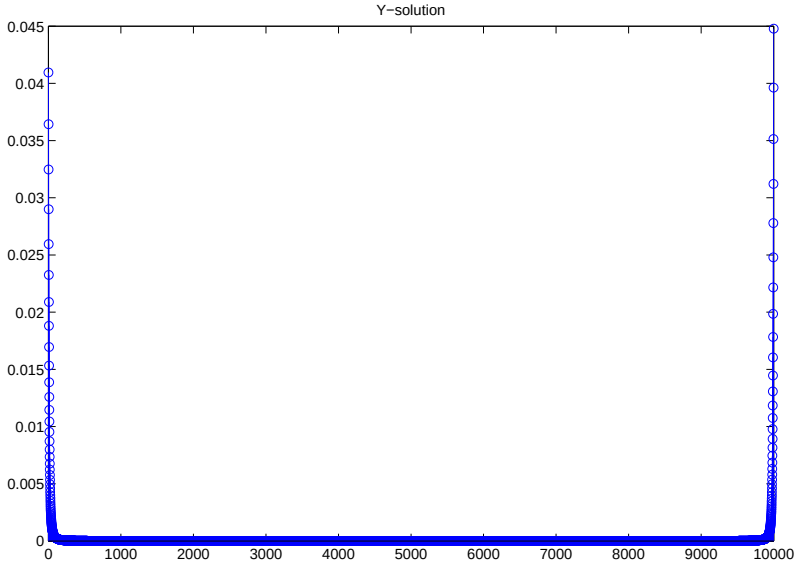
Les figures qui suivent illustrent les résultats obtenus numériquement. La première donne la précision Δ_N obtenue en dimension 10000 en fonction du nombre d'itérations effectuées. La seconde donne la précision obtenue pour 1600 itérations en fonction de la dimension du problème considéré. Ces deux figures ont été réalisées en effectuant 100 simulations pour chaque paire (dimension du problème, nombre d'itérations considéré). Enfin les deux dernières figures représentent les coordonnées (x, y) d'un vecteur solution obtenu après 1600 itérations en dimension 10000.

Ces simulations confortent les résultats théoriques que nous avions obtenus : la précision obtenue est bonne et les résultats dépendent peu de la dimension du problème considéré.


Dimension 1e4
x 10^-3
Precision
Iterations
200
400
800
1600
3200


1600 iterations
x 10^-3
Precision
Problem dimension
100
1000
10000


X−solution
x 10^−3
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000


Y−solution
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

2.5.3 Problème de localisation minmax

Introduction

Nous allons nous intéresser au problème suivant : étant donné un certain nombre d'usagers répartis sur un territoire donné et une ressource qu'ils doivent partager, où faut-il placer cette ressource ? Le problème est de première importance quand la ressource en question est un moyen d'intervention d'urgence tel qu'une caserne de pompiers, un hôpital, ... Dans ce cas, il faudra garantir la meilleure qualité de service possible en pire cas : c'est-à-dire que nous chercherons à minimiser la distance maximale entre la ressource et chacun des usagers. C'est cet objectif en terme de service qui a donné ce problème le nom de problème de localisation minmax. Bien entendu, dans les problèmes réels que nous avons donnés en exemple, la position des différents usagers n'est pas fixe : le problème qui nous intéresse est donc stochastique par nature.
Il existe diverses versions de ce problème. Nous nous intéresseront à deux d'entre elles, la première est empruntée à [21] et la seconde est inspirée de [6].

Premier modèle

On se place dans un domaine convexe X de $\mathbb{R}^2$ où se trouvent m individus qui doivent partager une même ressource. La loi qui régit les déplacements de l'individu j dans le domaine X sera notée F_j. Si la ressource est placée au point $x \in X$, les "difficultés d'accès" à la ressource pour l'individu j sont données par la distance moyenne de x cet individu, c'est-à-dire par la fonction

$$\phi_j(x) = \int \| x - p \|_2 \, dF_j(p).$$

Notre problème est de trouver un emplacement pour la ressource qui minimise la distance au plus éloigné des usagers.
Mathématiquement parlant, il faut résoudre le problème stochastique minmax suivant :

$$\min_{x \in X} \varphi(x), \ \varphi(x) = \max_{j=1 \cdots m} \phi_j(x).$$

On suppose que la seule source d'information dont on dispose pour résoudre le problème est un échantillon

$$w_1 = (p_1^1, \cdots, p_m^1), \cdots, (p_1^N, \cdots, p_m^N)$$

où les variables aléatoires $(p_j^i)_{j,i}$ sont indépendantes et p_j^i suit la loi F_j.
Le problème minmax précédent peut alors être considérer comme un problème de point selle stochastique ayant pour fonction objectif

$$\Phi(x, y) = \sum_{j=1}^{m} y^{(j)} \phi_j(x) : X \times Y \to \mathbb{R},$$

où $Y = \{y \in \mathbb{R}^m | y \geq 0, \sum_{j=1}^m y^{(j)} = 1\}$.
Les versions stochastiques Ψ_x et Ψ_y des dérivées partielles de Φ qui interviennent dans nos algorithmes de point-selle sont données ici par

$$\begin{aligned} \Psi_Y(x, y) &= (\| x - p_1 \|_2, \cdots, \| x - p_m \|_2), \\ \Psi_X(x, y) &= \sum_{j=1}^{m} y^{(j)} \frac{x - p_j}{\| x - p_j \|_2}, \end{aligned}$$

On peut vérifier que

$$\begin{aligned}
\mathbb{E}[\Psi_X(x,y)] &= \sum_{j=1}^{m} y^{(j)} \partial\phi_j(x) = \frac{\partial\Phi}{\partial x}(x,y), \\
\mathbb{E}[\Psi_Y(x,y)] &= (\phi_1(x), \cdots, \phi_m(x)) = \frac{\partial\Phi}{\partial y}(x,y).
\end{aligned}$$

Précisons maintenant les normes et les fonctions proximales associées que nous allons utiliser. Au vu du problème considéré, le choix le plus naturel pour la norme sur X est la norme $\| \cdot \|_2$ et celui pour la fonction proximale associée est $V_X(x) = \frac{1}{2} \| x - x_0 \|_2$ (où x_0 sera le point de recherche initial de l'algorithme) qui est fortement convexe de paramètre $\alpha_{X,0} = 1$. On notera dans la suite Π_X la projection orthogonale sur X. Y étant un simplexe, nous choisirons $\| \, . \, \|_Y = \| \, . \, \|_1$ et utiliserons la fonction proximale suivante : $V_Y(y) = \ln(m) + \sum_{j=1}^{m} y^{(j)} \ln(y^{(j)})$. L'algorithme 2.5 devient alors

Algorithme 2.10

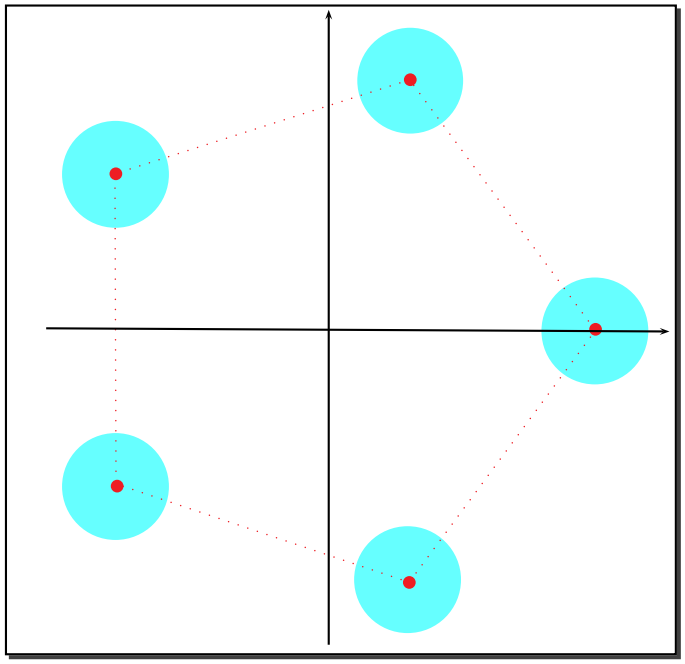
- *Initialisation* $x_0 \in X$, $y_0 = (1/m, \cdots, 1/m) \in Y$ *et* $(\zeta_0, \xi_0) = 0 \in \mathbb{R}^2 \times \mathbb{R}^m$.
- *Pour* $i = 1, \cdots, N$ *faire*

$$\begin{aligned}
\zeta_i &= \zeta_{i-1} - \gamma_i \sum_{j=1}^{m} y_{i-1}^{(j)} \frac{x_{i-1} - p_j^i}{\| x_{i-1} - p_j^i \|_2}, \\
\xi_i &= \xi_{i-1} + \gamma_i (\| x_{i-1} - p_1^i \|_2, \cdots, \| x_{i-1} - p_m^i \|_2), \\
x_i &= \Pi_X(x_0 + \frac{\zeta_i}{\beta_{X,i}}), \\
y_i^{(j)} &= e^{\xi_i^{(j)}/\beta_{Y,i}} \left(\sum_{k=1}^{m} e^{\xi_i^{(k)}/\beta_{Y,i}} \right)^{-1}, \; j = 1, \cdots, m.
\end{aligned}$$

- *Sortie*

$$x^N = \frac{\sum_{i=1}^{N} \gamma_i \, x_{i-1}}{\sum_{i=1}^{N} \gamma_i}, \; y^N = \frac{\sum_{i=1}^{N} \gamma_i \, y_{i-1}}{\sum_{i=1}^{N} \gamma_i}.$$

Il nous faut maintenant préciser les paramètres du problème. Nous allons nous placer dans un cas simple. Un exemple plus réaliste sera donné lors de l'étude du second modèle. Nous supposerons que X est le pavé $[-10, 10] \times [-10, 10] \in \mathbb{R}^2$. Les déplacements des cinq individus considérés ont lieu de manière uniforme dans des disques (de rayon r) centrés aux sommets d'un pentagone régulier. Ce pentagone régulier a pour centre de symétrie le point $(0, 0)$ et un de ses sommets au point $(5, 0)$. Le dessin ci-dessous shématise le problème dans le cas $r = 1$.



On vérifie aisément $M_X \leq 1$, $M_Y \leq 20$, $\overline{V}_X \leq 400$, $\overline{V}_Y \leq \ln(5)$. Il nous faut ensuite évaluer σ_X^2 et σ_Y^2. Pour tout $y \in \Delta$ et tout $x \in X$, on a

$$\begin{aligned}
\mathbb{E}[\| \eta_X(x,y) \|^2] &= \int \left\| \sum_{j=1}^m y^{(j)} \left(\frac{x-p_j}{\| x-p_j \|_2} - \int \frac{x-q_j}{\| x-q_j \|_2} dF_j(q_j) \right) \right\|_2^2 \prod_{j=1}^m dF_j(p_j), \\
&\leq \int \left(\sum_{j=1}^m y^{(j)} \int \left\| \left(\frac{x-p_j}{\| x-p_j \|_2} - \frac{x-q_j}{\| x-q_j \|_2} \right) dF_j(q_j) \right\|_2 dF_j(q_j) \right)^2 \prod_{j=1}^m dF_j(p_j), \\
&\leq 4,
\end{aligned}$$

si bien que $\sigma_X^2 \leq 4$. Majorons maintenant l'espérance de $\| \eta_Y(x,y) \|_\infty$ pour tout x, y. On a, pour tout $1 \leq j \leq m$ et tout p_j dans le support de la mesure dF_j,

$$\begin{aligned}
\left| \| x-p_j \|_2 - \int \| x-q_j \|_2 \, dF_j(q_j) \right| &\leq \int |\| x-p_j \|_2 - \| x-q_j \|_2| \, dF_j(q_j), \\
&\leq \int \| q_j - p_j \|_2 \, dF_j(q_j), \\
&\leq 2r.
\end{aligned}$$

D'où $\sigma_Y^2 \leq 4r^2$.
Pour résoudre le problème de localisation minmax dans le cas où le modèle retenu est celui que nous venons de décrire, nous allons mettre en œuvre l'algorithme 2.5 avec un nombre N de pas fixé à l'avance. par conséquent nous ferons les choix de paramètres suivants pour les suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{y,i})_i$

$$\forall i_\gamma_i = 1, \; \beta_{X,i} = \frac{(M_X + \sigma_X)\sqrt{N}}{2\alpha_{X,0}\overline{V}_X} = \frac{3\sqrt{N}}{20\sqrt{2}}, \; \beta_{Y,i} = \frac{(M_Y + \sigma_Y)\sqrt{N}}{2\alpha_{Y,0}\overline{V}_Y} = \frac{4(5+r)\sqrt{N}}{\sqrt{2\ln(5)}}.$$

Remarquons que l'hypothèse 2.12 est vérifiée avec $\varkappa = 1$. Nous avons donc

Théorème 2.3 *Bornes pour le premier modèle du problème de localisation minmax :*

$$\mathbb{E}[\Delta_N] \leq \frac{\sqrt{2}100 + \sqrt{2\ln(m)}(5+r)}{\sqrt{N}}.$$

Pour tout $\alpha > \exp\left(-\frac{N}{4}\right)$, *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \frac{2\sqrt{2}(35+(15+r)\sqrt{\ln(m)}) + 8(10+r)\sqrt{\ln(4\alpha^{-1})}}{\sqrt{N}} + 6\sqrt{14}\left(4\sqrt{5} + \sqrt{\frac{\ln(m)}{5}}\right)\frac{\sqrt{\ln(4\alpha^{-1})}}{N},$$

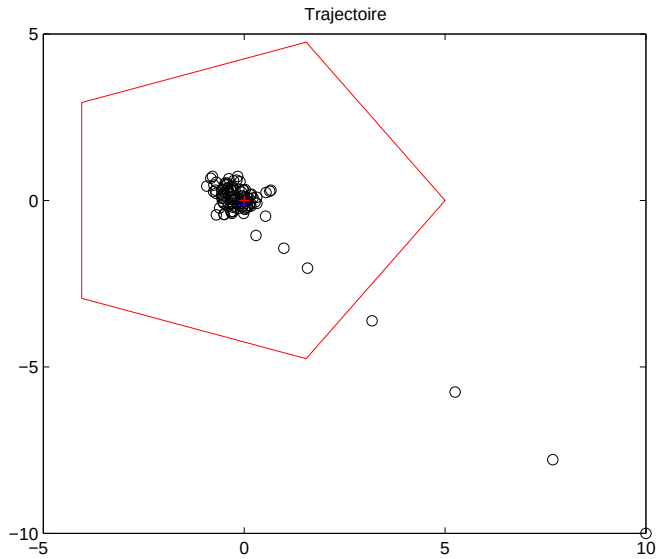
on a

$$\mathbb{P}(\Delta_N > \lambda) \leq \alpha.$$

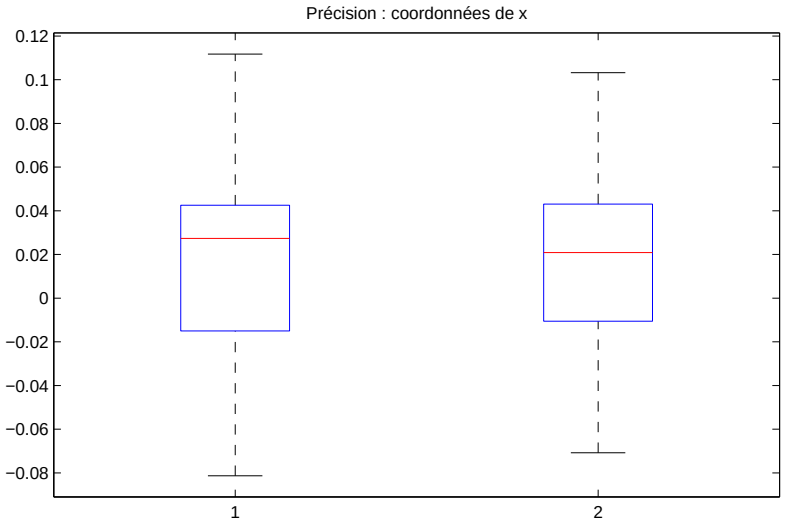
Notons que les bornes obtenues dans le théorème 2.3 ne dépendent que faiblement du nombre m de villes puisque celui-ci n'apparaît que sous la forme $\sqrt{\ln(s)}$. L'algorithme 2.10 est donc particulièrement bien adapté au cas où il y a un grand nombre de villes.

Simulations

Nous présentons maintenant des résultats de simulations de ce premier modèle. Les trajectoires ont toutes été initialisées au point $x_0 = (10, -10)$. La première figure donne une trajectoire type $(x_i)_i$ lorsque le nombre d'itérations est égal à 100. la seconde figure donne les coordonnées du point solution x obtenu sur 100 réalisations avec un nombre d'itérations égal à 200.



Nous pouvons observer une convergence rapide de la trajectoire vers la solution du problème et la précision obtenue est bonne.


Précision : coordonnées de x
0.12
0.1
0.08
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
1
2

Deuxième modèle : une version plus réaliste

Le modèle que nous avons étudié était très simpliste et visait seulement à illustrer la convergence de l'algorithme dans le cadre de la recherche de point selle. Quelques modifications simples nous paraissent souhaitables si l'algorithme doit être utilisé avec des données réelles. Nous en proposons trois :

- donner des poids différents aux individus,
- ne recevoir que la position d'un seul individu à chaque pas de l'algorithme, la fréquence pour chaque individu étant proportionnelle au poids qui lui a été attribué ,
- permettre que les lois qui régissent les déplacement des individus ne soient pas toutes identiques.

Voici quelques raisons d'être pour ces modifications. La première peut être le résultat de regroupements d'individus ayant des propriétés similaires (afin de réduire la taille du problème par exemple) ou de l'importance accordée à chacun des individus (s'il s'agit d'un entrepot au service de deux boutiques ayant des chiffres de vente très différents, il peut être intéressant de se placer plus près de celle qu'il faut réapprovisionner le plus souvent). Dans la mesure où seuls les poids relatifs des différents individus ont de l'importance, nous supposerons dans la suite que la somme totale de tous les poids est égale à 1, ainsi les différents poids pourront être considérés comme une loi de probabilité sur les différents individus.
Pour ce qui est de la seconde, elle permet de se placer dans un contexte plus réaliste : s'il s'agit par exemple de trouver un emplacement pour une station d'ambulance, les demandes arrivent une par une et sont en général plus fréquentes en provenance des zones les plus peuplées. Notons que si les poids accordés aux différents individus sont proportionnels aux fréquences des demandes qui en émanent, cette hypothèse nous permettra de ne pas supposer que les poids sont connus. Nous nous placerons d'ailleurs dans la suite dans le cas où l'individu d'où provient la demande est choisi selon la loi de probabilité donnée par les poids. Enfin cette hypothèse permet un traitement on line des données.
Enfin la troisième modification permet de supposer que les individus ont des comportements différents, certains pouvant être plus mobiles que d'autres.

Le nouveau modèle est le suivant : dans une partie convexe X de $\mathbb{R}^2$ se trouvent m individus qui se déplacent de manière indépendante les uns des autres. On note F_j la loi qui régit le déplacement du j-ème individu. Comme dans le modèle précédent, les m individus partagent une ressource commune. La distance moyenne qui sépare l'individu j de la ressource si celle-ci est située en $x \in X$ est donnée par la fonction

$$\phi_j(x) = \int \| x - p \|_2 \, dF_i(p).$$

On note λ_j la proportion du nombre total d'accès à la ressource effectué par j ($0 < \lambda_j \leq 1$, $\sum_{j=1}^m \lambda_j = 1$). Les coefficients $(\lambda_j)_j$ ne sont pas supposés connus.
On cherche à savoir où placer la ressource pour minimiser le pire cas en tenant compte des pondérations, c'est-à-dire à résoudre le problème suivant :

$$\min_{x \in X} \varphi(x), \ \varphi(x) = \max_{j=1 \cdots m} \lambda_j \phi_j(x).$$

La seule source d'information dont on dispose est un échantillon $w_k = (i_k, p_k)$ où $i_k \in \{1 \cdots m\}$ avec $\mathbb{P}(i_k = j) = \lambda_j$ et la loi conditionnelle de p_k sachant i_k est F_{i_k}.

La fonction objectif de notre problème de point selle devient alors

$$\Phi(x,y) = \sum_{j=1}^{m} y^{(j)} \lambda_j \phi_j(x) : X \times Y \to \mathbb{R},$$

où $Y = \{y \in \mathbb{R}^n | y \geq 0, \sum_{j=1}^{m} y^{(j)} = 1\}$.
Et les versions stochastiques Ψ_X et Ψ_Y des dérivées de Φ sont maintenant données par

$$\begin{aligned} \Psi_Y(x,y) &= \| x - p_k \|_2 \, (\delta_{1,i_k}, \cdots, \delta_{n,i_k}), \\ \Psi_X(x,y) &= y^{(i_k)} \frac{x - p_k}{\| x - p_k \|_2}, \end{aligned}$$

où $\delta_{i,j}$ vaut 1 si $i = j$ et 0 sinon. On a bien

$$\begin{aligned} \mathbb{E}[\Psi_X(x,y)] &= \sum_{j=1}^{m} y^{(j)} \lambda_j \partial \phi_j(x) = \frac{\partial \Phi}{\partial x}(x,y), \\ \mathbb{E}[\Psi_Y(x,y)] &= (\lambda_1 \phi_1(x), \cdots, \lambda_n \phi_n(x)) = \frac{\partial \Phi}{\partial y}(x,y). \end{aligned}$$

Les normes et les fonctions proximales associées que nous allons utiliser sont les mêmes que dans le modèle précédent.
La norme sur X est la norme $\| \cdot \|_2$ et la fonction proximale associée est $V_X(x) = \frac{1}{2} \| x - x_0 \|_2$ (où x_0 sera le point de recherche initial de l'algorithme) . On notera dans la suite Π_X la projection orthogonale sur X. Sur Y nous considérerons $\| \, . \, \|_Y = \| \, . \, \|_1$ et utiliserons la fonction proximale : $V_Y(y) = \ln(m) + \sum_{j=1}^{m} y^{(j)} \ln(y^{(j)})$. L'algorithme utilisé pour résoudre le problème est alors

Algorithme 2.11

- *Initialisation* $x_0 \in X$, $y_0 = (1/m, \cdots, 1/m) \in Y$, $(\zeta_0, \xi_0) = 0 \in \mathbb{R}^2 \times \mathbb{R}^m$, $\beta_{X,0} > 0$ *et* $\beta_{X,0} > 0$.
- *Pour* $i = 1, \cdots, N$ *faire*

$$\begin{aligned}
\zeta_i &= \zeta_{i-1} - \sum_{j=1}^{m} y_{i-1}^{(j)} \frac{x_{i-1} - p_j^i}{\| x_{i-1} - p_j^i \|_2}, \\
\xi_i &= \xi_{i-1} + (\| x_{i-1} - p_1^i \|_2, \cdots, \| x_{i-1} - p_m^i \|_2), \\
\beta_{X,i} &= \beta_{X,i-1} + \frac{1}{\overline{V}_X \beta_{X,i-1}} \| \sum_{j=1}^{m} y_{i-1}^{(j)} \frac{x_{i-1} - p_j^i}{\| x_{i-1} - p_j^i \|_2} \|_2^2, \\
\beta_{Y,i} &= \beta_{Y,i-1} + \frac{1}{\overline{V}_Y \beta_{Y,i-1}} \| (\| x_{i-1} - p_1^i \|_2, \cdots, \| x_{i-1} - p_m^i \|_2) \|_\infty^2, \\
x_i &= \Pi_X(x_0 + \frac{\zeta_i}{\beta_{X,i}}), \\
y_i^{(j)} &= e^{\xi_i^{(j)}/\beta_{Y,i}} \left(\sum_{k=1}^{m} e^{\xi_i^{(k)}/\beta_{Y,i}} \right)^{-1}, \ j = 1, \cdots, m.
\end{aligned}$$

- *Sortie*

$$x^N = \frac{1}{N} \sum_{i=1}^{N} x_{i-1}, \ y^N = \frac{1}{N} \sum_{i=1}^{N} y_{i-1}.$$

Donnons maintenant les bornes théoriques pour la vitesse de convergence de cet algorithme

Théorème 2.4 *Bornes pour le second modèle du problème de localisation minmax*

$$\begin{aligned}
\mathbb{E}[\Delta_N] \leq\ & \frac{4}{\sqrt{2N}}(80 + (3r + 20)\sqrt{\ln(m)}) + \frac{800\beta_{X,0} + 2\ln(m)\beta_{Y,0}}{N} \\
& + \frac{\ln(N)^2}{\sqrt{2}N^{\frac{3}{2}}} \left[\frac{160}{3} \left(1 + \frac{1}{25\beta_{X,0}} \right) + \frac{4r^2\sqrt{\ln(m)}}{10 + r} \left(1 + \frac{16r^2}{\beta_{Y,0}\ln(m)} \right) \right] + \frac{C}{N^{\frac{3}{2}}},
\end{aligned}$$

où C *est une constante.*
pour tout $\alpha \geq 8\exp\left(-\frac{N}{4}\right)$, *tout* $\beta_{X,0} > 0$, *tout* $\beta_{Y,0} > 0$ *et tout* $\lambda > \lambda^*(N, \alpha)$ *où*

$$\begin{aligned}
\lambda^*(N, \alpha) =\ & \frac{40\sqrt{2} + 2r\sqrt{2\ln(m)}}{\sqrt{N}} + \frac{4(10 + r)}{\sqrt{N}}\lambda_1(N, \alpha) + \frac{40 + 2r\sqrt{\ln(m)}}{\sqrt{N}}\lambda_2(N, \alpha) \\
& + \frac{800}{N}\sqrt{\beta_{X,0}^2 + \frac{N}{100}(5 + 4\lambda_2(N, \alpha)) + \left(\frac{1}{50\beta_{X,0}}\right)^2 \lambda_{3,X}(N, \alpha) + \frac{\lambda_{4,X}(N, \alpha)}{100\beta_{X,0}}} \\
& \frac{2\ln(m)}{N}\sqrt{\beta_{Y,0}^2 + \frac{16N}{\ln(m)}(100 + r^2(1 + \lambda_2(N, \alpha))) + \left(\frac{8r^2}{\ln(m)\beta_{Y,0}}\right)^2 \lambda_{3,Y}(N, \alpha) + \frac{4r^2\lambda_{4,Y}(N, \alpha)}{\ln(m)\beta_{Y,0}}},
\end{aligned}$$

avec

$$\lambda_1(N,\alpha) = \sqrt{4\ln(8\alpha^{-1})},\ \lambda_2(N,\alpha) = \sqrt{\frac{28}{5N}\ln(8\alpha^{-1})},$$

$$\lambda_{3,X}(N,\alpha) = \left(\frac{1}{4} + \ln\left(\frac{8N}{e\alpha}\right)\right)^2,\ \lambda_{3,Y}(N,\alpha) = \left(\frac{100}{r^2} + \ln\left(\frac{8N}{e\alpha}\right)\right)^2,$$

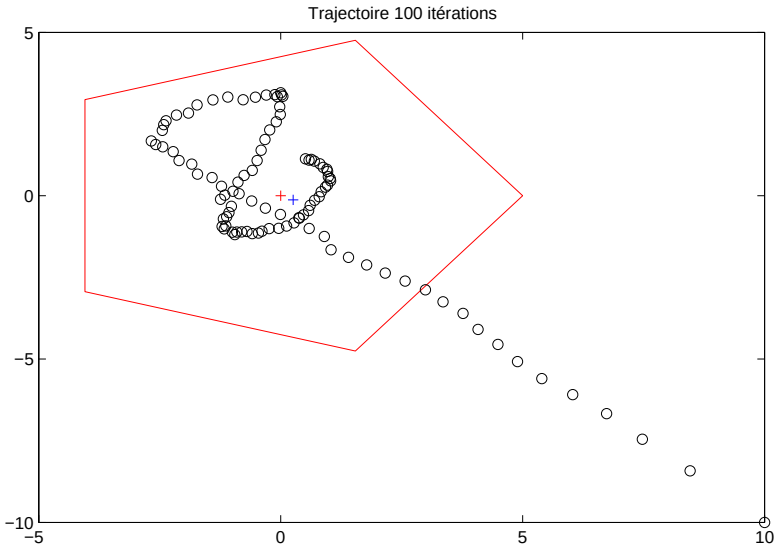
$$\lambda_{4,X} = \left\{\ln\left[\left(1 + \frac{1}{100\beta_{X,0}^2}\right)\exp\left(\frac{1}{4}\right)\frac{8N}{e\alpha}\right]\right\}^2 \ \textit{et}\ \lambda_{4,Y} = \left\{\ln\left[\left(1 + \frac{16r^2}{\ln(m)\beta_{Y,0}^2}\right)\exp\left(\frac{100}{r^2}\right)\frac{8N}{e\alpha}\right]\right\}^2,$$

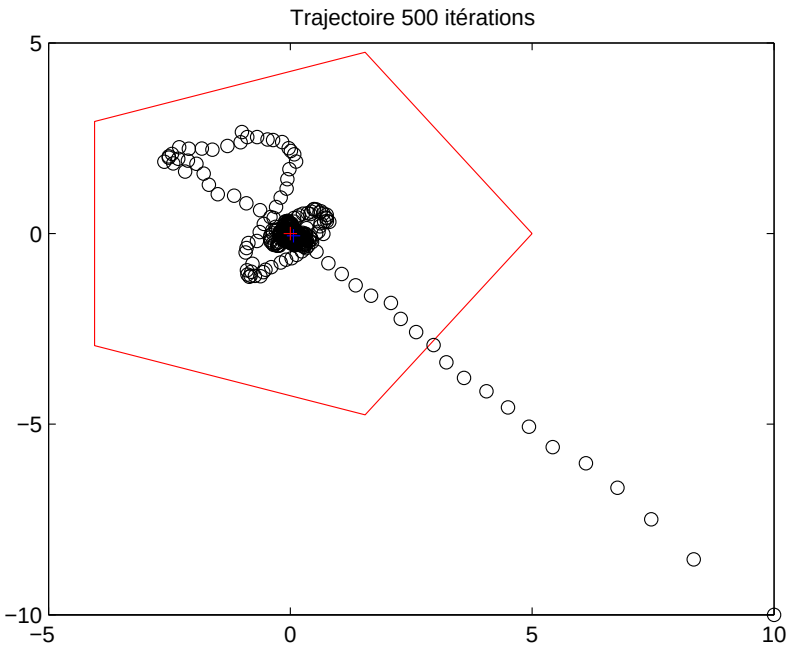
on a

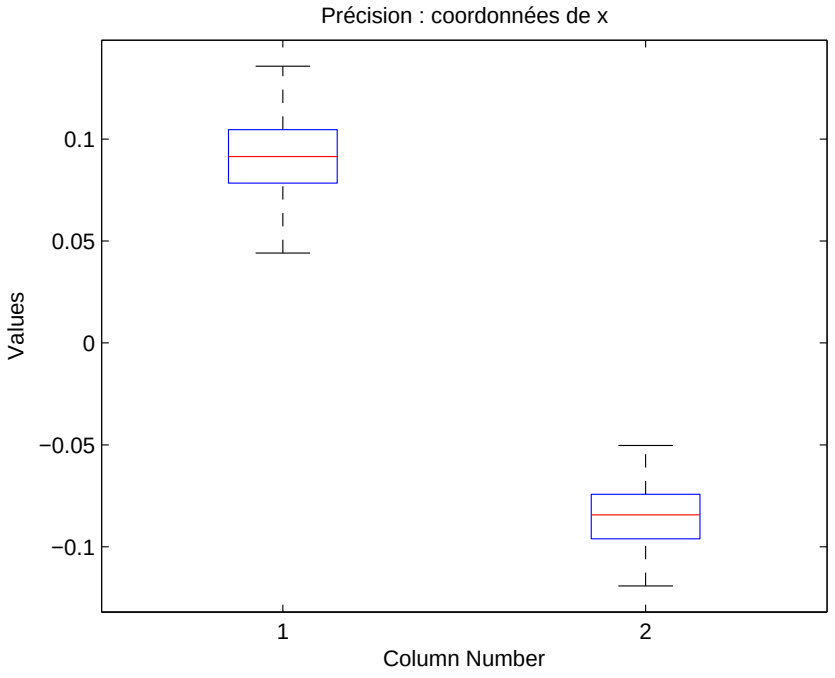
$$\mathbb{P}(\Delta_N \geq \lambda) \leq \alpha.$$

Simulations

Donnons maintenant quelques résultats de simulation pour ce problème. Les deux premières figures représentent deux exemples de trajectoires, la première est obtenue après 100 itérations et la seconde après 500 itérations (dans la mesure où la seconde version de l'algorithme ne demande que la position d'un seul des cinq sommets du pentagone à chaque itération, c'est la seconde trajectoire qu'il faut comparer avec celle obtenue pour le premier modèle). La croix bleue marque l'emplacement de la solution calculée par l'algorithme, la croix rouge la solution théorique.
Les résultats obtenus sont conformes à la théorie. Nous pouvons constater sur la troisième figure que la première coordonnée du point x solution est systématiquement positive et la seconde systématiquement négative : cet effet est lié au choix de x_0, très éloigné de la solution x^* et qui "marque" la trajectoire. ce phénomène tend à disparaître lorsque le nombre d'itérations augmente ou lorsque le point initial x_0 est choisi plus proche du pentagone.




Trajectoire 500 itérations
5
0
−5
−10
−5
0
5
10


Précision : coordonnées de x
0.1
0.05
0
−0.05
−0.1
Values
1
2
Column Number

Chapitre 3

Algorithme d'extrapolation dual

3.1 Introduction

L'objet de ce chapitre est d'adapter au cas stochastique l'algorithme d'extrapolation dual introduit par Nesterov dans [24].
Comme dans le chapitre précédent, nous nous intéresserons à trois grandes classes de problèmes : les problèmes de minimisation sans contrainte, les problèmes d'inégalités variationnelles à opérateur monotone et les problèmes de point selle. Nous montrerons ensuite que, comme pour l'algorithme de descente en miroir, il est possible de mettre en place une stratégie de choix du gain qui puisse s'adapter dans une certaine mesure à l'oracle utilisé.

3.2 Algorithme pour les problèmes de minimisation sans contrainte

3.2.1 Algorithme

Nous considérons de nouveau le premier des problèmes qui nous ont intéressés dans le chapitre précédent, à savoir

$$\min_{x\in Q} f(x),$$

où Q est un convexe fermé et f une fonction convexe sur Q. L'ensemble Q est inclus dans un espace vectoriel de dimension finie E muni de la norme $\|\cdot\|$. Nous noterons E^* l'espace dual de E, $\|\cdot\|_*$ la norme duale associée et $\langle\cdot,\cdot\rangle$ le produit scalaire sur $E\times E^*$.
Les outils nécessaires à la mise en oeuvre de ce nouvel algorithme sont sensiblement les mêmes que pour l'algorithme de descente en miroir. Nous supposons connue une fonction fortement convexe d de paramètre $\alpha_0>0$ et differentiable sur Q. Nous noterons V la divergence de Bregman associée à d qui est donnée par

$$\forall x,y\ V(x,y)=d(y)-d(x)-\langle d'(x),y-x\rangle.$$

Comme d est fortement convexe, pour tous $x,\ y\in Q$, $V(x,y)\geq\frac{\alpha_0}{2}\ \|\ x-y\ \|^2$. Choisissons arbitrairement un point $\overline{x}$ comme 'centre'de l'ensemble Q ($\overline{x}$ aura sensiblement le même rôle que celui du point x_0 dans l'algorithme de descente en miroir) et $\overline{V}>0$ tel que $x^*\in\mathcal{F}_{\overline{V}}=\{x\in Q,\ V(\overline{x},x)<\overline{V}\}$.
Nous utiliserons dans la suite les deux fonctions de support suivantes :

$$\begin{aligned}\forall s\in E^*\ \ \xi_{\overline{V}}(s) &= \max_x\left\{\langle s,x-\overline{x}\rangle,\ x\in\mathcal{F}_{\overline{V}}\right\},\\ \forall(z,s)\in E\times E^*,\ \beta>0\ \ W_\beta(z,s) &= \max_{x\in Q}\left\{\langle s,x-z\rangle-\beta V(z,x)\right\}.\end{aligned}$$

Nous supposerons dans la suite que la fonction f est continue sur Q et pour $x\in Q$ nous noterons $f'(x)$ un élément quelconque du sous-gradient $\nabla f(x)$ de f en x. Nous supposerons que l'une des deux hypothèses parmi 2.13 et l'hypothèse suivante est vérifiée.

Hypothèse 3.1 *f' est à variations bornées sur Q. Il existe une constante $M>0$ telle que*

$$\forall x,y\in Q\ \ \|\ f'(x)-f'(y)\ \|_*\leq M.$$

Nous disposons de plus d'un oracle qui nous fournit une version aléatoire (bruitée) Ψ de f' et on note $\eta=\Psi-f'$. Les hypothèses à faire concernant η seront discutées ultérieurement.
Nous allons maintenant étudier le pas de mise à jour de l'algorithme que nous allons considérer. Il met en jeu, comme pour l'algorithme de descente en miroir, deux paramètres strictement positifs β

et γ. Le pas $\mathcal{E}_{\beta,\gamma}$ que nous allons expliciter, transforme un point quelconque de l'espace dual E^* en un nouveau point $s_+ \in E^*$. Plus précisément, notons

$$T_\beta(z,s) = \operatorname*{argmax}_x \{\langle s, x-z\rangle - \beta V(z,x) \,:\, x \in Q\},$$

alors $\mathcal{E}_{\beta,\gamma}(s) = (x, y, s_+)$ avec

$$\begin{aligned} x &= T_\beta(\overline{x}, s), \\ y &= T_\beta(x, -\gamma\Psi(x)), \\ s_+ &= s - \gamma\Psi(y). \end{aligned}$$

La mise à jour du paramètre s nécessite la construction de deux points intermédiaires dans l'espace primal contre un seul pour l'algorithme de descente de descente en miroir. En fait, la différence entre les deux algorithmes réside dans la construction du point y, le pas de l'algorithme de descente en miroir peut en effet se réécrire

$$\begin{aligned} x &= T_\beta(x_0, \zeta), \\ \zeta_+ &= \zeta - \gamma\Psi(x). \end{aligned}$$

Tout se passe donc, dans l'algorithme d'extrapolation duale, comme si, après avoir obtenu le point x à partir des points de recherche y précédents par un pas semblable à celui de la méthode de descente, nous essayions de 'deviner' le point de recherche suivant en réinitialisant l'algorithme de descente en miroir avec $x_0 = x$ et en conservant le premier point $x_1 = y$ de la nouvelle trajectoire construite. Le pas $\mathcal{E}_{\beta,\gamma}(s)$ met alors à jour la valeur de s en utilisant le point y extrapolé à partir du début de la trajectoire, plutôt que le point x qui serait fournit par l'algorithme de descente en miroir.

Décrivons maintenant l'algorithme obtenu. Le choix des suites $(\gamma_i)_i$ et $(\beta_i)_i$ sera discuté ultérieurement. Nous supposerons seulement pour le moment que les deux suites sont à valeurs strictement positives et que la suite $(\beta_i)_i$ est croissante.

Algorithme 3.1

- *Initialisation :* $s_{-1} = 0 \in E^*$.
- *Pour* $k = 0$ *à* N *faire* $(x_k, y_k, s_k) = \mathcal{E}_{\beta_k,\gamma_k}(s_{k-1})$.
- *Sortie :*

$$y^N = \frac{\sum_{i=0}^N \gamma_i y_i}{\sum_{i=0}^N \gamma_i}$$

3.2.2 Bornes en espérance

Nous voulons maintenant étudier la convergence de l'algorithme. Remarquons d'abord que pour tout $\beta > 0$, $s \in E^*$

$$\xi_{\overline{V}}(s) \leq \beta\overline{V} + W_\beta(\overline{x}, s).$$

La démonstration de la convergence del'algorithme repose principalement sur le lemme suivant

Lemme 3.1 *Soient s, $\delta \in E^*$ et $z \in Q$. Alors*

$$W_\beta(z, s+\delta) \leq W_\beta(z,s) + \langle \delta, T_\beta(z,s) - z\rangle + W_\beta(T_\beta(z,s), \delta).$$

Preuve
Notons pour simplifier les expressions $T_\beta = T_\beta(z,s)$. Par définition de W_β et de T_β :

$$\begin{aligned} W_\beta(z, s+\delta) &= \max_{x\in Q} \{\langle s+\delta, x-z\rangle - \beta V(z,x)\}, \\ &= \max_{x\in Q} \{\langle s, T_\beta - z\rangle - \beta V(z,T_\beta) + \langle \delta, T_\beta - z\rangle + \langle s+\delta, x - T_\beta\rangle + \beta V(z,T_\beta) - \beta V(z,x)\}, \\ &= W_\beta(z,s) + \langle \delta, T_\beta - z\rangle + \max_{x\in Q} \{\langle s+\delta, x-T_\beta\rangle + \beta V(z,T_\beta) - \beta V(z,x)\}. \end{aligned}$$

Par définition de V, pour tout $x \in Q$,

$$V(z,T_\beta) - V(z,x) = d(x) - d(T_\beta) - \langle \nabla d(z), T_\beta - x\rangle.$$

De plus, comme T_β est solution du problème de maximisation qui définit $W_\beta(z,s)$, il satisfait à la condition d'optimalité du premier ordre qui s'écrit :

$$\langle s - \beta(\nabla d(T_\beta) - \nabla d(z)), x - T_\beta\rangle \leq 0 \ \forall x \in Q$$

si bien que

$$\begin{aligned} \max_{x\in Q} \{\langle s+\delta, x-T_\beta\rangle + \beta V(z,T_\beta) - \beta V(z,x)\} &\leq \max_{x\in Q} \{\langle \delta, x-T_\beta\rangle - \beta(d(T_\beta) - d(x) - \langle \nabla d(T_\beta), x - T_\beta\rangle)\}, \\ &\leq W_\beta(T_\beta, \delta). \end{aligned}$$

□

Remarque : Comme $W'_{s,\beta}(z,s) = T_\beta(z,s) - z \in E$, le lemme 3.1 peut se réécrire

$$W_\beta(z, s+\delta) \leq W_\beta(z,s) + \langle \delta, W'_{s,\beta}(z,s)\rangle + W_\beta(T_\beta(z,s), \delta).$$

Comme pour l'algorithme de descente en miroir l'étude de la convergence de l'algorithme d'extrapolation dual se ramène à l'étude des accroissements de la fonction $W_\beta(\overline{x}, \cdot)$ le long de la trajectoire $(s_i)_i$. Le lemme suivant donne le principal résultat qui sera utilisé

Lemme 3.2 *Soit $s \in E^*$ et $(x, y, s_+) = \mathcal{E}_{\beta,\gamma}(s)$. Alors*

$$\gamma\langle \Psi(y), y - \overline{x}\rangle + W_\beta(\overline{x}, s_+) \leq W_\beta(\overline{x}, s) + \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x) - \Psi(y) \parallel_*^2 - \frac{\alpha_0\beta}{2} \parallel x - y \parallel^2 .$$

Preuve
Choisissons $z = \overline{x}$ et $\delta = -\gamma\Psi(y)$ dans le lemme 3.1. Il vient

$$W_\beta(\overline{x}, s - \gamma\Psi(y)) \leq W_\beta(\overline{x}, s) + \langle -\gamma\Psi(y), T_\beta(\overline{x}, s) - \overline{x}\rangle + W_\beta(T_\beta(\overline{x}, s), -\gamma\Psi(y)).$$

C'est-à-dire

$$W_\beta(\overline{x}, s_+) \leq W_\beta(\overline{x}, s) + \langle -\gamma\Psi(y), x - \overline{x}\rangle + W_\beta(x, -\gamma\Psi(y)). \tag{3.1}$$

Choisissons ensuite $z = x$, $s = -\gamma\Psi(x)$, $\delta = \gamma(\Psi(x) - \Psi(y))$ dans le lemme 3.1. Alors

$$W_\beta(x, -\gamma\Psi(y)) \leq W_\beta(x, -\gamma\Psi(x)) + \langle \gamma(\Psi(x) - \Psi(y)), T_\beta(x, -\gamma\Psi(x)) - x\rangle + W_\beta(T_\beta(x, -\gamma\Psi(x)), \gamma(\Psi(x) - \Psi(y))).$$

Soit

$$\begin{aligned} W_\beta(x,-\gamma\Psi(y)) &\leq W_\beta(x,-\gamma\Psi(x)) + \langle\gamma(\Psi(x)-\Psi(y)),y-x\rangle + W_\beta(y,\gamma(\Psi(x)-\Psi(y))), \\ &\leq W_\beta(x,-\gamma\Psi(x)) + +\langle\gamma(\Psi(x)-\Psi(y)),y-x\rangle + \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x)-\Psi(y) \parallel_*^2 \end{aligned} \tag{3.2}$$

Comme y a été choisit de telle sorte que $W_\beta(x,-\gamma\Psi(x)) = \langle -\gamma\Psi(x), y-x\rangle - \beta V(x,y)$, et comme on a $V(x,y) \geq \frac{1}{2}\alpha_0 \parallel x-y \parallel^2$. On déduit alors de (3.1) et (3.2) que

$$\begin{aligned} \gamma\langle\Psi(y),y-\overline{x}\rangle + W_\beta(\overline{x},s_+) - W_\beta(\overline{x},s) &\leq \gamma\langle\Psi(y),y-\overline{x}\rangle - \gamma\langle\Psi*(y),x-\overline{x}\rangle + W_\beta(x,-\gamma\Psi(y)), \\ &\leq \gamma\langle\Psi(y),y-x\rangle + W_\beta(x,-\gamma\Psi(x)) + \gamma\langle\Psi(x)-\Psi(y),y-x\rangle \\ &\quad + \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x)-\Psi(y) \parallel_*^2, \\ &\leq \gamma\langle\Psi(x),y-x\rangle - \gamma\langle\Psi(x),y-x\rangle - \beta V(x,y) \\ &\quad + \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x)-\Psi(y) \parallel_*^2, \\ &\leq \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x)-\Psi(y) \parallel_*^2 - \frac{\alpha_0\beta}{2} \parallel x-y \parallel^2 . \end{aligned}$$

□

Nous souhaitons maintenant majorer l'erreur $\varepsilon_N = f(y^N) - f(x^*)$.

$$\begin{aligned} \left(\sum_{i=0}^N \gamma_i\right)\varepsilon_N &\leq \sum_{i=0}^N \gamma_i(f(y_i)-f(x^*)), \\ &\leq \sum_{i=0}^N \gamma_i\langle\nabla f(y_i),y_i-x^*\rangle, \\ &\leq \sum_{i=0}^N \gamma_i\langle f'(y_i)-\Psi(y_i),y_i-x^*\rangle + \sum_{i=0}^N \gamma_i\langle\Psi(y_i),y_i-\overline{x}\rangle + \sum_{i=0}^N \gamma_i\langle\Psi(y_i),\overline{x}-x^*\rangle, \\ &\leq -\sum_{i=0}^N \gamma_i\langle\eta(y_i),y_i-x^*\rangle + \sum_{i=0}^N \gamma_i\langle\Psi(y_i),y_i-\overline{x}\rangle + \xi_{\overline{V}}(s_N), \end{aligned}$$

avec $\xi_{\overline{V}}(s_N) \leq \beta_N\overline{V} + W_{\beta_N}(\overline{x},s_N)$. On note

$$a_k = \sum_{i=0}^k \gamma_i\langle\Psi(y_i),y_i-\overline{x}\rangle + \beta_N\overline{V} + W_{\beta_k}(\overline{x},s_k).$$

Nous déduisons alors du lemme 3.2 et du fait que $(\beta_i)_i$ soit croissante que

$$\begin{aligned} a_k &= a_{k-1} + \gamma_k\langle\Psi(y_k),y_k-\overline{x}\rangle + W_{\beta_k}(\overline{x},s_k) - W_{\beta_{k-1}}(\overline{x},s_{k-1}), \\ &\leq a_{k-1} + \gamma_k\langle\Psi(y_k),y_k-\overline{x}\rangle + W_{\beta_k}(\overline{x},s_k) - W_{\beta_k}(\overline{x},s_{k-1}), \\ &\leq a_{k-1} + \frac{\gamma_k^2}{2\alpha_0\beta_k} \parallel \Psi(x_k)-\Psi(y_k) \parallel_*^2 - \frac{\alpha_0\beta_k}{2} \parallel x_k-y_k \parallel^2 . \end{aligned}$$

Ensuite, puisque $a_{-1} = \beta_N\overline{V}$, nous obtenons

Proposition 3.1

$$\left(\sum_{i=0}^{N}\gamma_i\right)\varepsilon_N \leq \beta_N\overline{V} - \sum_{i=0}^{N}\gamma_i\langle\eta(y_i), y_i - x^*\rangle + \sum_{i=0}^{N}\left(\frac{\gamma_i^2}{2\alpha_0\beta_i}\parallel\Psi(x_i)-\Psi(y_i)\parallel_*^2 - \frac{\alpha_0\beta_i}{2}\parallel x_i - y_i\parallel^2\right). \tag{3.3}$$

Nous allons maintenant utiliser l'équation (3.3) pour obtenir une borne en espérance pour ε_N. Comme notre algorithme est stochastique, il nous aurons besoin de quelques hypothèses portant sur les variables aléatoires $\eta(x_i)$ et $\eta(y_i)$ pour étudier son comportement.
Notons $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$ et pour tout $i \geq 0$, $\mathcal{F}_i$ est la tribu engendrée par $\mathcal{F}_{i-1}$ et les variables aléatoires $\eta(x_i)$ et $\eta(y_i)$. Nous supposerons que l'hypothèse suivante est satisfaite

Hypothèse 3.2 *Il existe* $0 < \sigma^2 < +\infty$ *telle que pour tout* i,

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0 \ p.s., \ \mathbb{E}[\eta(y_i)|\mathcal{F}_i] = 0 \ p.s.,$$

$$\mathbb{E}[\parallel\eta(x_i)\parallel_*^2|\mathcal{F}_i] \leq \sigma^2 \ p.s. \ et \ \mathbb{E}[\parallel\eta(y_i)\parallel_*^2|\mathcal{F}_i] \leq \sigma^2 \ p.s.$$

Commençons par traiter le cas où l'hypothèse 3.1 est vérifiée. Puisque pour tout i

$$\parallel\Psi(x_i)-\Psi(y_i)\parallel_* \leq (M + \parallel\eta(x_i)\parallel_* + \parallel\eta(y_i)\parallel_*), \tag{3.4}$$

nous déduisons de (3.3) que

$$\mathbb{E}[\varepsilon_N] \leq \frac{1}{\sum_{i=0}^{N}\gamma_i}\left(\beta_N\overline{V} + \sum_{i=0}^{N}\frac{\gamma_i^2(2\sigma+M)^2}{2\alpha_0\beta_i}\right). \tag{3.5}$$

Proposition 3.2 *La borne (3.5) est minimale lorsque* $\gamma_i \equiv \gamma$ *et* $\beta_i \equiv \beta$ *avec la relation*

$$\frac{\beta}{\gamma} = \frac{(M+2\sigma)\sqrt{(N+1)}}{\sqrt{2\alpha_0\overline{V}}}.$$

On a alors

$$\mathbb{E}[\varepsilon_N] \leq \sqrt{\frac{2\overline{V}}{\alpha_0}}(M+2\sigma)\frac{1}{\sqrt{N+1}}.$$

Si on ne connaît pas à l'avance le nombre de pas à effectuer dans l'algorithme, on peut choisir $(\gamma_i)_i$ et $(\beta_i)_i$ de façon à ce qu'ils dépendent de i et non plus de N. Nous pouvons par exemple choisir $\gamma_i \equiv 1$ et $\beta_i = (M+2\sigma)(\alpha_0\overline{V})^{\frac{-1}{2}}\sqrt{i+1}$. Il vient alors

$$\mathbb{E}[\varepsilon_N] \leq 2\sqrt{\frac{\overline{V}}{\alpha_0}}(M+2\sigma)\frac{\sqrt{N+2}}{N+1}.$$

Supposons maintenant que l'hypothèse 2.13 est vérifiée. Alors

$$\begin{aligned}\parallel\Psi(x_i)-\Psi(y_i)\parallel_*^2 &\leq 3\left(\parallel\Psi(x_i)-f'(x_i)\parallel_*^2 + \parallel f'(x_i)-f'(y_i)\parallel_*^2 + \parallel f'(y_i)-\Psi(y_i)\parallel_*^2\right),\\ &\leq 3\left(\parallel\eta(x_i)\parallel_*^2 + L^2\parallel x_i - y_i\parallel^2 + \parallel\eta(y_i)\parallel_*^2\right),\end{aligned} \tag{3.6}$$

et (3.3) devient

$$\left(\sum_{i=0}^{N}\gamma_i\right)\varepsilon_N \leq \beta_N\overline{V} - \sum_{i=0}^{N}\gamma_i\langle\eta(y_i), y_i - x^*\rangle + \sum_{i=0}^{N}\left(\frac{3\gamma_i^2L^2}{2\alpha_0\beta_i} - \frac{\alpha_0\beta_i}{2}\right)\parallel x_i - y_i \parallel^2 + \sum_{i=0}^{N}\frac{3\gamma_i^2}{2\alpha_0\beta_i}\left(\parallel \eta(x_i) \parallel_*^2 + \parallel \eta(y_i) \parallel_*^2\right).$$

En choisissant $(\beta_i)_i$ et $(\gamma_i)_i$ telles que $\beta_i\gamma_i^{-1} \geq \sqrt{3}L\alpha_0^{-1}$ pour tout i, le troisième terme du membre de gauche de l'inégalité précédente est négatif et

$$\mathbb{E}[\varepsilon_N] \leq \frac{1}{\sum_{i=0}^{N}\gamma_i}\left(\beta_N\overline{V} + \frac{3\sigma^2}{\alpha_0}\sum_{i=0}^{N}\frac{\gamma_i^2}{\beta_i}\right).$$

Deux cas de figure se présentent alors

- Si $N+1 \geq \frac{L^2\overline{V}}{\sigma^2\alpha_0}$, les choix de paramètres qui permettent de minimiser la borne en espérance sont $\gamma_i \equiv \gamma$ et $\beta_i \equiv \beta$ avec la relation

$$\frac{\beta}{\gamma} = \frac{\sigma\sqrt{3(N+1)}}{\sqrt{\alpha_0\overline{V}}},$$

  et nous obtenons alors

$$\mathbb{E}[\varepsilon_N] \leq 2\sqrt{\frac{3\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}}. \tag{3.7}$$

- Sinon, en choisissant $\gamma_i \equiv \gamma$ et $\beta_i \equiv \beta$ avec la relation $\beta\gamma^{-1} = \sqrt{3}L\alpha_0^{-1}$,il vient

$$\mathbb{E}[\varepsilon_N] \leq \frac{\sqrt{3}L}{(N+1)}\frac{\overline{V}}{\alpha_0} + \sqrt{\frac{3\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}}. \tag{3.8}$$

Nous avons donc obtenu :

Proposition 3.3 *On suppose que l'hypothèse 2.13 est vérifiée. Alors*

$$\mathbb{E}[\varepsilon_N] \leq \sqrt{\frac{3\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}} + \max\left(\sqrt{\frac{3\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}},\ \frac{\sqrt{3}L}{(N+1)}\frac{\overline{V}}{\alpha_0}\right).$$

Cette borne est obtenue lorsque les deux suites $(\beta_i)_i$ et $(\gamma_i)_i$ sont constantes, égales à β et γ respectivement, avec

$$\frac{\beta}{\gamma} = \max\left(\frac{\sigma\sqrt{3(N+1)}}{\sqrt{\alpha_0\overline{V}}},\ \frac{\sqrt{3}L}{\alpha_0}\right).$$

Remarque : Nous retrouvons dans (3.8), comme dans le cas de l'algorithme de descente en miroir, une vitesse de convergence qui dépend différemment de L et de σ. C'est là un des intérêts de l'algorithme d'extrapolation dual.

Dans le cas où le nombre de pas à effectuer dans l'algorithme 3.1 n'est pas connu à l'avance,

– si $\sqrt{2\alpha_0}\sigma \geq \sqrt{\overline{V}}L$ le choix $\gamma_i \equiv 1$ et $\beta_i = \sqrt{6\sigma^2\alpha_0^{-1}\overline{V}^{-1}(i+1)}$ permet d'obtenir

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\sigma\sqrt{6\overline{V}(N+2)}}{\sqrt{\alpha_0}(N+1)};$$

– sinon, en posant $\gamma_i \equiv 1$ et $\beta_i = \sqrt{\frac{3}{2}}\frac{L}{\alpha_0}\sqrt{i+1}$, nous obtenons

$$\mathbb{E}[\varepsilon_N] \leq \sqrt{\frac{3}{2}}\left(\frac{\overline{V}L}{\alpha_0} + \frac{4\sigma^2}{L}\right)\frac{1}{\sqrt{N+1}}.$$

3.2.3 Bornes en déviation, intervalles de confiance

Nous allons maintenant donner des bornes pour les déviations de ε_N dans le cas où l'hypothèse 3.2, l'hypothèse suivante et une des deux hypothèses 3.1 et 2.13 sont satisfaites.

Hypothèse 3.3 *Il existe une constante* $0 < \varkappa < \infty$ *telle que pour tout* $i \geq 0$, $\mathbb{E}\left[\exp\left(\varkappa\frac{\|\eta(x_i)\|_*^2}{\sigma_i^2}\right)\Big|\,\mathcal{F}_i\right] \leq e$ *p.s. et* $\mathbb{E}\left[\exp\left(\varkappa\frac{\|\eta(y_i)\|_*^2}{\sigma_i^2}\right)\Big|\,\mathcal{F}_i\right] \leq e$ *p.s.*

Commençons par le cas où l'hypothèse 3.1 est vérifiée. D'après (3.4) et (3.3), en utilisant les choix pour $(\beta_i)_i$ et $(\gamma_i)_i$ donnés dans la proposition 3.2,

$$\begin{aligned}\varepsilon_N \quad \leq \quad & \frac{1}{N+1}\left(\frac{(M+2\sigma)\sqrt{\overline{V}(N+1)}}{2\alpha_0} - \sum_{i=0}^{N}\langle\eta(y_i), y_i - x^*\rangle + \frac{3\sqrt{\overline{V}(N+1)}(M^2+2\sigma^2)}{\sqrt{2\alpha_0(M+2\sigma)}}\right.\\ & \left. + \sum_{i=0}^{N}\frac{3\sqrt{\overline{V}}}{\sqrt{2\alpha_0(N+1)}(M+2\sigma)}\left[(\|\,\eta(x_i)\,\|_*^2 - \sigma^2) + (\|\,\eta(y_i)\,\|_*^2 - \sigma^2)\right]\right).\end{aligned}$$

En appliquant la proposition B.5 à la suite de variables aléatoires $(-\frac{\sqrt{\alpha_0}}{2\sqrt{2\overline{V}}}\langle\eta(y_i), y_i - x^*\rangle)_i$ et la proposition B.6 aux variables aléatoires $(\|\,\eta(x_i)\,\|_*)_i$ et $(\|\,\eta(x_i)\,\|_*)_i$, nous obtenons la proposition suivante

Proposition 3.4 *On suppose que l'hypothèse 3.1 est vérifiée. Alors pour tout* $\alpha \geq 3\exp\left(-\varkappa\frac{(N+1)}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N+1,\alpha)$ *où*

$$\begin{aligned}\lambda^*(N+1,\alpha) \quad = \quad & \sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{(M+2\sigma)}{\sqrt{N+1}} + 2\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}}\lambda_1^*(N+1,\alpha) + \sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{3(M^2+2\sigma^2)}{\sqrt{N+1}(M+2\sigma)}\\ & + \sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{3\sigma^2}{(M+2\sigma)\sqrt{N+1}}\lambda_2^*(N+1,\alpha),\end{aligned}$$

avec

$$\lambda_1^*(N+1,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(3\alpha^{-1})} \text{ et } \lambda_2^*(N+1,\alpha) = \sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(3\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N \geq \lambda) \leq \alpha.$$

Considérons maintenant le cas où l'hypothèse 2.13 est vérifiée et où les choix de paramètres effectués sont ceux donnés dans la proposition 3.3. Nous avons alors

$$\begin{aligned}\varepsilon_N &\leq \frac{\overline{V}}{N+1}\max\left(\frac{\sigma\sqrt{3(N+1)}}{\sqrt{\alpha_0\overline{V}}},\ \frac{\sqrt{3}L}{\alpha_0}\right)-\frac{1}{N+1}\sum_{i=0}^{N}\langle\eta(y_i),y_i-x^*\rangle\\ &\quad+\frac{1}{N+1}\sum_{i=0}^{N}\frac{\sqrt{3\overline{V}}}{2\sigma\sqrt{\alpha_0(N+1)}}(\parallel\eta(x_i)\parallel_*^2+\parallel\eta(y_i)\parallel_*^2).\end{aligned}$$

En appliquant alors la proposition B.5 à la suite de variables aléatoires $(-\frac{\sqrt{\alpha_0}}{2\sqrt{2\overline{V}}}\langle\eta(y_i),y_i-x^*\rangle)_i$ et la proposition B.6 aux variables aléatoires $(\parallel\eta(x_i)\parallel_*)_i$ et $(\parallel\eta(x_i)\parallel_*)_i$, nous obtenons

Proposition 3.5 *On suppose que l'hypothèse 2.13 est vérifiée. Alors pour tout* $\alpha\geq 3\exp\left(-\varkappa\frac{(N+1)}{4}\right)$ *et tout* $\lambda\geq\lambda^*(N+1,\alpha)$ *où*

$$\begin{aligned}\lambda^*(N+1,\alpha) &= \frac{\overline{V}}{N+1}\max\left(\frac{\sigma\sqrt{3(N+1)}}{\sqrt{\alpha_0\overline{V}}},\ \frac{\sqrt{3}L}{\alpha_0}\right)+2\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}}\lambda_1^*(N+1,\alpha)\\ &\quad+\sqrt{\frac{3\overline{V}}{\alpha_0}}\frac{\sigma}{\sqrt{N+1}}(1+\lambda_2^*(N+1,\alpha)),\end{aligned}$$

avec

$$\lambda_1^*(N+1,\alpha)=\sqrt{\frac{4}{\varkappa}\ln(3\alpha^{-1})}\ \textit{et}\ \lambda_2^*(N+1,\alpha)=\sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(3\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N\geq\lambda)\leq\alpha.$$

3.3 Algorithme pour les inégalités variationnelles

Comme pour l'algorithme de descente en miroir, l'algorithme d'extrapolation dual peut être appliqué à de larges classes de problèmes. Nous abordons ici le cas des problèmes variationnels à opérateur monotone que nous avons déjà abordé. A savoir nous nous plaçons dans un ensemble convexe fermé Q inclus dans l'espace vectoriel de dimension finie $(E,\|\cdot\|)$ et nous notons $(E^*,\|\cdot\|_*)$ l'espace dual et la norme duale associée. Nous voulons résoudre le problème

$$\text{Trouver } x^*\in Q:\ \langle g(x),x-x^*\rangle\geq 0\ \forall x\in Q,$$

où $g:Q\to E^*$ est un opérateur continu et monotone. Nous supposons dans ce qui suit que l'une des deux hypothèses suivantes est vérifiée.

Hypothèse 3.4 *L'opérateur g est à variations bornées sur Q. Il existe une constante $M>0$ telle que*

$$\forall x,y\in Q\ \ \parallel g(x)-g(y)\parallel_*\leq M.$$

Hypothèse 3.5 *L'opérateur g est lipschitzien sur Q, il existe une constante $L>0$ telle que*

$$\parallel g(x)-g(y)\parallel_*\leq L\parallel x-y\parallel\ \ \forall x,y\in Q.$$

Nous disposons en outre d'un oracle qui nous fournit une version aléatoire Ψ de g et on note $\eta = \Psi - g$. Les hypothèses à faire concernant Ψ seront discutées ultérieurement.
Comme précédemment, nous nous donnons une fonction d fortement convexe sur Q de paramètre α_0 pour la norme $\|\cdot\|$.
Soit d une fonction fortement convexe sur Q de paramètre α_0 que nous supposons différentiable sur Q. Nous noterons V la divergence de Bregman associée, c'est-à-dire

$$\forall x, y \in Q \; V(x,y) = d(y) - d(x) - \langle d'(x), y - x\rangle.$$

Fixons-nous un point $\overline{x}$ comme 'centre' de l'ensemble Q. Nous nous intéressons dans la suite à la fonction de mérite restreinte suivante :

$$\epsilon_{\overline{V}}(x) = \max_{y \in Q}\{\langle g(y), y - x\rangle \,|V(\overline{x}, y) \leq \overline{V}\},$$

où $\overline{V} > 0$ est un paramètre fixé.
Nous aurons besoin des fonctions de support de Q définies sur E^* suivantes :

$$\begin{aligned} \xi_{\overline{V}}(s) &= \max_x\{\langle s, x - \overline{x}\rangle \,|V(\overline{x}, x) \leq \overline{V}\}, \\ W_\beta(z,s) &= \max_x\{\langle s, x - \overline{x}\rangle - \beta V(z,x) \,|V(\overline{x}, x) \leq \overline{V}\}, \end{aligned}$$

où $\beta, \overline{V} > 0$ et $z \in Q$ et nous noterons $T_\beta(z,s) = \arg\max_x\{\langle s, x - z\rangle - \beta V(z,x)\}$.
Étant données deux constantes strictement positives β et γ, le pas d'extrapolation duale $\mathcal{E}_{\beta,\gamma}(s)$ qui transforme un point arbitraire $s \in E^*$ en un nouveau point s_+ est définit par :

$$(x, y, s_+) = \mathcal{E}_{\beta,\gamma}(s)_- \Leftrightarrow \left\{ \begin{array}{rcl} x &=& T_\beta(\overline{x}, s), \\ y &=& T_\beta(x, -\gamma\Psi(x)), \\ s_+ &=& s - \gamma\Psi(y). \end{array} \right.$$

L'algorithme qui nous intéresse est alors donné par

Algorithme 3.2

- *Initialisation :* $s_{-1} = 0 \in E^*$.
- *Pas : pour* $k = 0$ *à* N *faire*

$$\begin{aligned} x_k &= T_{\beta_k}(\overline{x}, s_{k-1}), \\ y_k &= T_{\beta_k}(x_k, -\gamma_k\Psi(x_k)), \\ s_k &= s_{k-1} - \gamma_k\Psi(y_k). \end{aligned}$$

- *Sortie*

$$y^N = \frac{\sum_{i=0}^N \gamma_i y_i}{\sum_{i=0}^N \gamma_i}$$

Nous allons maintenant étudier la convergence de l'algorithme. Notons

$$\epsilon_{\overline{V}}(y^N) = \max\{\langle g(x), y^N - x\rangle \,|V(\overline{x}, x) \leq \overline{V}\}.$$

Proposition 3.6 *$\forall N \geq 0$ et $\forall \beta > 0$:*

$$\left(\sum_{i=0}^{N} \gamma_i\right) \epsilon_{\overline{V}}(y^N) \leq (\beta + \beta_N)\overline{V} + \sum_{i=0}^{N} \gamma_i \langle \eta(y_i), \check{y}_i - y_i \rangle + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_0 \beta} \parallel \eta(y_i) \parallel_*^2$$
$$+ \sum_{i=0}^{N} \left(\frac{\gamma_i^2}{2\alpha_0 \beta_i} \parallel \Psi(x_i) - \Psi(y_i) \parallel_*^2 - \frac{\alpha_0 \beta_i}{2} \parallel x_i - y_i \parallel^2 \right),$$

où

$$\check{y}_i = T_\beta \left(\overline{x}, \sum_{j=0}^{i-1} \gamma_i \eta(y_i) \right) - \overline{x}.$$

Cherchons maintenant des bornes en espérance pour $\epsilon_{\overline{V}}(y^N)$. Nous nous plaçons dans le contexte probabiliste suivant : $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$ et pour tout $i \geq 0$, $\mathcal{F}_i$ est la tribu engendrée par $\mathcal{F}_{i-1}$ et les variables aléatoires $\eta(x_i)$ et $\eta(y_i)$. Nous supposerons dans la suite que l'hypothèse suivante est satisfaite

Hypothèse 3.6 *Il existe $0 < \sigma^2 < +\infty$ telle que pour tout i,*

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0 \text{ p.s., } \mathbb{E}[\eta(y_i)|\mathcal{F}_i] = 0 \text{ p.s.,}$$

$$\mathbb{E}[\parallel \eta(x_i) \parallel_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s. et } \mathbb{E}[\parallel \eta(y_i) \parallel_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s.}$$

Proposition 3.7 *Supposons que l'hypothèse 3.4 est vérifiée. Alors si on choisit alors les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes et satisfaisant*

$$\gamma_N = 1 \text{ et } \beta_N = (2\sigma + M)\sqrt{\frac{(N+1)}{2\alpha_0 \overline{V}}},$$

on a

$$\mathbb{E}[\epsilon_{\overline{V}}(y^N)] \leq \sqrt{2}(M + 3\sigma)\sqrt{\frac{\overline{V}}{\alpha_0(N+1)}}.$$

Preuve
Il suffit de prendre l'espérance de l'expression de $\epsilon_{\overline{V}}(N)$ donnée dans la proposition 3.6 en choisissant $\beta = \sigma\sqrt{\frac{N+1}{2\alpha_0 \overline{V}}}$ et en remarquant que pour tout i, la variable aléatoire $\check{y}_i - y_i$ est $\mathcal{F}_{i-1}$-mesurable. □

Supposons maintenant que l'hypothèse 3.5 est vérifiée. Alors
– Si $L^2 \leq \frac{(N+1)\sigma^2\alpha_0}{\overline{V}}$, en choisissant les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes et telles que

$$\gamma_N = 1 \text{ et } \beta_N = \sigma\sqrt{\frac{3(N+1)}{\alpha_0 \overline{V}}},$$

nous obtenons

$$\mathbb{E}[\epsilon_{\overline{V}}(y^N)] \leq \frac{(2\sqrt{3} + \sqrt{2})\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}}.$$

– Sinon, en choisissant les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes et telles que

$$\gamma_N = 1 \text{ et } \beta_N = \sqrt{\frac{3}{\alpha_0}}L,$$

il vient

$$\mathbb{E}[\epsilon_{\overline{V}}(y^N)] \leq \frac{\sqrt{3}L}{(N+1)}\frac{\overline{V}}{\alpha_0} + \frac{(\sqrt{3}+\sqrt{2})\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}}.$$

Nous avons donc obtenu le résultat suivant

Proposition 3.8 *On suppose que l'hypothèse 3.5 est vérifiée. Alors on a*

$$\mathbb{E}[\epsilon_{\overline{V}}(y^N)] \leq \frac{(\sqrt{3}+\sqrt{2})\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}} + \max\left(\frac{\sqrt{3}L}{(N+1)}\frac{\overline{V}}{\alpha_0}, \frac{\sqrt{3}\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}}\right).$$

Cette borne est obtenue lorsque les suites $(\gamma_i)_i$ et $(\beta_i)_i$ sont constantes et égales à 1 et $\max\left(\sigma\sqrt{\frac{3(N+1)}{\alpha_0\overline{V}}}, \sqrt{\frac{3L}{\alpha_0}}\right)$ *respectivement.*

Donnons maintenant des bornes pour les déviations de $\epsilon_{\overline{V}}(y^N)$ l'hypothèse 3.6, l'hypothèse suivante et l'une des hypothèses 3.4 et 3.5 sont vérifiées.

Hypothèse 3.7 *Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$, $\mathbb{E}\left[\exp\left(\varkappa\frac{\|\eta(x_i)\|_*^2}{\sigma_i^2}\right)\Big|\mathcal{F}_i\right] \leq e$ p.s. et $\mathbb{E}\left[\exp\left(\varkappa\frac{\|\eta(y_i)\|_*^2}{\sigma_i^2}\right)\Big|\mathcal{F}_i\right] \leq e$ p.s.*

Commençons par traiter le cas où l'hypothèse 3.4 est vérifiée. En utilisant la borne obtenue dans la proposition 3.6 avec les valeurs des paramètres donnés dans la proposition 3.7,nous obtenons

$$\begin{aligned}\epsilon_{\overline{V}}(y^N) \leq{}& \frac{(M+4\sigma)}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{1}{N+1}\sum_{i=0}^{N}\langle\eta(y_i), \check{y}_i - y_i\rangle + \sqrt{\frac{\overline{V}}{2\alpha_0}}\sum_{i=0}^{N}\frac{\| \eta(y_i) \|_*^2 - \sigma^2}{\sigma(N+1)^{\frac{3}{2}}} \\ &+ \frac{3}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{(M^2+2\sigma^2)}{(M+2\sigma)} + \sum_{i=0}^{N}\frac{3}{(M+2\sigma)(N+1)^{\frac{3}{2}}}\sqrt{\frac{\overline{V}}{2\alpha_0}}\left(\| \eta(y_i) \|_*^2 \| \eta(x_i) \|_*^2 - 2.\sigma^2\right)\end{aligned}$$

Appliquons maintenant la proposition B.5 à la suite de variables aléatoires $(\frac{\sqrt{\alpha_0}}{2\sqrt{2\overline{V}}}\langle\eta(y_i), \check{y}_i - y_i\rangle)_i$ et la proposition B.6 aux variables aléatoires $(\| \eta(x_i) \|_*)_i$ et $(\| \eta(x_i) \|_*)_i$. Nous obtenons

Proposition 3.9 *On suppose que l'hypothèse 3.4 est vérifiée. Alors pour tout $\alpha \geq 3\exp\left(-\varkappa\frac{(N+1)}{4}\right)$ et tout $\lambda \geq \lambda^*(N+1, \alpha)$ où*

$$\lambda^*(N+1,\alpha) = \frac{4M+7\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{4\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{2\alpha_0}}\sqrt{\frac{4}{\varkappa}\ln(3\alpha^{-1})} + (4\sigma)\sqrt{\frac{\overline{V}}{2\alpha_0(N+1)}}\sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(3\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N \geq \lambda) \leq \alpha.$$

Proposition 3.10 *On suppose que l'hypothèse 3.5 est vérifiée. Alors pour tout* $\alpha > 3\exp\left(-\varkappa\frac{N+1}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N+1,\alpha)$ *où*

$$\begin{aligned}\lambda^*(N+1,\alpha) &= \frac{(\sqrt{2}+\sqrt{3})\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}} + \frac{\overline{V}}{N+1}\max\left\{\sigma\sqrt{\frac{3(N+1)}{2\alpha_0\overline{V}}},\sqrt{\frac{3L}{\alpha_0}}\right\} \\ &\quad + \frac{2}{\sqrt{N+1}}\sqrt{\frac{2\overline{V}}{\alpha_0}}\sqrt{\frac{4}{\varkappa}\ln(3\alpha^{-1})} + \frac{(\sqrt{2}+2\sqrt{3})\sigma}{2\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}}\sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(3\alpha^{-1})},\end{aligned}$$

on a

$$\mathbb{P}(\epsilon_{\overline{V}}(y^N) \geq \lambda) \leq \alpha.$$

3.4 Algorithme pour les problèmes de point-selle

Nous nous intéressons de nouveau au problème de recherche de point selle d'une fonction convexe-concave $L(x,y)$ (L est convexe en x et concave en y) où les variables x et y appartiennent respectivement à des ensembles convexes fermés X et Y. Le problème à résoudre est donc le suivant :

$$\min_{x\in X}\max_{y\in Y} L(x,y).$$

X et Y sont supposés inclus dans des espaces vectoriels de dimension finie E_X et E_Y munis respectivement des normes $\|\cdot\|_X$ et $\|\cdot\|_Y$. Nous notons $(E_X^*, \|\cdot\|_{X,*})$ et $(E_X^*, \|\cdot\|_{X,*})$ les espaces duaux et les normes duales associés.
Nous supposons que f est continue et nous notons $\partial_x L(x,y)$ le sous-différentiel de la fonction $x \to L(x,y)$ et $\partial_y L(x,y)$ le super-différentiel de la fonction $y \to L(x,y)$; dans ce qui suit, $L'_X(x,y)$ (respectivement $L'_Y(x,y)$) désignera un élément quelconque de $\partial_x L(x,y)$ (respectivement $\partial_y L(x,y)$).
Nous supposons de plus que L vérifie une des deux hypothèses suivantes

Hypothèse 3.8 *L'_X et L'_Y sont à variations bornées sur $Q = X \times Y$, c'est-à-dire que*

$$\forall (x,y),(w,z) \in Q, \left\{\begin{array}{rcl} \| L'_X(x,y) - L'_X(w,z) \|_{X,*} & \leq & M_X, \\ \| L'_Y(x,y) - L'_Y(w,z) \|_{Y,*} & \leq & M_Y, \end{array}\right.$$

où M_X et M_Y sont deux constantes strictement positives.

Hypothèse 3.9 *L'_X et L'_Y sont lipschitziens sur $Q = X \times Y$, c'est-à-dire que*

$$\forall (x,y),(w,z) \in Q, \left\{\begin{array}{rcl} \| L'_X(x,y) - L'_X(w,z) \|_{X,*} & \leq & M_{X,X} \| x-w \|_X + M_{X,Y} \| y-z \|_Y, \\ \| L'_Y(x,y) - L'_Y(w,z) \|_{Y,*} & \leq & M_{Y,X} \| x-w \|_X + M_{Y,Y} \| y-z \|_Y \end{array}\right.$$

où $M_{X,X}$, $M_{X,Y}$, $M_{Y,X}$ et $M_{Y,Y}$ sont quatre constantes strictement positives.

Les réalisations de ces différentes fonctions (c'est-à-dire L'_X et L'_Y) nous sont fournies par un oracle stochastique. Nous noterons $\eta_X(x,y) = \Phi(x,y) - L'_X(x,y)$ et $\eta_Y(x,y) = \Psi(x,y) - L'_X(x,y)$. Les hypothèses portant sur Φ et Ψ seront données lors de l'étude de la précision de l'algorithme.
Comme dans le cas de l'algorithme de descente en miroir, nous ne traiterons pas le problème du point selle comme un simple cas particulier du problème d'inégalité variationnelle, mais nous chercherons à particulariser les deux variables x et y.

Pour mettre en oeuvre l'algorithme, nous avons aussi besoin de deux fonctions fortement convexes d_X et d_Y de paramètres respectifs $\alpha_{0,X}$ et $\alpha_{0,Y}$. Nous supposons que d_X et d_Y sont différentiables sur leurs domaines X et Y et nous définissons

$$\begin{aligned} V_X(x,u) &= d_X(x) - d_X(u) + \langle d'_X(x), u-x\rangle, \\ V_Y(y,v) &= d_Y(y) - d_Y(v) + \langle d'_Y(y), v-y\rangle. \end{aligned}$$

Fixons arbitrairement deux points $\overline{x}$ et $\overline{y}$ pour 'centres' des ensembles X et Y respectivement et $\overline{V}_X$, $\overline{V}_Y > 0$ tels que $x^* \in \mathcal{F}_{\overline{V}_X}$ et $y^* \in \mathcal{F}_{\overline{V}_Y}$ où (x^*, y^*) est le point-selle de L et

$$\begin{aligned} \mathcal{F}_{\overline{V}_X} &= \{x \in X | V_X(\overline{x}, x) < \overline{V}_X\}, \\ \mathcal{F}_{\overline{V}_Y} &= \{y \in Y | V_Y(\overline{y}, y) < \overline{V}_Y\}. \end{aligned}$$

Nous serons amenés par la suite à considérer les fonctions de support suivantes :

$$\begin{aligned} \forall s \in E_X^* \;\; \xi_{\overline{V}_X}(s) &= \max_x \left\{ \langle s, x-\overline{x}\rangle,\; x \in \mathcal{F}_{\overline{V}_X} \right\}, \\ \forall t \in E_Y^* \;\; \xi_{\overline{V}_Y}(t) &= \max_y \left\{ \langle t, y-\overline{y}\rangle,\; y \in \mathcal{F}_{\overline{V}_Y} \right\}, \\ \forall (u,s) \in E_X \times E_X^*,\; \beta > 0 \;\; W_{X,\beta}(u,s) &= \max_{x \in X} \left\{ \langle s, x-u\rangle - \beta V_X(u,x) \right\}, \\ \forall (v,t) \in E_Y \times E_Y^*,\; \beta > 0 \;\; W_{Y,\beta}(v,t) &= \max_{y \in Y} \left\{ \langle t, y-v\rangle - \beta V_Y(v,y) \right\}. \end{aligned}$$

Notons $T_\beta^X(z,s) = \mathrm{argmax}_x\{\langle s, x-z\rangle - \beta V_X(z,x) \,:\, x \in X\}$, donnons-nous $\beta, \gamma > 0$, $y, v \in Y$ et considérons le pas suivant $\mathcal{E}_{\beta,\gamma}^X$, qui transforme un point arbitraire $s \in E_X^*$ en un nouveau point s_+. Plus précisément, $\mathcal{E}_{\beta,\gamma}^X(s,y,v) = (x,u,s_+)$ avec

$$\begin{aligned} x &= T_\beta^X(\overline{x}, s), \\ u &= T_\beta^X(x, -\gamma\Phi(x,y)), \\ s_+ &= s - \gamma\Phi(u,v). \end{aligned}$$

De manière analogue, notons pour tous $\zeta \in Y$, $t \in E_Y^*$, β et $\gamma > 0$, $T_\beta^Y(\zeta,t) = \mathrm{argmax}_y\{\langle t, v-\zeta\rangle - \beta V_Y(\zeta,y) \,:\, y \in Y\}$ et on considérons le pas $\mathcal{E}_{\beta,\gamma}^X(t,x,u) = (y,v,t_+)$ avec $t, t_+ \in E_Y^*$ et $x \in X$ donné par

$$\begin{aligned} y &= T_\beta^Y(\overline{y}, t), \\ v &= T_\beta^Y(y, \gamma\Psi(x,y)), \\ t_+ &= t + \gamma\Psi(u,v). \end{aligned}$$

Nous avons maintenant tous les éléments en main pour pouvoir décrire l'algorithme. Le choix des suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ sera discuté ultérieurement. Nous supposerons seulement pour le moment que chacune de ces trois suites est strictement positive et que $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ sont décroissantes.

Algorithme 3.3

– *Initialisation :* $s_{-1}, = 0 \in E_X^*$, $t_{-1}, = 0 \in E_Y^*$.
– *Pas : pour* $k = 0$ *à* N *faire*

$$\begin{aligned}
x_k &= T^X_{\beta_{X,k}}(\overline{x}, s_{k-1}),\\
y_k &= T^Y_{\beta_{Y,k}}(\overline{y}, t_{k-1}),\\
u_k &= T^X_{\beta_{X,k}}(x_k, -\gamma_k \Phi(x_k, y_k)),\\
v_k &= T^Y_{\beta_{Y,k}}(y_k, +\gamma_k \Psi(x_k, y_k)),\\
s_k &= s_{k-1} - \gamma_k \Phi(u_k, v_k),\\
t_k &= t_{k-1} + \gamma_k \Psi(u_k, v_k).
\end{aligned}$$

– *Sortie*

$$u^N = \frac{\sum_{i=0}^N \gamma_i u_i}{\sum_{i=0}^N \gamma_i}, v^N = \frac{\sum_{i=0}^N \gamma_i v_i}{\sum_{i=0}^N \gamma_i}$$

Nous souhaitons borner l'erreur $\varepsilon_N = \max_{y \in \mathcal{F}_{\overline{V}_Y}} L(u^N, y) - \min_{x \in \mathcal{F}_{\overline{V}_X}} L(x, v^N)$.

Proposition 3.11 *Pour tous* $\beta_X > 0$ *et* $\beta_Y > 0$,

$$\begin{aligned}
\left(\sum_{i=0}^N \gamma_i\right) \varepsilon_N \leq{}& (\beta_X + \beta_{X,N})\overline{V}_X + \sum_{i=0}^N \gamma_i \langle \eta_X(u_i, v_i), \check{u}_i - u_i\rangle + \sum_{i=0}^N \frac{\gamma_i^2}{2\alpha_{X,0}\beta_X} \parallel \eta_X(u_i, v_i) \parallel^2_{X,*}\\
&+ (\beta_Y + \beta_{Y,N})\overline{V}_Y - \sum_{i=0}^N \gamma_i \langle \eta_Y(u_i, v_i), \check{v}_i - v_i\rangle + \sum_{i=0}^N \frac{\gamma_i^2}{2\alpha_{Y,0}\beta_Y} \parallel \eta_Y(u_i, v_i) \parallel^2_{Y,*}\\
&+ \sum_{i=0}^N \left(\frac{\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}} \parallel \Phi(x_i, y_i) - \Phi(u_i, v_i) \parallel^2_{X,*} - \frac{\alpha_{X,0}\beta_{X,i}}{2} \parallel x_i - u_i \parallel^2_X\right)\\
&+ \sum_{i=0}^N \left(\frac{\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}} \parallel \Psi(x_i, y_i) - \Psi(u_i, v_i) \parallel^2_{Y,*} - \frac{\alpha_{Y,0}\beta_{Y,i}}{2} \parallel y_i - v_i \parallel^2_Y\right), \quad (3.9)
\end{aligned}$$

où $\check{u}_i = T^X_{\beta_X}\left(\overline{x}, \sum_{j=0}^{i-1} \gamma_j \eta_X(u_j, v_j)\right)$ *et* $\check{v}_i = T^Y_{\beta_Y}\left(\overline{y}, -\sum_{j=0}^{i-1} \gamma_j \eta_Y(u_j, v_j)\right)$.

Donnons maintenant des bornes en espérance pour ε_N. Pour cela notons $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$ et pour tout $i \geq 0$, notons $\mathcal{F}_i$ la tribu engendrée par $\mathcal{F}_{i-1}$ et les variables aléatoires $\eta_X(u_i, v_i)$, $\eta_X(x_i, y_i)$, $\eta_Y(u_i, v_i)$ et $\eta_Y(x_i, y_i)$. Nous supposerons que l'hypothèse suivante est satisfaite

Hypothèse 3.10 *Il existe deux constantes strictement positives* σ_X *et* σ_Y *telles que pour tout* $i \geq 0$ *les relations suivantes sont vraies presque sûrement*

$$\begin{array}{ll}
\mathbb{E}[\eta_X(u_i, v_i)|\mathcal{F}_i] = 0, & \mathbb{E}[\eta_X(x_i, y_i)|\mathcal{F}_i] = 0,\\
\mathbb{E}[\eta_Y(u_i, v_i)|\mathcal{F}_i] = 0, & \mathbb{E}[\eta_Y(x_i, y_i)|\mathcal{F}_i] = 0,\\
\mathbb{E}[\parallel \eta_X(u_i, v_i) \parallel^2_{X,*} |\mathcal{F}_i] \leq \sigma_X^2, & \mathbb{E}[\parallel \eta_X(x_i, y_i) \parallel_{X,*} |\mathcal{F}_i] \leq \sigma_X^2,\\
\mathbb{E}[\parallel \eta_Y(u_i, v_i) \parallel^2_{Y,*} |\mathcal{F}_i] \leq \sigma_Y^2, & \mathbb{E}[\parallel \eta_Y(x_i, y_i) \parallel_{Y,*} |\mathcal{F}_i] \leq \sigma_Y^2.
\end{array}$$

Commençons par le cas où cette hypothèse et l'hypothèse 3.8 sont vérifiées. Nous avons alors

$$\begin{aligned} \| \Phi(x_i, y_i) - \Phi(u_i, v_i) \|_{X,*} &\leq M_X + \| \eta_X(x_i, y_i) \|_{X,*} + \| \eta_X(u_i, v_i) \|_{X,*}, \\ \| \Psi(x_i, y_i) - \Psi(u_i, v_i) \|_{Y,*} &\leq M_Y + \| \eta_Y(x_i, y_i) \|_{Y,*} + \| \eta_Y(u_i, v_i) \|_{Y,*} . \end{aligned}$$

En prenant l'espérance dans (3.9) il vient

$$\begin{aligned} \left(\sum_{i=0}^{N} \gamma_i\right) \varepsilon_N \leq (\beta_X + \beta_{X,N})\overline{V}_X + (\beta_Y + \beta_{Y,N})\overline{V}_Y + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_{X,0}\beta_X}\sigma_X^2 + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_{Y,0}\beta_Y}\sigma_Y^2 \\ + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}(M_X + 2\sigma_X)^2 + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}}(M_Y + 2\sigma_Y)^2. \end{aligned}$$

Cette borne est minimale lorsque les trois suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{X,i})_i$ sont constantes et que $\gamma_i = 1$,

$$\begin{aligned} \beta_X &= \frac{\sigma_X\sqrt{N+1}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}, \qquad \beta_{X,i} = \frac{(M_X + 2\sigma_X)\sqrt{N+1}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}, \\ \beta_Y &= \frac{\sigma_Y\sqrt{N+1}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}, \qquad \beta_{Y,i} = \frac{(M_Y + 2\sigma_Y)\sqrt{N+1}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}. \end{aligned}$$

Alors

Proposition 3.12

$$\mathbb{E}[\varepsilon_N] \leq \frac{\sqrt{2}(M_X + 3\sigma_X)}{\sqrt{N+1}}\sqrt{\frac{\overline{V}_X}{\alpha_{X,0}}} + \frac{\sqrt{2}(M_Y + 3\sigma_Y)}{\sqrt{N+1}}\sqrt{\frac{\overline{V}_Y}{\alpha_{Y,0}}}.$$

Dans le cas où l'hypothèse 3.9 est vérifiée, nous avons

$$\begin{aligned} \| \Phi(x_i, y_i) - \Phi(u_i, v_i) \|_{X,*} &\leq M_{X,X} \| x_i - u_i \|_X + M_{X,Y} \| y_i - v_i \|_Y + \| \eta_X(x_i, y_i) \|_{X,*} + \| \eta_X(u_i, v_i) \|_{X,*}, \\ \| \Psi(x_i, y_i) - \Psi(u_i, v_i) \|_{Y,*} &\leq M_{Y,X} \| x_i - u_i \|_X + M_{Y,Y} \| y_i - v_i \|_Y + \| \eta_Y(x_i, y_i) \|_{Y,*} + \| \eta_Y(u_i, v_i) \|_{Y,*} . \end{aligned}$$

Et

$$\begin{aligned}\left(\sum_{i=0}^{N}\gamma_i\right)\varepsilon_N \leq{}& (\beta_X+\beta_{X,N})\overline{V}_X+\sum_{i=0}^{N}\gamma_i\langle\eta_X(u_i,v_i),\check{u}_i-u_i\rangle+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{X,0}\beta_X}\parallel\eta_X(u_i,v_i)\parallel_{X,*}^2\\ &+(\beta_Y+\beta_{Y,N})\overline{V}_Y-\sum_{i=0}^{N}\gamma_i\langle\eta_Y(u_i,v_i),\check{v}_i-v_i\rangle+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{Y,0}\beta_Y}\parallel\eta_Y(u_i,v_i)\parallel_{Y,*}^2\\ &+\sum_{i=0}^{N}\frac{3\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}\left(\parallel\eta_X(u_i,v_i)\parallel_{X,*}+\parallel\eta_X(x_i,y_i)\parallel_{X,*}\right)^2\\ &+\sum_{i=0}^{N}\frac{3\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}}\left(\parallel\eta_Y(u_i,v_i)\parallel_{Y,*}+\parallel\eta_Y(x_i,y_i)\parallel_{Y,*}\right)^2\\ &+\sum_{i=0}^{N}\left(\frac{3\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}M_{X,X}^2+\frac{3\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}}M_{Y,X}^2-\frac{\alpha_{X,0}\beta_{X,i}}{2}\right)\parallel x_i-u_i\parallel_X^2\\ &+\sum_{i=0}^{N}\left(\frac{3\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}}M_{Y,Y}^2+\frac{3\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}M_{X,Y}^2-\frac{\alpha_{Y,0}\beta_{Y,i}}{2}\right)\parallel y_i-v_i\parallel_Y^2 . \quad (3.10)\end{aligned}$$

Considérons alors le système d'équations suivant

$$(S)\quad\left\{\begin{array}{rcl}\alpha_{X,0}\beta_{X,i} & \geq & \frac{3\gamma_i^2}{\alpha_{X,0}\beta_{X,i}}M_{X,X}^2+\frac{3\gamma_i^2}{\alpha_{Y,0}\beta_{Y,i}}M_{Y,X}^2\\ \alpha_{Y,0}\beta_{Y,i} & \geq & \frac{3\gamma_i^2}{\alpha_{Y,0}\beta_{Y,i}}M_{Y,Y}^2+\frac{3\gamma_i^2}{\alpha_{X,0}\beta_{X,i}}M_{X,Y}^2\end{array}\right.$$

Si (S) est satisfait, nous obtenons en prenant l'espérance dans (3.10)

$$\begin{aligned}\mathbb{E}[\varepsilon_N]\leq\left(\sum_{i=0}^{N}\gamma_i\right)^{-1}\Bigg[&(\beta_X+\beta_{X,N})\overline{V}_X+(\beta_Y+\beta_{Y,N})\overline{V}_Y+\left(\frac{\sigma_X^2}{2\alpha_{X,0}\beta_X}+\frac{\sigma_Y^2}{2\alpha_{Y,0}\beta_Y}\right)\sum_{i=0}^{N}\gamma_i^2\\ &+\sum_{i=0}^{N}\gamma_i^2\left(\frac{6\sigma_X^2}{\alpha_{X,0}\beta_{X,i}}+\frac{6\sigma_Y^2}{\alpha_{Y,0}\beta_{Y,i}}\right)\Bigg] . \quad (3.11)\end{aligned}$$

La borne (3.11) est minimale lorsque les trois suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{X,i})_i$ sont constantes et que $\gamma_i=1$,

$$\beta_X=\frac{\sigma_X\sqrt{N+1}}{\sqrt{2\alpha_{X,0}\overline{V}_X}},\qquad \beta_{X,i}=\frac{\sigma_X\sqrt{6(N+1)}}{\sqrt{\alpha_{X,0}\overline{V}_X}},$$
$$\beta_Y=\frac{\sigma_Y\sqrt{N+1}}{\sqrt{2\alpha_{Y,0}Y_X}},\qquad \beta_{Y,i}=\frac{\sigma_Y\sqrt{6(N+1)}}{\sqrt{\alpha_{Y,0}\overline{V}_Y}}.$$

Par conséquent, si ces valeurs satisfont (S), il vient

Proposition 3.13

$$\mathbb{E}[\varepsilon_N]\leq\frac{\sqrt{2}(1+2\sqrt{3})}{\sqrt{N+1}}\left(\sigma_X\sqrt{\frac{\overline{V}_X}{\alpha_{X,0}}}+\sigma_Y\sqrt{\frac{\overline{V}_Y}{\alpha_{Y,0}}}\right).$$

Sinon, il nous faut minimiser (3.11) sous contrainte (S). Remarquons que la solution est encore atteinte lorsque les les trois suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{X,i})_i$ sont constantes et par conséquent on peut garder les valeurs de $(\gamma_i)_i$, β_X et β_Y données ci-dessus. Trouver des valeurs adéquates pour $(\beta_{X,i})_i$ et $(\beta_{X,i})_i$ revient donc à résoudre un problème de dimension 2 ce qui peut être fait efficacement au cas par cas.

Afin de fixer les idées sur la borne obtenue, nous allons imposer la contrainte supplémentaire $\beta_{X,i} = \beta_{Y,i} = \beta_i$ pour tout i. Le système (S) à satisfaire devient alors

$$(S') \;\; \beta_i^2 \geq 3\gamma_i^2 \max\left(\frac{M_{X,X}^2}{\alpha_{X,0}^2} + \frac{M_{Y,X}^2}{\alpha_{X,0}\alpha_{Y,0}}, \frac{M_{Y,Y}^2}{\alpha_{Y,0}^2} + \frac{M_{X,Y}^2}{\alpha_{X,0}\alpha_{Y,0}}\right).$$

Nous choisirons comme précédemment les deux suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes et égales à 1 et β respectivement où :

$$\beta = \begin{cases} \beta^* = \sqrt{\frac{6(N+1)}{(\overline{V}_X+\overline{V}_Y)}\left(\frac{\sigma_X^2}{\alpha_{X,0}} + \frac{\sigma_Y^2}{\alpha_{Y,0}}\right)} & \text{si } \beta^* \text{ satisfait } (S'), \\ \tilde{\beta} = \sqrt{3\max\left(\frac{M_{X,X}^2}{\alpha_{X,0}^2} + \frac{M_{Y,X}^2}{\alpha_{X,0}\alpha_{Y,0}}, \frac{M_{Y,Y}^2}{\alpha_{Y,0}^2} + \frac{M_{X,Y}^2}{\alpha_{X,0}\alpha_{Y,0}}\right)} & \text{sinon.} \end{cases}$$

Ce choix pour les paramètres permet d'obtenir

$$\mathbb{E}[\varepsilon_N] \leq \frac{\sigma_X}{\sqrt{N+1}}\sqrt{\frac{2\overline{V}_X}{\alpha_X,0}} + \frac{\sigma_Y}{\sqrt{N+1}}\sqrt{\frac{2\overline{V}_Y}{\alpha_Y,0}} + \sqrt{\frac{6(\overline{V}_X+\overline{V}_Y)}{N+1}\left(\frac{\sigma_X^2}{\alpha_{X,0}} + \frac{\sigma_Y^2}{\alpha_{Y,0}}\right)}$$
$$+ \max\left(\sqrt{\frac{6(\overline{V}_X+\overline{V}_Y)}{N+1}\left(\frac{\sigma_X^2}{\alpha_{X,0}} + \frac{\sigma_Y^2}{\alpha_{Y,0}}\right)}, \frac{\tilde{\beta}(\overline{V}_X+\overline{V}_Y)}{N+1}\right).$$

Remarque : La borne précédente n'est pas la meilleure que l'on puisse obtenir dans le cas où les deux suites $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ sont égales. En effet, forcer cette égalité revient à se replacer dans plus général des problèmes variationnels à opérateur monotone que nous avons déjà abordés en choisissant pour normes sur $E = E_X \times E_Y$ et sur $E^* = E_X^* \times E_Y^*$ les normes $\|\cdot\|$ et $\|\cdot\|_*$ données respectivement par

$$\begin{aligned} \forall (x,y) \in E_X \times E_Y \;\; \| (x,y) \| &= \max(\| x \|_X, \| y \|_Y), \\ \forall (z,t) \in E_X^* \times E_Y^* \;\; \| (z,t) \|_* &= \| z \|_{X,*} + \| t \|_{Y,*} . \end{aligned}$$

Plutôt que de conserver les deux fonctions d_X et d_Y utilisées pour construire la méthode, il est alors plus judicieux, comme dans [27], de les renormaliser pour construire une fonction d mieux adaptée au problème en posant

$$\forall (x,y) \in X \times Y \; d((x,y)) = \frac{1}{\overline{V}_X} d_X(x) + \frac{1}{\overline{V}_Y} d_Y(y),$$

qui est fortement convexe de paramètre $\alpha_0 = \min(\frac{\sigma_X}{\overline{V}_X}, \frac{\sigma_Y}{\overline{V}_Y})$.

Donnons maintenant des bornes pour les déviations de ε_N.

Proposition 3.14 *Supposons que l'hypothèse 3.8 est vérifiée. Alors pour tout* $\alpha > 6\exp\left(-\varkappa\frac{(N+1)}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = \sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{(4M_X+7\sigma_X)}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{(4M_Y+7\sigma_Y)}{\sqrt{N+1}}$$
$$+\left(\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{4\sigma_X}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{4\sigma_Y}{\sqrt{N+1}}\right)\lambda_1^*(N,\alpha)+\left(\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{4\sigma_X}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{4\sigma_Y}{\sqrt{N+1}}\right)\lambda_2^*(N,\alpha),$$

avec

$$\lambda_1^*(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(6\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(6\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N \geq \lambda) \leq \alpha.$$

Proposition 3.15 *Supposons que l'hypothèse 3.9 est vérifiée et que les valeurs* $\beta_{X,i} = \frac{\sigma_X\sqrt{6(N+1)}}{\sqrt{\alpha_{X,0}\overline{V}_X}}$ *et* $\beta_{Y,i} = \frac{\sigma_Y\sqrt{6(N+1)}}{\sqrt{\alpha_{Y,0}\overline{V}_Y}}$ *satisfont* (S). *Alors pour tout* $\alpha > 6\exp\left(-\varkappa\frac{(N+1)}{4}\right)$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = (1+\sqrt{2}+2\sqrt{3})\left(\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{\sigma_X}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{\sigma_Y}{\sqrt{N+1}}\right)$$
$$+\left(\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{4\sigma_X}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{4\sigma_Y}{\sqrt{N+1}}\right)\lambda_1^*(N,\alpha)+(1+2\sqrt{3})\left(\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{\sigma_X}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{\sigma_Y}{\sqrt{N+1}}\right)\lambda_2^*(N,\alpha),$$

avec

$$\lambda_1^*(N,\alpha) = \sqrt{\frac{4}{\varkappa}\ln(6\alpha^{-1})} \text{ et } \lambda_2^*(N,\alpha) = \sqrt{\frac{28}{5\varkappa^2(N+1)}\ln(6\alpha^{-1})},$$

on a

$$\mathbb{P}(\varepsilon_N \geq \lambda) \leq \alpha.$$

3.5 Choix de gain adaptatif

3.5.1 Introduction

Dans le chapitre consacré à l'algorithme de descente en miroir nous avons vu quels étaient les avantages à effectuer le choix de la suite des gains $(\beta_i)_i$ de manière adaptative, à savoir :

- le traitement de données *on line* ;
- la possibilité d'améliorer la vitesse de convergence des algorithmes lorsque la fonction objectif était plus régulière ;
- le fait qu'il n'était plus nécessaire de connaître le niveau du bruit σ pour mettre en oeuvre l'algorithme considéré.

Dans la mesure où l'algorithme que nous étudions dans ce chapitre est déjà capable, lorsque le choix du gain est déterministe, de prendre en compte la régularité de la fonction objectif, les enjeux seront ici plus restreints que dans le chapitre précédent. Néanmoins les deux autres points (implémentation *on line* et non nécessité de connaître σ) justifient à eux seuls l'intérêt de la construction que nous nous proposons d'effectuer.

3.5.2 Construction et bornes pour l'algorithme

Plaçons-nous dans le cadre d'étude de la section 3.2. La suite $(\gamma_i)_i$ sera dans la suite choisie constante et égale à 1 pour tout i. Il nous faut déterminer comment construire la suite $(\beta_i)_i$ à l'aide des informations fournies par l'oracle. Par analogie avec ce qui a été fait pour l'algorithme de descente en miroir, nous supposons que β_0 est une constante fixée (dont nous déterminerons une valeur adéquate ultérieurement) et nous définissons la suite $(\beta_i)_i$ récursivement par

$$\beta_{i+1} = \beta_i + \frac{u_i}{\beta_i} \text{ où } u_i = \frac{1}{2\alpha_0 \overline{V}} \parallel \Psi(x_i) - \Psi(y_i) \parallel_*^2 . \tag{3.12}$$

Cette construction est naturelle au vu de la borne (3.3) et de ce qui a déjà été fait au chapitre précédent : la mise à jour du gain s'effectue en ajoutant le terme quadratique à sa valeur courante. Fixons alors un point $\overline{x} \in Q$ et notons alors pour tout $\beta > 0$ et tout $s \in E^*$, $\mathcal{E}_\beta(s) = (x, y, s_+)$ avec

$$\begin{aligned} x &= T_\beta(\overline{x}, s), \\ y &= T_\beta(x, -\Psi(x)), \\ s_+ &= s - \Psi(y). \end{aligned}$$

L'algorithme obtenu s'écrit (β_0 est supposé connu)

Algorithme 3.4

- *Initialisation :* $s_{-1} = 0 \in E^*$.
- *Pour* $k = 0$ *à* N *faire*
  1. $(x_k, y_k, s_k) = \mathcal{E}_{\beta_k, \gamma_k}(s_{k-1})$,
  2. $\beta_{k+1} = \beta_k + \frac{1}{2\alpha_0 \overline{V} \beta_k} \parallel \Psi(x_k) - \Psi(y_k) \parallel_*^2$,
- *Sortie :*

$$y^N = \frac{\sum_{k=0}^N y_k}{N+1}$$

Le résultat suivant se déduit alors de la borne 3.3.

Proposition 3.16 *Pour tout N, on a*

$$(N+1)\varepsilon_N \leq 2\beta_{N+1}\overline{V} - \sum_{i=0}^N \langle \eta(y_i), y_i - x^* \rangle - \sum_{i=0}^N \frac{\alpha_0 \beta_i}{2} \parallel x_i - y_i \parallel^2 . \tag{3.13}$$

Nous allons maintenant établir une borne en espérance pour ε_N lorsque l'hypothèse 2.2 est vérifiée, le contexte probabiliste étant le même que dans la section 3.2 . Notons que dans ce cas, le second terme du membre de droite de l'équation (3.13) est d'espérance nulle et le troisième est toujours négatif, en conséquence nous allons concentrer nos efforts sur l'étude de β_{N+1}.
Il découle de l'équation de récurrence (3.12) que

$$\beta_{N+1}^2 = \beta_0^2 + 2\sum_{i=0}^{N} u_i + \sum_{i=0}^{N} \frac{u_i^2}{\beta_i^2} \leq \beta_0^2 + 2\sum_{i=0}^{N} u_i + \sum_{i=0}^{N} \frac{u_i^2}{\beta_0^2}.$$

Par conséquent si nous supposons que les variables aléatoires $\eta(x_i)$ ont un moment d'ordre 4 fini, nous obtiendrons une borne pour $\mathbb{E}[\varepsilon_N]$. Plus précisément

Proposition 3.17 *Supposons que l'hypothèse 3.1 qu'il existe une constante $\mu > 0$ finie telle que pour tout $i \geq 0$,*

$$\mathbb{E}[\| \eta(x_i) \|_*^4 |\mathcal{F}_i] \leq \mu^4.$$

Alors pour tout $\beta_0 > 0$, alors

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\overline{V}}{N+1}\sqrt{\beta_0^2 + \frac{8(N+1)}{\alpha_0\overline{V}}(M+\sigma)^2 + \frac{16(N+1)}{\alpha_0^2\overline{V}^2\beta_0^2}(M+\mu)^4}. \tag{3.14}$$

Remarque : La borne (3.14) est en général très grossière. D'une part nous avons majoré β_i^{-1} par β_0^{-1} et cette approximation peut être très mauvaise. D'autre part, nous n'avons pas tenu compte du terme $-\sum_{i=0}^{N} \frac{\alpha_0\beta_i}{2} \| x_i - y_i \|^2$ qui est toujours négatif et contribue donc à réduire l'espérance de ε_N.

Nous allons donc nous donner des hypothèses supplémentaires pour affiner cette borne. En particulier nous supposerons que l'hypothèse 3.3 est vérifiée. Nous noterons dans la suite $u_N^* = \sup_{i=0\cdots N} u_i$.
D'après la proposition 3.16 et le lemme C.3, nous avons

$$(N+1)\varepsilon_N \leq 2\overline{V}\sqrt{\beta_0^2 + \sum_{i=0}^{N} u_i + \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0^2}\sum_{i=0}^{N} u_i\right)} - \sum_{i=0}^{N}\langle \eta(y_i), y_i - x^*\rangle - \sum_{i=0}^{N}\frac{\alpha_0\beta_i}{2} \| x_i - y_i \|^2 . \tag{3.15}$$

Nous supposons à partir de maintenant que l'hypothèse 3.1 est satisfaite. Nous avons donc pour tout $0 \leq i \leq N$:

$$u_i = \frac{1}{2\alpha_0\overline{V}} \| \Psi(x_i) - \Psi(y_i) \|_*^2 \leq \frac{3}{2\alpha_0\overline{V}}(M^2 + \| \eta(x_i) \|_*^2 + \| \eta(y_i) \|_*^2).$$

Lemme 3.3 *Pour tout $N \geq 1$*

$$\mathbb{E}\left[\frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0^2}\sum_{i=0}^{N} u_i\right)\right] \leq \frac{9\sigma^2}{4\varkappa\alpha_0\overline{V}}\left[\ln\left(\left(1 + \frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp\left(\frac{\varkappa M^2}{\sigma^2}\right) + 2e\right)(N+1)\right)\right]^2.$$

Preuve
On a d'une part

$$\begin{aligned}
\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=0}^{N}u_i\right) &\leq \ln\left(1+\frac{3}{2\alpha_0\overline{V}\beta_0^2}(M^2+\parallel\eta(x_i)\parallel_*^2+\parallel\eta(y_i)\parallel_*^2)\right),\\
&\leq \ln\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left[(N+1)\exp\left(\frac{\varkappa M^2}{\sigma^2}\right)\right.\right.\\
&\qquad\left.\left.+\sum_{i=0}^{N}\left(\exp\left(\frac{\varkappa\parallel\eta(x_i)\parallel_*^2}{\sigma^2}\right)+\exp\left(\frac{\varkappa\parallel\eta(y_i)\parallel_*^2}{\sigma^2}\right)\right)\right]\right),
\end{aligned}$$

et d'autre part

$$\begin{aligned}
\frac{u_N^*}{2}=\frac{1}{2}\sup_{i=0\cdots N}u_i &\leq \frac{3}{4\alpha_0\overline{V}}\sup_{i=0\cdots N}(M^2+\parallel\eta(x_i)\parallel_*^2+\parallel\eta(y_i)\parallel_*^2),\\
&\leq \frac{9\sigma^2}{4\varkappa\alpha_0\overline{V}}\ln\left((N+1)\exp\left(\frac{\varkappa M^2}{\sigma^2}\right)+\sum_{i=0}^{N}\left(\exp\left(\frac{\varkappa\parallel\eta(x_i)\parallel_*^2}{\sigma^2}\right)+\exp\left(\frac{\varkappa\parallel\eta(y_i)\parallel_*^2}{\sigma^2}\right)\right)\right).
\end{aligned}$$

Si bien que

$$\begin{aligned}
\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=0}^{N}u_i\right) \leq \frac{9\sigma^2}{4\varkappa\alpha_0\overline{V}}&\left[\ln\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left[(N+1)\exp\left(\frac{\varkappa M^2}{\sigma^2}\right)\right.\right.\\
&\left.\left.\left.+\sum_{i=0}^{N}\left(\exp\left(\frac{\varkappa\parallel\eta(x_i)\parallel_*^2}{\sigma^2}\right)+\exp\left(\frac{\varkappa\parallel\eta(y_i)\parallel_*^2}{\sigma^2}\right)\right)\right]\right)\right]^2.
\end{aligned}$$

Remarquons alors que la fonction $x\to\ln(x)^2$ est concave sur l'ensemble $[e,+\infty[$; par conséquent nous pouvons appliquer l'inégalité de Jensen à l'inégalité précédente dès que $N\geq 1$ ce qui permet d'obtenir le résultat annoncé. □

Donnons maintenant des bornes en espérance pour ε_N.

Proposition 3.18 *Supposons que l'hypothèse 3.1 est vérifiée. Alors pour tout $N\geq 1$ et pour tout $\beta_0>0$*

$$\mathbb{E}[\varepsilon_N]\leq\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{(M+2\sigma)}{\sqrt{N+1}}+\frac{2\overline{V}\beta_0}{N+1}+\frac{18\sigma^2}{M+2\sigma}\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{1}{(N+1)^{\frac{3}{2}}}\ln(N+1)^2+\frac{C}{(N+1)^{\frac{3}{2}}},$$

où C est une constante.

Preuve
En passant à l'espérance dans (3.3), il vient grâce au lemme précédent

$$\begin{aligned}
\mathbb{E}[\varepsilon_N] &\leq \frac{2\overline{V}}{N+1}\mathbb{E}[\beta_N],\\
&\leq \frac{2\overline{V}}{N+1}\sqrt{\beta_0^2+(N+1)\frac{(M+2\sigma)^2}{2\overline{V}\alpha_0}+\frac{9\sigma^2}{4\varkappa\alpha_0\overline{V}}\ln^2\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp(\varkappa\frac{M^2}{\sigma^2})+2e\right)(N+1)\right)}.
\end{aligned}$$

Utilisons maintenant le fait que pour tous $x,\ y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$:

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\overline{V}\beta_0}{N+1} + \sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{(M+2\sigma)}{\sqrt{N+1}} + \tilde{\varepsilon}_N,$$

où

$$\tilde{\varepsilon}_N = \frac{9\sigma^2}{M+2\sigma}\sqrt{\frac{2\overline{V}}{\alpha_0}}\frac{1}{(N+1)^{\frac{3}{2}}}\ln^2\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp(\varkappa\frac{M^2}{\sigma^2})+2e\right)(N+1)\right),$$

dont on déduit immédiatement le résultat annoncé. □

Proposition 3.19 *Supposons que les hypothèses 2.13 et 3.1 sont vérifiées. Alors pour tout $N \geq 1$ et tout $\beta_0 \geq \frac{\sqrt{6}L}{\alpha_0}$*

$$\mathbb{E}[\varepsilon_N] \leq \frac{2\overline{V}\beta_0}{N+1} + \sqrt{\frac{6\overline{V}}{\alpha_0}}\frac{2\sigma}{\sqrt{N+1}} + \sqrt{\frac{\overline{V}}{6\alpha_0}}\frac{9\sigma}{2\varkappa(N+1)^{\frac{3}{2}}}\ln(N+1)^2 + \frac{C}{(N+1)^{\frac{3}{2}}},$$

où C est une constante.

Preuve
Comme l'hypothèse 2.13 est vérifiée, on a

$$\begin{aligned}
2\sum_{i=0}^{N} u_i &\leq \frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}\left(\| \eta(x_i) \|_*^2 + \| f'(x_i) - f'(y_i) \|_*^2 + \| \eta(y_i) \|_*^2\right), \\
&\leq \frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}\left(\| \eta(x_i) \|_*^2 + L^2 \| x_i - y_i \|^2 + \| \eta(y_i) \|_*^2\right).
\end{aligned}$$

L'équation (3.15) devient alors

$$\begin{aligned}
(N+1)\varepsilon_N &\leq -\sum_{i=0}^{N}\langle \eta(y_i), y_i - x^* \rangle - \sum_{i=0}^{N}\frac{\alpha_0\beta_i}{2}\| x_i - y_i \|^2 \\
&\quad +2D\sqrt{\beta_0^2 + \frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}(\| \eta(x_i) \|_*^2 + L^2 \| x_i - y_i \|^2 + \| \eta(y_i) \|_*^2) + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)}, \\
&\leq -\sum_{i=0}^{N}\langle \eta(y_i), y_i - x^* \rangle - \sum_{i=0}^{N}\frac{\alpha_0\beta_i}{2}\| x_i - y_i \|^2 + 2D\sqrt{\beta_0^2 + \frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}L^2 \| x_i - y_i \|^2} \\
&\quad +2\overline{V}\sqrt{\frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}(\| \eta(x_i) \|_*^2 + \| \eta(y_i) \|_*^2) + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)}.
\end{aligned}$$

On utilise alors le fait que pour tous $x,\ y>0$, $\sqrt{x+y}\leq\sqrt{x}+\frac{y}{2\sqrt{x}}$

$$\begin{aligned}(N+1)\varepsilon_N &\leq -\sum_{i=0}^{N}\langle\eta(y_i),y_i-x^*\rangle-\sum_{i=0}^{N}\frac{\alpha_0\beta_i}{2}\parallel x_i-y_i\parallel^2+2\overline{V}\left(\beta_0+\frac{3}{2\alpha_0\overline{V}\beta_0}\sum_{i=0}^{N}L^2\parallel x_i-y_i\parallel^2\right)\\ &\quad+2\overline{V}\sqrt{\frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}(\parallel\eta(x_i)\parallel_*^2+\parallel\eta(y_i)\parallel_*^2)+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)},\\ &\leq 2\overline{V}\sqrt{\frac{3}{\alpha_0\overline{V}}\sum_{i=0}^{N}(\parallel\eta(x_i)\parallel_*^2+\parallel\eta(y_i)\parallel_*^2)+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)}+2\overline{V}\beta_0\\ &\quad-\sum_{i=0}^{N}\langle\eta(y_i),y_i-x^*\rangle,\end{aligned}$$

puisque pour tout i, par construction , $\beta_i\geq\beta_0\geq\frac{\sqrt{6}L}{\alpha_0}$. On utilise alors la concavité de la fonction $x\to\sqrt{x}$ et l'inégalité de Jensen. On obtient

$$\mathbb{E}[\varepsilon_N]\leq\frac{2\overline{V}\beta_0}{N+1}+\frac{2\overline{V}}{N+1}\sqrt{\frac{6(N+1)\sigma^2}{\alpha_0\overline{V}}+\mathbb{E}\left[\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)\right]}.$$

On utilise une nouvelle fois le fait que pour tous $x,\ y>0$, $\sqrt{x+y}\leq\sqrt{x}+\frac{y}{2\sqrt{x}}$ et le lemme 3.3 pour obtenir

$$\mathbb{E}[\varepsilon_N]\leq\frac{2\overline{V}\beta_0}{N+1}+\sqrt{\frac{6\overline{V}}{\alpha_0}}\frac{2\sigma}{\sqrt{N+1}}+\tilde{\varepsilon}_N,$$

où

$$\tilde{\varepsilon}_N=\sqrt{\frac{\overline{V}}{6\alpha_0}}\frac{9\sigma}{4\varkappa(N+1)^{\frac{3}{2}}}\ln^2\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp(\varkappa\frac{M^2}{\sigma^2})+2e\right)(N+1)\right),$$

ce qui achève la preuve . □

Donnons maintenant des bornes en déviation pour ε_N. Nous commençons par un lemme qui nous permettra de borner les déviations des termes u_N^* et $\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)$.

Lemme 3.4 (Inégalité de Markov) *Soit X une variable aléatoire à valeurs positives. Alors pour tout $\lambda>0$ on a*

$$\mathbb{P}(X>\lambda)\leq\frac{\mathbb{E}[X]}{\lambda}.$$

En particulier pour tout $\lambda>0$, on a

$$\mathbb{P}\left(\sum_{i=0}^{N}\left[\exp\left(\frac{\varkappa\parallel\eta(x_i)\parallel_*^2}{\sigma^2}\right)+\exp\left(\frac{\varkappa\parallel\eta(y_i)\parallel_*^2}{\sigma^2}\right)\right]\geq\lambda\right)\leq\frac{2(N+1)e}{\lambda}.$$

Proposition 3.20 *Supposons que les hypothèses 3.1 et 3.3 sont vérifiées. Alors pour tout* $\alpha > 4\exp\left(-\frac{\varkappa(N+1)}{4}\right)$*, tout* $\beta_0 > 0$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = 4\sigma\sqrt{\frac{2\overline{V}\ln(4\alpha^{-1})}{\alpha_0\varkappa(N+1)}} + \sqrt{\frac{2\overline{V}(3M^2+2\sigma^2)}{\alpha_0(N+1)}} + \frac{\sqrt{3}\sigma}{(N+1)^{\frac{3}{4}}}\sqrt{\frac{2\overline{V}}{\varkappa\alpha_0}}\left(\frac{28}{5}\ln(4\alpha^{-1})\right)^{\frac{1}{4}}$$
$$+ \frac{3\sigma}{N+1}\sqrt{\frac{D}{\varkappa\alpha_0}}\ln\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp(\varkappa M^2\sigma^{-2}) + 8e(N+1)\alpha^{-1}\right)(N+1)\right) + \frac{2\overline{V}\beta_0}{N+1},$$

on a

$$\mathbb{P}(\varepsilon_N > \lambda) \leq \alpha.$$

Proposition 3.21 *Supposons que les hypothèses 2.13, 3.1 et 3.3 sont vérifiées. Alors pour tout* $\alpha > 4\exp\left(-\frac{\varkappa(N+1)}{4}\right)$*, tout* $\beta_0 \geq \frac{\sqrt{6}L}{\alpha_0}$ *et tout* $\lambda \geq \lambda^*(N,\alpha)$ *où*

$$\lambda^*(N,\alpha) = 4\sigma\sqrt{\frac{2\overline{V}\ln(4\alpha^{-1})}{\alpha_0\varkappa(N+1)}} + \frac{2\sigma}{\sqrt{N+1}}\sqrt{\frac{6\overline{V}}{\alpha_0}} + \frac{\sqrt{3}\sigma}{(N+1)^{\frac{3}{4}}}\sqrt{\frac{2\overline{V}}{\varkappa\alpha_0}}\left(\frac{28}{5}\ln(4\alpha^{-1})\right)^{\frac{1}{4}}$$
$$+ \frac{3\sigma}{N+1}\sqrt{\frac{\overline{V}}{\varkappa\alpha_0}}\ln\left(\left(1+\frac{3\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\left(\exp(\varkappa M^2\sigma^{-2}) + 8e(N+1)\alpha^{-1}\right)(N+1)\right) + \frac{2\overline{V}\beta_0}{N+1},$$

on a

$$\mathbb{P}(\varepsilon_N > \lambda) \leq \alpha.$$

Chapitre 4

Optimisation stochastique sous contraintes

4.1 Introduction

Les deux chapitres précédents étaient dédiés à l'étude de plusieurs algorithmes destinés à résoudre certaines classes de problèmes variationnels. L'objet de ce chapitre est d'étudier des problèmes d'optimisation stochastique sujets à des contraintes fonctionnelles. Plus précisément, nous allons chercher à résoudre des problèmes de la forme

$$\min_{x\in X} f_0(x) \text{ sous contraintes } f_j(x) \leq 0 \ , \ j = 1, \dots, m \, ; \tag{4.1}$$

où les $f_j(x)$, $j = 0, \dots, m$ sont des fonctions convexes continues et X est un ensemble convexe compact inclus dans un espace vectoriel de dimension n. Nous supposerons dans toute la suite que ce problème a des solutions, c'est-à-dire que l'ensemble $\{x \in X | f_j(x) \leq 0 \ , \ j = 1, \dots, m\}$ est non vide. Comme dans les deux chapitres précédents, nous nous plaçons dans le cas stochastique : les valeurs de tout ou partie des fonctions f_j et de leurs sous-gradients ne sont pas accessibles directement, nous ne recevons que des versions bruitées fournies par un oracle. Les hypothèses (portant sur l'oracle utilisé et sur les fonctions f_j) qui nous serons nécessaires seront introduites au fil de l'étude.
Considérons maintenant la fonction de Lagrange $L(x, \lambda)$ (où $x \in X$ et $\lambda = (\lambda^{(1)}, \cdots, \lambda^{(m)}) \in \mathbb{R}^m_+$) associée au problème (4.1) donnée par

$$L(x, \lambda) = f_0(x) + \sum_{j=1}^{m} \lambda^{(j)} f_j(x).$$

De part sa construction, cette fonction est convexe en x et concave en λ, nous allons donc nous intéresser au problème de point selle

$$\min_{x\in X} \max_{\lambda\in\mathbb{R}^m_+} L(x, \lambda). \tag{4.2}$$

Il est connu (voir par exemple [10]) que si (x^*, λ^*) est un point selle de L, alors x^* est une solution du problème (4.1). Il est alors tentant de vouloir résoudre le problème qui nous intéresse actuellement en utilisant un des algorithmes étudiés dans les chapitres précédents. Malheureusement cette idée se heurte à quelques difficultés qui ont pour origine le fait que nous n'ayons pas de bornes a priori sur les multiplicateurs de Lagrange $(\lambda_j)_{j=1\cdots m}$. En effet, rappelons que les algorithmes que nous avons décrits nécessitent tous pour leur mise en oeuvre une ou plusieurs fonctions proximales (notées V) qui doivent être, entre autres propriétés, fortement convexes. Ceci implique en particulier, dans le cas qui nous interesse, qu'une telle fonction V est non bornée sur $\mathbb{R}^m$. Or la valeur maximale $\overline{V}$ de V intervient à la fois dans nos bornes pour l'erreur commise par les différents algorithmes et dans la plupart des choix de paramètres nécessaires à leur mise en oeuvre. De plus, l'absence de borne pour les λ_j risque de provoquer une "explosion" du niveau du bruit. Plus précisément, nous avons supposé que les valeurs des fonctions f_j et f'_j ne sont pas directement accessibles, mais qu'un oracle nous en transmet des versions bruitées : $f_j(x) + \eta_j(x)$ et $f'_j(x) + \eta'_j(x)$. Considérons maintenant la version stochastique de $\frac{\partial L}{\partial x}$ qui intervient à la fois dans l'étude et dans la mise en oeuvre des algorithmes qui nous intéressent. Avec les notations précédentes,

$$\Psi_X(x, \lambda) = \frac{\partial L}{\partial x}(x, \lambda) + \eta_X(x, \lambda) = f'_0(x) + \eta'_0(x) + \sum_{i=1}^{m} \lambda_j (f_j(x) + \eta_j(x)).$$

Les coefficients λ_j agissent donc multiplicativement sur les variables aléatoires $\eta_j(x)$ et $\eta'_j(x)$, et par conséquent ils agissent de la même façon sur le niveau (l'écart type) du bruit. Comme les coefficients λ_j sont potentiellement grands et que le niveau du bruit intervient directement dans la précision de l'algorithme, celle-ci risque de diminuer fortement. Les inconvénients liés à la recherche du point selle de la fonction L sont donc nombreux et à la fois d'ordre théorique et pratique. Nous allons donc chercher à mettre en oeuvre d'autres méthodes pour résoudre le problème (4.1).
Ce chapitre est organisé de la manière suivante. Nous allons dans un premier temps réduire le problème (4.1) à un nouveau problème unidimensionnel et nous envisagerons deux algorithmes pour le résoudre. Nous proposerons ensuite un algorithme pour résoudre le problème initial dont nous étudierons la précision et la complexité. Enfin nous traiterons en guise d'exemple le problème de Markowitz.

4.2 Réduction à un problème unidimensionnel

L'objet de cette section est de réduire le problème (4.1) à un problème de dimension 1, à savoir chercher un zéro d'une fonction f^* de $\mathbb{R}$ dans $\mathbb{R}$. Nous commençons par définir et étudier la fonction f^* en question. Nous proposons ensuite deux algorithmes pour résoudre le nouveau problème obtenu : une méthode de dichotomie et une variante de l'algorithme de Newton. Enfin nous faisons le lien avec les problèmes de point selle précédemment étudiés.

4.2.1 Définition et propriétés de la fonction f^*

Oublions dans un premier temps que nous avons un problème stochastique et concentrons-nous sur le problème d'optimisation sous contraintes sous-jacent. Notons

$$f(t,x) = \max\{f_0(x) - t,\ f_i(x)\,;\ i = 1 \dots m\}$$

où $x \in X$ et $t \in \mathbb{R}$ et considérons la fonction

$$f^*(t) = \min_{x \in X} f(t,x).$$

Nous commençons par donner quelques propriétés élémentaires de la fonction f^* qui vont nous permettre de reformuler le problème (4.1) en utilisant uniquement la fonction f^*.

Lemme 4.1 *Soit t^* la valeur optimale du problème (4.1). Alors*

$$\begin{array}{rcl} f^*(t) & \leq & 0 \ si\ t \geq t^* \\ f^*(t) & > & 0 \ si\ t < t^*. \end{array}$$

Preuve
Soit $t \geq t^*$. Pour tout $x \in X$: $f(t,x) \leq f(t^*,x)$, donc $f^*(t) \leq f(t^*,x)$. Si x^* est solution de (4.1) on a $f(t^*,x^*) \leq 0$ et $f^*(t) \leq 0$.
Supposons que $t < t^*$ et que $f^*(t) \leq 0$. Alors il existe $x \in X$ tel que $f_0(x) \leq t < t^*$ et que $f_j(x) \leq 0$, $i = 1 \dots m$, ce qui contredit le fait que t^* soit la valeur optimale du problème. □

Ainsi, résoudre (4.1) est équivalent à chercher un zéro de la fonction f^* (comme nous allons le voir par la suite, f^* est convexe et le point qui nous intéresse est la borne inférieure de l'ensemble

des zéros de f^*). Cette dernière remarque est fondamentale pour la suite : elle nous permet de ramener le problème (4.1) à un problème 'simple' unidimensionnel.
Cette remarque étant faite, nous donnons maintenant des propriétés de la fonction f^* qui vont nous permettre de mettre en œuvre des algorithmes adéquats pour résoudre le problème qui nous intéresse. Les deux propositions suivantes sont empruntées à [28].

Proposition 4.1 *Pour tout $\Delta \geq 0$ et $t \in \mathbb{R}$*

$$f^*(t) - \Delta \leq f^*(t + \Delta) \leq f^*(t).$$

La fonction f^* est donc décroissante et lipschitzienne de constante 1.

Preuve
On a

$$\begin{aligned} f^*(t+\Delta) &= \min_{x\in X} \max_{1\leq i\leq m} \{f_0(x) - t - \Delta, f_j(x)\}, \\ &\leq \min_{x\in X} \max_{1\leq j\leq m} \{f_0(x) - t, f_j(x)\} = f^*(t). \end{aligned}$$

et

$$\begin{aligned} f^*(t+\Delta) &= \min_{x\in X} \max_{1\leq j\leq m} \{f_0(x) - t, f_j(x) + \Delta\} - \Delta, \\ &\geq \min_{x\in X} \max_{1\leq j\leq m} \{f_0(x) - t, f_j(x)\} - \Delta = f^*(t) - \Delta. \end{aligned}$$

□

Proposition 4.2 *La fonction f^* est convexe.*

Preuve
Soit $t_0, t_2 \in \mathbb{R}$ et $\alpha \in [0, 1]$. On note $t_1 = (1-\alpha)t_0 + \alpha t_2$ et $x_\alpha = (1-\alpha)x^*(t_0) + \alpha x^*(t_2)$, où $x^*(t_i)$ est solution de $f^*(t_i) = f(t_i, x^*(t_i))$. Alors

$$\begin{aligned} f^*(t_1) &\leq \max_{1\leq j\leq m} \{f_0(x_\alpha) - t_1, f_j(x_\alpha)\} \\ &\leq \max_{1\leq j\leq m} \{(1-\alpha)(f_0(x^*(t_0)) - t_0) + \alpha(f_0(x^*(t_2)) - t_2), (1-\alpha)f_j(x^*(t_0) + \alpha f_j(x^*(t_2)\} \\ &\leq (1-\alpha) \max_{1\leq j\leq m} \{f_0(x^*(t_0)) - t_0, f_j(x^*(t_0))\} + \alpha \max_{1\leq j\leq m} \{f_0(x^*(t_2)) - t_2, f_j(x^*(t_2))\} \\ &\leq (1-\alpha)f^*(t_0) + \alpha f^*(t_2). \end{aligned}$$

□

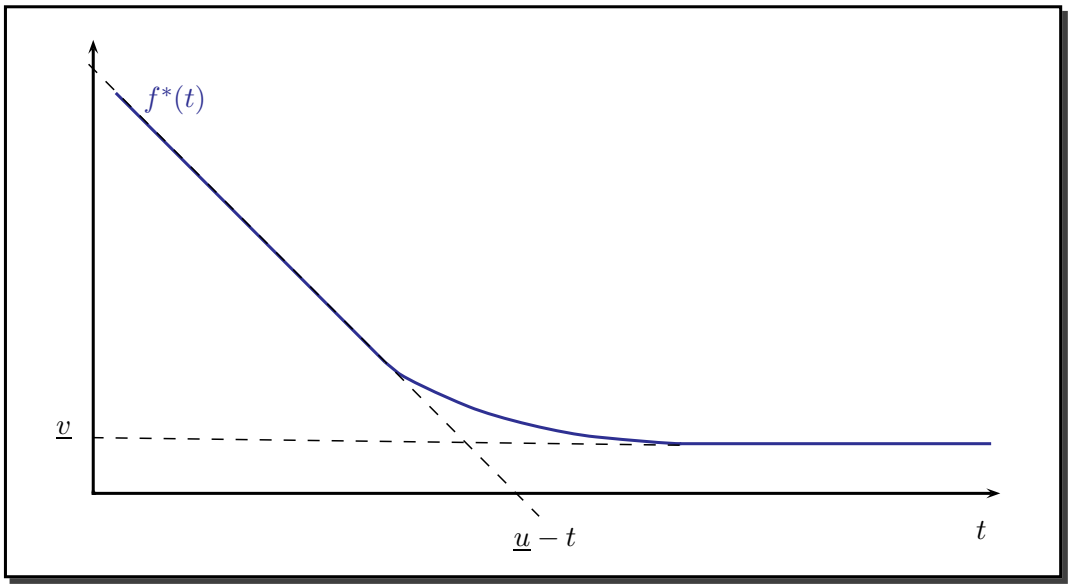
Les deux propositions précédentes résument les propriétés qui nous seront utiles pour résoudre notre problème. Néanmoins il est possible d'affiner l'étude de la fonction f^* afin d'en établir un "profil type". Plus précisément, remarquons que la fonction f^* est de la forme $\min_{x\in X} \max(u(x) - t, v(x))$ où $u = f_0$ et $v = \max_{1\leq j\leq m} f_j$ sont des fonctions convexes et continues. Comme X est compact, il existe des constantes $\underline{u}$, $\overline{u}$, $\underline{v}$ et $\overline{v}$ telles que pour tout $x \in X$

$$\underline{u} \leq u(x) \leq \overline{u},\ \underline{v} \leq v(x) \leq \overline{v},$$

et chacune de ces bornes est atteinte sur X. Alors il est clair que

– si $-t \geq \overline{v} - \underline{u}$, $f^*(t) = \underline{u} - t$,
– si $t \geq \overline{u} - \underline{v}$, $f^*(t) = \underline{v}$.

Il en découle en particulier que la valeur 1 est la meilleure constante de Lipschitz possible pour f^*. Le dessin ci-dessous donne un exemple de fonction f^* ayant toutes les propriétés que nous avons énoncées :



Nous avons vu que la fonction f^* est continue, car lipschitzienne. Nous ne pourrons en général rien affirmer de plus sur sa régularité, comme l'illustre l'exemple suivant.
Plaçons-nous dans $\mathbb{R}^m$ et considérons le problème

$$\min_{x \in X} f_0(x) \text{ sous contraintes } f_j(x) \leq 0 \ , \ j = 1, \ldots, m \, ;$$

où X est la boule euclidienne de $\mathbb{R}^m$ de centre 0 et de rayon 1 : $X = \{x \in \mathbb{R}^m | \parallel x \parallel_2 \leq 1\}$, $f_0(x) = \parallel x \parallel_2^2$ et pour tout $j = 1 \cdots m$, $f_j(x) = \langle x, e_j \rangle$ où e_j désigne le $j$ème vecteur de la base canonique de $\mathbb{R}^m$. Alors la fonction f^* est donnée par

$$\begin{array}{ll} \text{si } t \leq 0, & f^*(t) = -t, \\ \text{si } 0 \leq t \leq 1 + \frac{1}{\sqrt{m}}, & f^*(t) = \frac{1}{2}\left(\frac{-1}{\sqrt{m}} + \sqrt{\frac{1}{m} + 4t}\right) - t, \\ \text{si } t \geq 1 + \frac{1}{\sqrt{m}}, & f^*(t) = \frac{-1}{\sqrt{m}}. \end{array}$$

La fonction f^* obtenue est bien continue, mais pas dérivable.

Nous allons maintenant proposer deux algorithmes de recherche du plus petit zéro t^* de f^*. Rappelons-nous que notre objectif est de résoudre un problème stochastique, par conséquent nous devons prendre en compte le fait que ni les valeurs de f^*, ni celles de son sous-gradient ne sont connues de manière exacte a priori.

4.2.2 Résolution à l'aide d'une variante de l'algorithme de Newton

Le premier algorithme que nous proposons ici est une adaptation de l'algorithme de Newton classique à notre problème. Cette version de l'algorithme de Newton est inspirée des algorithmes donnés dans le chapitre 2 de [28] que nous avons repris et modifiés pour exploiter au mieux les différentes propriétés du problème que nous étudions. L'algorithme proposé est le suivant :

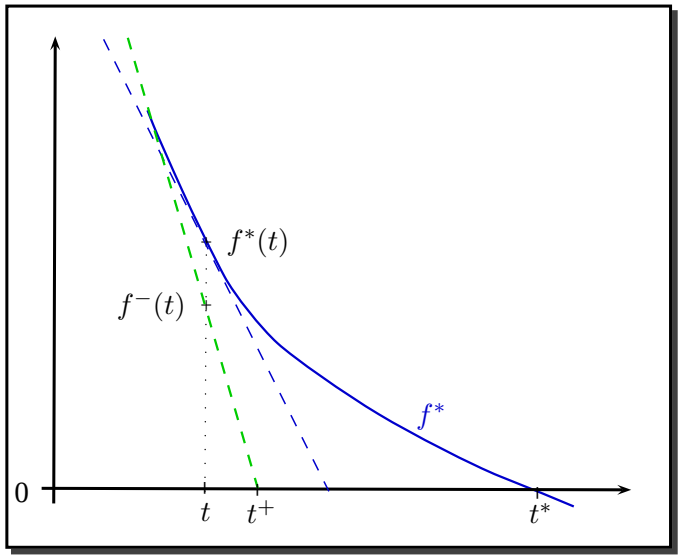
Algorithme 4.1

- *Initialisation : choisir* $t_0 < t^*$, $\kappa \in]0, \frac{1}{2}[$ *et* $\epsilon > 0$.
- *Pas* k *(*$k \geq 0$*) :*
  1. *Trouver deux nombres* $f^+(t_k)$ *et* $f^-(t_k)$ *tels que :*

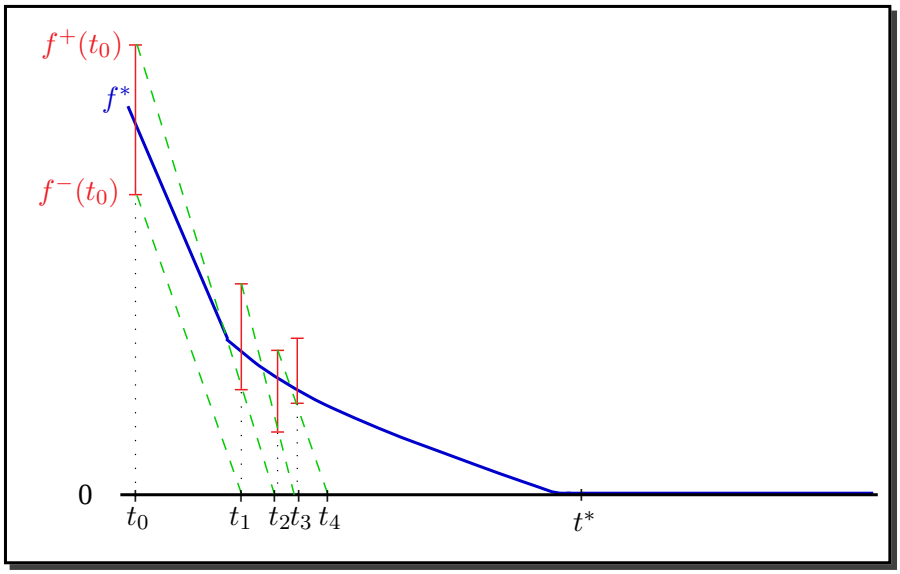
     $$f^+(t_k) \geq f^*(t_k) \geq f^-(t_k) \text{ et } f^-(t_k) \geq (1-\kappa) f^+(t_k).$$

     Si $f^+(t_k) \leq \epsilon$*, fin de l'algorithme.*
  2. $t_{k+1} = \begin{cases} t_0 + f^-(t_0) & \text{si } k = 0, \\ t_k + f^-(t_k) \frac{t_k - t_{k-1}}{f^+(t_{k-1}) - f^-(t_k)} & \text{sinon.} \end{cases}$

Cet algorithme cherche à imiter l'algorithme de Newton classique (pour lequel la construction d'un nouveau point de recherche à partir du point courant t est donné par la droite en pointillés bleus dans le dessin ci-après), mais puisque les valeurs de $f^*(t)$ et de son sous-gradient $\partial f^*(t)$ en un point t donné ne sont pas connues, quelques adaptations sont nécessaires. Nous allons conserver le schéma général de l'algorithme, mais il faut trouver des valeurs adéquates à substituer à $f^*(t)$ et $\partial f^*(t)$. Notre objectif étant de trouver le plus petit zéro t^* de la fonction f^*, nous voudrions construire une suite de points de recherche $(t_k)_k$ qui soit croissante et majorée par t^*. Ces conditions peuvent être réalisées si nous savons calculer un minorant de f^* strictement positif et un minorant de son sous-gradient en un point $t < t^*$ donné. En effet, pour un tel point t, si $f^-(t)$ est un minorant strictement positif de $f^*(t)$ et $p(t)$ un minorant de $\partial f^*(t)$, alors le zéro t^+ de la droite passant par le point $(t, f^-(t))$ et de pente $p(t)$ (droite en pointillés verts sur le dessin) vérifie , puisque f^* est convexe et décroissante, $t \leq t^+ \leq t^*$.



Dans l'algorithme proposé, nous effectuons une construction de ce type. Plus précisément, en chaque point de recherche t_k, pour calculer une valeur de $p(t_k)$ adéquate, l'algorithme procède ainsi : si $k = 0$, puisque f^* est lipschitzienne de constante 1, nous pouvons borner inférieurement son sous-gradient au point t_0 par $p(t_0) = -1$; si $k \geq 1$, nous pouvons minorer la valeur du sous-gradient de f^* en t_k par le coefficient directeur de la droite contenant les points $(t_{k-1}, f^+(t_{k-1}))$ et $(t_k, f^-(t_k))$.



La proposition suivante évalue la complexité de l'algorithme 4.1

Proposition 4.3 *L'algorithme 4.1 s'arrête après au plus $N(\epsilon)$ iterations où*

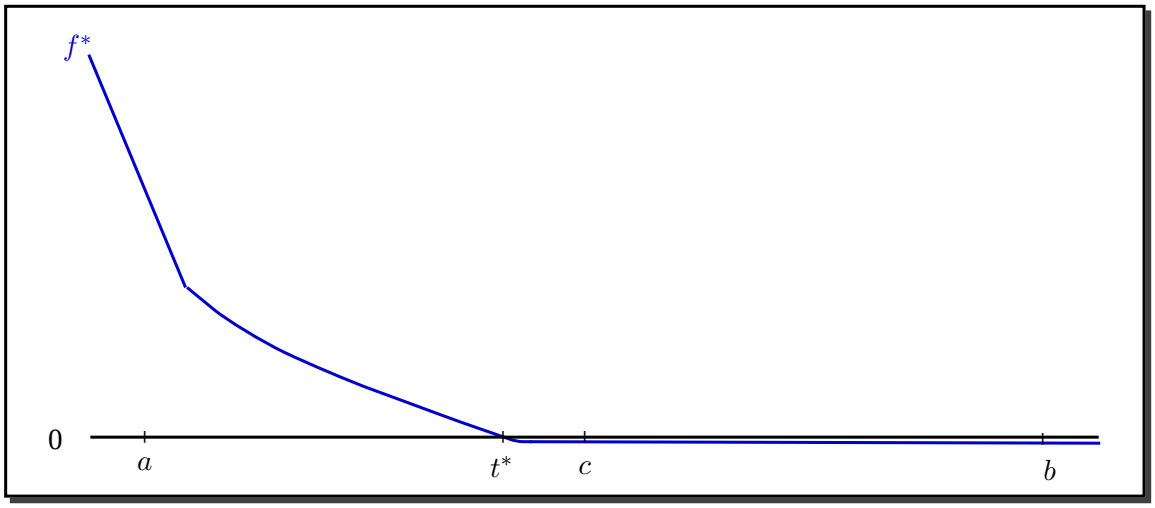
$$N(\epsilon) = \frac{1}{\ln(2(1-\kappa))} \ln \frac{t^* - t_0}{(1-\kappa)\epsilon}.$$

Preuve
Voir l'annexe C.

4.2.3 Résolution par dichotomie

Afin d'offrir une alternative à l'algorithme 4.1, nous proposons maintenant un second algorithme pour résoudre notre problème de recherche de zéro pour f^*. Comme pour l'algorithme de Newton, le contexte stochastique dans lequel nous nous trouvons va nécessiter d'adapter quelque peu la version usuelle de l'algorithme de dichotomie à notre problème. Plus précisément :

- L'évaluation de la fonction f^* aux différents points de recherche est remplacée par une estimation de f^*, puisque nous ne connaissons pas les valeurs des différentes fonctions f_j de manière exacte.
- La seconde modification concerne le critère d'arrêt : comme nous ne connaissons pas $\underline{v}$, il n'est pas raisonnable de stopper l'algorithme lorsque nous trouvons un point t tel que $f^*(t)$ est proche de 0 (comme l'illustre le dessin ci-après). Par conséquent, nous arrêterons l'algorithme lorsque l'amplitude de l'intervalle de recherche sera égale ou inférieure à la précision ϵ souhaitée : comme f^* est lipschitzienne de constante 1, tout point t de l'intervalle de recherche courant vérifiera lors $|f^*(t)| \leq \epsilon$.



L'algorithme proposé est le suivant :

Algorithme 4.2

- *Initialisation : choisir* $a_0 < t^*$, $b_0 > t^*$ *et* $\epsilon > 0$.
- $a = a_0$, $b = b_0$.
- *Tant que* $b - a > \epsilon$ *faire*
  - *construire un estimateur* $\hat{f}(c)$ *de* $f^*(c)$ *où* $c = \frac{a+b}{2}$,
  - *si* $\hat{f}(c) \geq \epsilon$ *faire* $a = c$,
  - *si* $\hat{f}(c) < \epsilon$ *faire* $b = c$.
- *Sortie* $\frac{a+b}{2}$.

Le principal problème soulevé par cet algorithme est de trouver un estimateur de $f^*(c)$ qui soit suffisamment précis pour nous permettre de déterminer si $f^*(c) > 0$ ou non avec une probabilité d'erreur fixée à l'avance. Les deux lignes

– si $\hat{f}(c) \geq \epsilon$ faire $a = c$,
– si $\hat{f}(c) < \epsilon$ faire $b = c$,

doivent donc être comprises comme un test d'hypothèse. Il faudra donc veiller à choisir convenablement $\hat{f}(c)$ pour contrôler efficacement les risques d'erreur.
Donnons maintenant la complexité de l'algorithme 4.2.

Proposition 4.4 *L'algorithme 4.2 termine après au plus $N(\epsilon)$ itérations où $N(\epsilon)$ est le plus petit entier supérieur ou égal à*

$$\frac{1}{\ln(2)} \ln\left(\frac{b_0 - a_0}{\epsilon}\right).$$

4.2.4 Remarques et comparaison des deux algorithmes

Le but de ce paragraphe est de donner quelques indications d'ordre technique concernant la mise en oeuvre des deux algorithmes que nous avons proposés pour résoudre le problème de recherche de zéro pour la fonction f^* ; puis de les comparer, d'une part en termes de complexité algorithmique et d'autre part en termes de robustesse aux erreurs éventuelles, afin de justifier le fait que nous privilégierons l'algorithme 4.1 dans la suite.

Initialisation des algorithmes

Le premier point concerne le choix de t_0 pour l'algorithme 4.1 et de a_0 et b_0 pour l'algorithme 4.2. Il s'agit donc de trouver une borne inférieure et une borne supérieure pour la solution du problème (4.1). Il est bien entendu envisageable d'avoir des valeurs "naturelles" pour ces constantes au vu du problème concret étudié. Notre propos ici est de donner une méthode pour construire des valeurs admissibles pour t_0, a_0 et b_0 lorsque la modélisation n'en fournit pas *a priori*.
Une borne inférieure peut être aisément obtenue en cherchant la solution du problème sans contrainte $\min_{x \in X} f_0(x)$, que nous savons résoudre à l'aide de l'algorithme 2.5. une fois cette borne inférieure connue, il suffit de remarquer que $\max_{x \in X} f_0(x) \leq \min_{x \in X} f_0(x) + L_0 D_X$ où L_0 désigne la constante de Lipschitz de la fonction f_0 (que nous supposerons connue) et D_X le diamètre de l'ensemble X pour obtenir une valeur convenable pour b_0.

Complexité

Les propositions 4.3 et 4.4 nous donnent le nombre d'itérations à effectuer en pire cas dans les algorithmes 4.1 et 4.2 respectivement. Dans les deux cas ce nombre d'itérations décroît comme $\ln\left(\frac{err_0}{\epsilon}\right)$ où err_0 est le manque de précision initial pour t^*, à savoir $t^* - t_0$ pour l'algorithme 4.1 et $b_0 - a_0$ pour l'algorithme 4.2. Le nombre d'itérations à effectuer est donc du même ordre de grandeur pour chacun des deux algorithmes. Le second point qui doit être pris en compte pour déterminer la complexité algorithmique de chacun des deux algorithmes est la "taille" des estimations : nous avons d'une part déjà mentionné le fait que les valeurs f^+ et f^- nécessaires pour le bon déroulement de l'algorithme 4.1 sont les bornes d'un intervalle de confiance pour f^* et d'autre part qu'il faudrait savoir contrôler les risques d'erreurs lors du test d'hypothèse qui intervient à chaque pas de l'algorithme 4.2. Or il est connu qu'en statistiques effectuer un test d'hypothèse ou construire un intervalle de confiance sont des problèmes très proches : il faut donc s'attendre à ce que l'effort, en termes de calculs, pour résoudre chacun de ces deux problèmes soit similaires. Par conséquent, les deux algorithmes 4.1 et 4.2 ont des temps de calculs théoriques en pire cas comparables. Nous

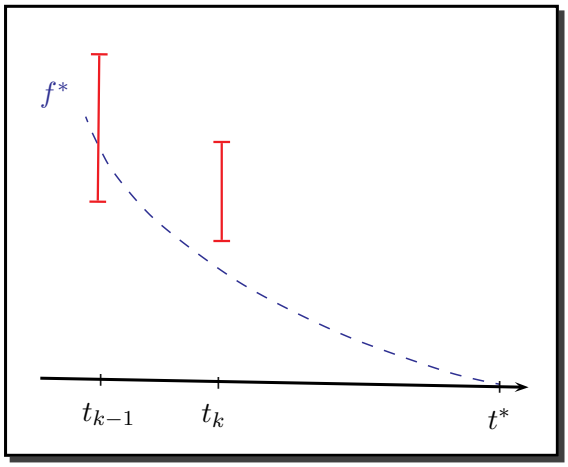
verrons cependant plus loin que dans la pratique, nous pouvons nous attendre à ce que le temps de calcul de l'algorithme 4.1 soit bien inférieur à la borne théorique en pire cas et par conséquent soit aussi inférieur à celui de l'algorithme 4.2.

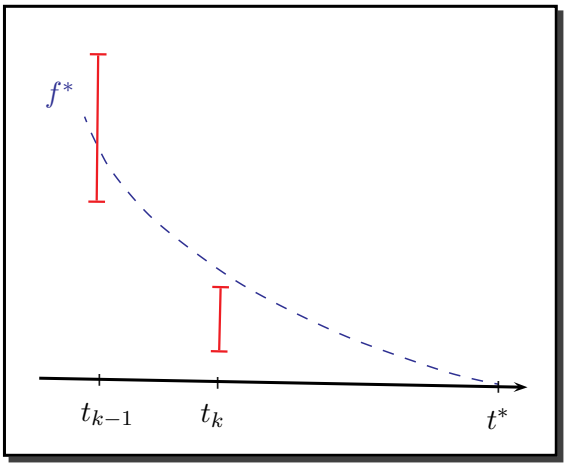
Robustesse aux erreurs

Les algorithmes 4.1 et 4.2 sont tous les deux itératifs et nécessitent à chaque étape une construction faisant appel au hasard :

- dans le cas de l'algorithme 4.1 il s'agit de construire un intervalle de confiance pour f^*,
- dans le cas de l'algorithme 4.2 il s'agit de choisir un nouveau point de recherche en fonction du résultat d'un test d'hypothèses.

Même avec des niveaux de confiance raisonnables, nous aurons de fortes chances d'observer des erreurs dans ces constructions au cours de l'exécution de l'algorithme que nous aurons choisi : par exemple avec un niveau de confiance de 99%, il se produira en moyenne deux erreurs si nous effectuons 200 itérations. Notre propos ici est de déterminer s'il est possible de détecter et, si il y a lieu, de corriger ces erreurs pour chacun des deux algorithmes 4.1 et 4.2.
Commençons par l'algorithme 4.1. Pour cet algorithme, une erreur se produit lorsque, au pas k, $f^*(t_k)$ n'appartient pas à l'intervalle de confiance $[f^-(t_k), f^+(t_k)]$ que nous avons construit. Afin de fixer les idées, nous supposerons qu'une seule erreur a lieu et qu'elle se produit au pas k où $k \neq 0$. Nous sommes alors dans une des deux situations suivantes





Commençons par traiter le premier cas, c'est-à-dire celui où $f^-(t_k) > f^*(t_k)$. Alors

- soit $f^-(t_k) \geq f^+(t_{k-1}))$, auquel cas nous pouvons immédiatement détecter la présence d'une erreur (de telles valeurs contredisent la décroissance de f^*) et nous pouvons la corriger en construisant un nouvel intervalle de confiance pour $f^*(t_k)$ et $f^*(t_{k-1}))$,
- soit $f^-(t_k) < f^+(t_{k-1}))$ auquel cas nous ne pouvons pas *a priori* détecter l'erreur et nous construisons un nouveau point de recherche t_{k+1} :
  - si $t_{k+1} \leq t^*$ (ce qui signifie que $f^-(t_k)$ est proche de $f^*(t_k)$ et sera le cas le plus courant), l'erreur commise est sans conséquence sur le déroulement ultérieur de l'algorithme,
  - sinon, en général l'algorithme s'arrête et renvoie t_{k+1} (solution incorrecte) ; il est encore possible de détecter la présence d'une erreur si nous obtenons une valeur $f^*(t_{k+1}) < 0$ puisque par construction tous les points t_j construits par l'algorithme doivent satisfaire $t_j \leq t^*$.

Venons-en maintenant au second cas, c'est-à-dire au cas où $f^+(t_k) < f^*(t_k)$. Alors l'erreur n'est généralement pas détectable avant de construire un intervalle de confiance pour le point de recherche suivant t_{k+1} (qui dans ce cas vérifie toujours $t_{k+1} < t^*$), cette construction ayant lieu pourvu que $f^+(t_k) > \epsilon$ (dans le cas contraire l'algorithme s'arrête prématurément et renvoie la solution incorrecte t_k). Alors

- si $f^+(t_k) \leq f^-(t_{k+1})$, l'erreur est détectée puisque ces valeurs contredisent la décroissante de f^* et il est possible de la corriger en calculant un nouvel intervalle de confiance pour $f^*(t_k)$ et $f^*(t_{k+1})$,
- sinon le point t_{k+2} est construit et
  - si $t_{k+2} \leq t^*$ l'erreur commise est sans conséquence sur le bon déroulement ultérieur de l'algorithme,
  - si $t_{k+2} > t^*$ l'algorithme s'arrête et renvoie la réponse incorrecte t_{k+2}.

En conclusion, lorsque l'algorithme 4.1 est utilisé, certaines erreurs commises lors de la construction des intervalles de confiance peuvent être corrigées et détectées.
En revanche, lorsque l'algorithme 4.2 est utilisé, toute conclusion erronée quant au test d'hypothèse mis en œuvre a pour conséquence que les intervalles de recherche suivants ne contiendront plus t^* : l'erreur n'est pas détectable et ne peut pas non plus être corrigée. En conséquence, nous utiliserons l'algorithme 4.1, qui est plus robuste aux erreurs éventuelles, dans la suite.

4.2.5 Lien avec les problèmes de point-selle stochastiques

L'algorithme 4.1 nous a permis de résoudre le problème de recherche de zéro de f^* pourvu que nous sachions comment trouver deux valeurs f^+ et f^- adéquates à chaque pas. C'est sur ce nouveau problème : comment calculer une borne supérieure $f^+(t)$ et une borne inférieure $f^-(t)$ pour $f^*(t)$ que nous nous penchons maintenant.
Supposons donc fixée un réel t et rappelons-nous que

$$f^*(t) = \min_{x \in X} \max_{j=0,\dots,m} f_{j,t}(x) \text{ où } f_{0,t}(x) = f_0(x) - t,\ f_{j,t}(x) = f_j(x) \text{ si } j > 0.$$

Notons $f_{(t)}(x)$ le vecteur $(f_{0,t}(x), \dots, f_{m,t}(x))^T$ et considérons la fonction $L_t(x, y)$ donnée par

$$L_t(x, y) = \sum_{j=0}^{m} y_j f_{j,t}(x) = y^T f_{(t)}(x) \tag{4.3}$$

où $y \in \Delta = \{y \in \mathbb{R}^{m+1} | y_j \geq 0, \sum_{j=0}^{m} y_j = 1\}$.
Remarquons que la fonction $L_t(x, y)$ est concave par rapport à la variable y et convexe par rapport à la variable x et que

Proposition 4.5

$$f^*(t) = \min_{x \in X} \max_{1=1,\dots,m} f_{i,t}(x) = \min_{x \in X} \max_{y \in \Delta} L_t(x, y).$$

Trouver une valeur approchée pour $f^*(t)$ est donc équivalent à résoudre le problème de recherche de point selle pour la fonction L_t, ce que nous savons faire en utilisant l'algorithme 2.5. Nous allons donc utiliser l'algorithme 2.5 pour calculer une valeur approchée pour $f^*(t)$ et utiliser ensuite les bornes en déviations obtenues dans le chapitre 2 pour l'erreur d'approximation faite dans cet algorithme afin d'en déduire un encadrement de $f^*(t)$. La borne inférieure et la borne supérieure ainsi obtenues joueront alors respectivement les rôles de f^- et f^+ dans l'algorithme 4.1.
Ainsi à chaque pas de l'algorithme 4.1, nous avons à résoudre un problème de point-selle, et résoudre le problème (4.1) est équivalent à résoudre une suite de problèmes de point selle stochastiques.

4.3 Algorithme

4.3.1 Introduction - notations

Revenons maintenant à notre problème initial. Pour rappel, nous voulons résoudre le problème d'optimisation suivant

$$\min_{x\in X} f_0(x) \text{ sous contraintes } f_j(x) \leq 0 \ , \ j = 1, \ldots, m \, ;$$

où les $f_j(x)$, $j = 0, \ldots, m$ sont des fonctions convexes, X est un ensemble convexe compact inclus dans l'espace vectoriel de dimension n E_X muni de la norme $\| \cdot \|_X$. Nous n'avons accès qu'à un oracle stochastique du premier ordre qui nous fournit des variables aléatoires indépendantes $\phi_j(x)$, $j = 0, \cdots, m$ et $\psi_j(x)$, $j = 0, \cdots, m$, telles que

$$\mathbb{E}\left[\psi_j(x)\right] = f_j(x), \ \mathbb{E}\left[\phi_j(x)\right] = f_j'(x),$$

où $f_j'(x)$ désigne un élément quelconque du sous-différentiel $\partial f_j(x)$ de f_j au point x.
Dans ce qui suit nous supposerons que le problème considéré possède des solutions, c'est-à-dire que l'ensemble $\{x \in X | f_j(x) \leq 0, \ j = 1 \cdots m\}$ est non vide. Dans la première partie de ce chapitre, nous avons expliqué comment décomposer le problème (4.1) en une suite de problèmes de points selle. L'algorithme 4.1 va nous fournir le schéma général, et l'algorithme 2.5 (avec quelques modifications que nous allons détailler) et les inégalités de déviations que nous avons obtenues, nous permettrons alors de calculer des valeurs adéquates pour f^+ et f^-. Nous voulons donc maintenant résoudre le problème

$$\min_{x\in X} \max_{j=0\cdots m} f_{j,t}(x) = \min_{x} \max_{y\in\Delta} L_t(x, y),$$

où $\Delta = \left\{ y \in \mathbb{R}^{m+1} \,|\, y \geq 0 \text{ et } \sum_{j=1}^{m+1} y^{(j)} = 1 \right\}$. Notons $F_t(x)$ la fonction convexe $\max_{y\in\Delta} L_t(x, y)$. A cause de la forme très particulière de la fonction L_t (voir (4.3)) nous avons seulement besoin de connaître les valeurs prises par les fonctions $f_{j,t}$ au point x pour calculer $F_t(x)$. Nous avons donc $Y = \Delta_1$, $\| \cdot \|_Y = \| \cdot \|_1$ et $\| \cdot \|_{Y,*} = \| \cdot \|_\infty$. L'algorithme que nous allons construire utilisera des versions stochastiques des sous/sur-différentiels de la fonction L_t, à savoir

$$\begin{aligned}
\Phi_t(x, y) &= \sum_{i=0}^{m} y^{(j)} \phi_j(x, y) = \frac{\partial L_t}{\partial x}(x, y) + \eta_X(x, y), \\
\Psi_t(x, y) &= (\psi_0(x, y) - t, \psi_1(x, y), \cdots, \psi_m(x, y)) = \frac{\partial L_t}{\partial y}(x, y) + \eta_Y(x, y).
\end{aligned}$$

Afin de mettre en oeuvre l'algorithme que nous avons esquissé, il reste à discuter les deux points suivants :

- A chaque pas de l'algorithme 4.1, nous avons besoin d'un intervalle de confiance pour $f^*(t_k) = F_{t_k}(x^*(t_k))$ où $x^*(t_k)$ est un élément quelconque de X tel que l'égalité $f^*(t_k) = F_{t_k}(x^*(t_k))$ ait lieu. Nous devons donc trouver comment le construire en utilisant l'algorithme 2.5 et les bornes en déviations pour l'erreur.
- Si nous choisissons de prendre (β_i) et (γ_i) constantes dans l'algorithme 2.5 (ce qui sera le cas dans la suite), il faudra connaître à l'avance combien de pas effectuer ce même algorithme et par conséquent définir une stratégie de choix pour le nombre de pas à effectuer.

Ces deux points font l'objet des deux paragraphes qui suivent.

4.3.2 Inégalités de déviations, compléments

Commençons par traiter le premier des deux points évoqués au paragraphe précédent, c'est à dire construire un intervalle de confiance pour $f^*(t) = F_t(x^*(t))$ où $x^*(t)$ désigne un point de l'ensemble X tel que l'égalité précédente ait lieu. Nous nous plaçons dans le cas où les hypothèses 2.12 et 2.10 sont satisfaites.
L'algorithme 2.5 nous permet obtenir une valeur approchée x^N pour $x^*(t)$, mais comme notre oracle est stochastique, nous n'avons pas directement accès à $F_t(x^N)$. Nous allons donc plutôt essayer d'approcher $f^*(t)$ par

$$S_N = \max_{y\in\Delta} \frac{1}{N}\sum_{i=1}^{N} L(x_{i-1}, y) = \frac{1}{N}\max_{j=0\cdots m}\sum_{i=1}^{N} f_{j,t}(x_{i-1}),$$

et plus précisément considérer sa version stochastique

$$\hat{S}_N = \frac{1}{N}\max_{j=0\cdots m}\sum_{i=1}^{N} \psi_{j,t}(x_{i-1})$$

avec $\psi_{0,t} = \psi_0 - t$ et $\psi_{j,t} = \psi_j$ si $j > 0$. Comme $\hat{S}_N$ ne dépend que de t et des informations fournies par l'oracle, cet estimateur de $f^*(t)$ peut être construit au cours de l'exécution de l'algorithme 2.5 et va par conséquent nous servir de point de départ pour construire un intervalle de confiance pour $f^*(t)$.
Remarquons maintenant que

$$S_N = \frac{1}{N}\max_{y\in\Delta}\sum_{i=1}^{N} L_t(x_{i-1}, y) \geq \frac{1}{N}\sum_{i=1}^{N} L_t(x_{i-1}, y^*) \geq \min_{x\in X} L_t(x, y^*) = F_t(x^*(t)) = f^*(t),$$

et par conséquent $S_N - f^*(t) \geq 0$. Ainsi, si nous trouvons deux réels positifs $\lambda(N,\alpha)$ et $\lambda'(N,\alpha)$ tels que, $\alpha > 0$ étant donné, nous ayons :

- d'une part $S_N - f^*(t) \leq \lambda(N,\alpha)$ avec une probabilité au moins égale à $1 - \frac{\alpha}{2}$,
- d'autre part $|\hat{S}_N - S_N| \leq \lambda'(N,\alpha)$ avec une probabilité au moins égale à $1 - \frac{\alpha}{2}$,

alors $f^*(t)$ appartiendra à l'intervalle

$$[\hat{S}_N - \lambda(N,\alpha) - \lambda'(N,\alpha), \hat{S}_N + \lambda'(N,\alpha)]$$

avec une probabilité au moins égale à $1 - \alpha$,et nous aurons construit notre intervalle de confiance $[f^-(t), f^+(t)]$ pour $f^*(t)$.
Commençons par traiter le terme $S_N - f^*(t)$. Nous avons

$$\begin{aligned} 0 \leq S_N - F_t(x^*(t)) &= \frac{1}{N}\max_{y\in\Delta}\sum_{i=1}^{N} L_t(x_{i-1}, y) - \min_{x\in X}\max_{y\in\Delta} L_t(x, y), \\ &\leq \frac{1}{N}\max_{y\in\Delta}\sum_{i=1}^{N} L_t(x_{i-1}, y) - \frac{1}{N}\min_{x\in X}\sum_{i=1}^{N} L_t(x, y_{i-1}). \end{aligned}$$

On déduit alors de l'inégalité 2.8 que

$$0 \leq N\left(S_N - F_t(x^*(t))\right) \leq \beta_{X,N}\overline{V}_X - \sum_{i=1}^{N}\langle x_{i-1}-\overline{x}, \eta_X(x_{i-1},y_{i-1})\rangle + \beta_{Y,N}\overline{V}_Y + \sum_{i=1}^{N}\langle y_{i-1}-\overline{y}, \eta_Y(x_{i-1},y_{i-1})\rangle$$
$$+\sum_{i=1}^{N}\left(\frac{\parallel \Phi(x_{i-1},y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}} + \frac{\parallel \Psi(x_{i-1},y_{i-1}) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}}\right). \quad (4.4)$$

avec $\begin{cases} \overline{x}=\text{argmax}_{x\in X}\sum_{i=1}^{N}\langle x, \eta_X(x_{i-1},y_{i-1})\rangle, \\ \overline{y}=\text{argmin}_{y\in\Delta_1}\sum_{i=1}^{N}\langle y, \eta_Y(x_{i-1},y_{i-1})\rangle. \end{cases}$

La proposition 2.11 permet alors d'obtenir

Proposition 4.6 *Pour tout* $\alpha > 8\exp\left(-\frac{\varkappa N}{4}\right)$ *et tout* $\lambda \geq \lambda(N,\alpha)$ *où*

$$\lambda(N,\alpha) = \frac{1}{\sqrt{N}}\left((3M_X+2\sigma_X)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + (3M_Y+2\sigma_Y)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + (D_X\sigma_X + D_Y\sigma_Y)\sqrt{\frac{4}{\varkappa}\ln(8\alpha^{-1})}\right.$$
$$\left. +3\sigma_X\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + 3\sigma_Y\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\right),$$

on a

$$\mathbb{P}(S_N - F_t(x^*(t)) > \lambda) \leq \frac{\alpha}{2}$$

Majorons maintenant le second terme, à savoir $|S_N - \hat{S}_N|$. Nous avons

$$\begin{aligned} N|S_N - \hat{S}_N| &= \max_{j=0,\ldots,m}\sum_{i=1}^{N} f_{j,t}(x_{i-1}) - \max_{j=0,\ldots,m}\sum_{i=1}^{N}\psi_{j,t}(x_{i-1}), \\ &\leq \max_{j=0,\ldots,m}\left|\sum_{i=1}^{N} f_{j,t}(x_{i-1}) - \psi_{j,t}(x_{i-1})\right|, \\ &\leq \max\left(\max_{y\in\Delta}\langle y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle, \max_{y\in\Delta} -\langle y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle\right). \end{aligned}$$

En utilisant le lemme C.1, il vient

$$\begin{aligned} \max_{y\in\Delta}\langle y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle &\leq \beta_Y\overline{V}_Y + \sum_{i=1}^{N}\langle \check{y}_{i-1}, \eta_Y(x_{i-1},y_{i-1})\rangle + \sum_{i=1}^{N}\frac{1}{2\alpha_{Y,0}\beta_Y}\parallel \eta_Y(x_{i-1},y_{i-1}) \parallel_{Y,*}^2, \\ \max_{y\in\Delta}\langle -y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle &\leq \beta_Y\overline{V}_Y + \sum_{i=1}^{N}\langle \tilde{y}_{i-1}, \eta_Y(x_{i-1},y_{i-1})\rangle + \sum_{i=1}^{N}\frac{1}{2\alpha_{Y,0}\beta_Y}\parallel \eta_Y(x_{i-1},y_{i-1}) \parallel_{Y,*}^2, \end{aligned}$$

où

$$\beta_Y = \sigma_Y\sqrt{\frac{N}{2\alpha_{Y,0}\overline{V}_Y}},\ \check{y}_i = W'_{\beta_Y}\left(-\sum_{k=1}^{i-1}\eta_Y(x_{k-1},y_{k-1})\right),\ \tilde{y}_i = W'_{\beta_Y}\left(\sum_{k=1}^{i-1}\eta_Y(x_{k-1},y_{k-1})\right).$$

Les propositions B.5 et B.6 permettent alors d'obtenir

Proposition 4.7 *Pour tout* $\alpha > 8\exp\left(-\frac{\varkappa N}{4}\right)$ *et tout* $\lambda \geq \lambda'(N,\alpha)$ *où*

$$\lambda'(N,\alpha) = \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}} + \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{4}{\varkappa}\ln(8\alpha^{-1})} + \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})},$$

on a

$$\mathbb{P}\left(\max_{y\in\Delta}\langle y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle > \lambda'\right) \leq \frac{\alpha}{4} \text{ et } \mathbb{P}\left(\max_{y\in\Delta}\langle y, \sum_{i=1}^{N}\eta_Y(x_{i-1},y_{i-1})\rangle > \lambda'\right) \leq \frac{\alpha}{4}.$$

Nous déduisons alors des propositions 4.6 et 4.7

Théorème 4.1 *Pour tout* $\alpha > 8\exp\left(-\frac{\varkappa N}{4}\right)$ *on a*

$$\mathbb{P}\left(f^*(t) \in [\hat{S}_N - \lambda(N,\alpha) - \lambda'(N,\alpha), \hat{S}_N + \lambda'(N,\alpha)]\right) \geq 1-\alpha$$

où

$$\begin{aligned}
\lambda(N,\alpha) &= \frac{1}{\sqrt{N}}\left((3M_X+2\sigma_X)\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + (3M_Y+2\sigma_Y)\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + (D_X\sigma_X + D_Y\sigma_Y)\sqrt{\frac{4}{\varkappa}\ln(8\alpha^{-1})}\right.\\
&\quad \left. +3\sigma_X\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + 3\sigma_Y\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\right), \qquad (4.5)\\
\lambda'(N,\alpha) &= \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}} + \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{4}{\varkappa}\ln(8\alpha^{-1})} + \frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\sqrt{\frac{28}{5\varkappa^2 N}\ln(8\alpha^{-1})}.
\end{aligned}$$

Nous avons maintenant en main tous les outils nécessaires pour construire un intervalle de confiance pour $f^*(t)$ étant donné t. Plus précisément il nous faut

- choisir un risque $\alpha > 0$ et $N \in \mathbb{N}$ satisfaisant les prémisses du théorème 4.1,
- effectuer N pas dans l'algorithme 2.5 et calculer $\hat{S}_N$.

Un intervalle de confiance au niveau $1-\alpha$ pour $f^*(t)$ est alors donné par $[f^-(t), f^+(t)] = [\hat{S}_N - \lambda(N,\alpha) - \lambda'(N,\alpha), \hat{S}_N + \lambda'(N,\alpha)]$. Ce sont ces valeurs que nous allons utiliser dans l'algorithme 4.1 après nous être assurés qu'elles vérifient l'inégalité $f^-(t) \geq (1-\kappa)f^+(t)$ puisqu'il faut satisfaire cette condition pour calculer la valeur suivante de t dans ce même algorithme.

Remarque : La valeur prise par la constante M_Y est susceptible de varier avec la valeur courante de t dans l'algorithme 4.1. Dans tout ce qui suit nous supposerons connue une borne supérieure $\overline{t}$ pour t^* et nous noterons $M_Y = \sup_{t\in[t_0,\overline{t}]} M_Y(t)$.

4.3.3 Complexité : préliminaires

Supposons fixé un niveau de confiance $1-\alpha$. Nous allons maintenant expliquer comment choisir le nombre de pas à effectuer dans l'algorithme 2.5. Commençons par nous donner quelques notations afin de simplifier les calculs à venir :

- N_0 est le plus petit entier supérieur à $\frac{4}{\varkappa}\ln(8\alpha^{-1})$,
- $\overline{\lambda}(\alpha) = \sup_{n\geq N_0}\sqrt{n}\lambda(n,\alpha)$ et $\overline{\lambda}'(\alpha) = \sup_{n\geq N_0}\sqrt{n}\lambda'(n,\alpha)$ où $\lambda(N,\alpha)$ et $\lambda'(N,\alpha)$ sont donnés par (4.5).

Remarquons que la proposition 4.6 est encore vraie si on remplace $\lambda(N,\alpha)$ et $\lambda'(N,\alpha)$ par $\frac{1}{\sqrt{N}}\overline{\lambda}(\alpha)$ et $\frac{1}{\sqrt{N}}\overline{\lambda}'(\alpha)$ respectivement. Afin de simplifier les calculs et les notations, nous allons donc utiliser ces deux valeurs dans la suite.

Notons n_k le nombre de pas à effectuer dans l'algorithme 2.5 pour obtenir l'intervalle de confiance pour t_k et passons en revue les contraintes que doit satisfaire n_k :

- Pour garantir un niveau de confiance au moins égal à $1-\alpha$, d'après la proposition 4.6, il faut choisir $n_k \geq N_0$.
- Notons $\Lambda_k = f^+(t_k) - f^-(t_k)$, c'est-à-dire

$$\Lambda_k = \frac{\overline{\lambda}(\alpha) + 2\overline{\lambda}(\alpha)}{\sqrt{n_k}}$$

par construction. Si l'inégalité $f^-(t_j) \geq (1-\kappa)f^+(t_j)$ est vraie pour $j \in [\![0,k]\!]$, on a

$$\begin{aligned}
\Lambda_k &= f^+(t_k) - f^-(t_k), \\
&= f^+(t_k)\left(1 - \frac{f^-(t_k)}{f^+(t_k)}\right), \\
&\leq \kappa f^+(t_k), \\
&\leq \frac{\kappa}{1-\kappa}\sqrt{f^*(t_0)(t^*-t_0)}\left[\frac{1}{2(1-\kappa)}\right]^k,
\end{aligned}$$

où la dernière inégalité a été montrée dans la preuve du lemme C.13. Notons alors

$$n_k^* = \left(\frac{1-\kappa}{\kappa}\right)^2 \frac{(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha))^2}{f^*(t_0)(t^*-t_0)}(2(1-\kappa))^{2k}. \tag{4.6}$$

Nous avons la borne inférieure $n_k \geq n_k^*$.

- On a $n_k \leq n(\epsilon,\alpha)$ où $n(\epsilon,\alpha) = \frac{1}{\kappa^2\epsilon^2}\left(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)\right)^2$. En effet : puisque

$$f^-(t_k) = f^+(t_k) - \frac{\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)}{\sqrt{n_k}},$$

si $f^+(t_k)(1-\kappa) > f^-(t_k)$ et $n_k \geq n(\epsilon,\alpha)$, alors il vient $f^+(t_k) < \epsilon$ et le critère d'arrêt de l'algorithme 4.1 est par conséquent satisfait.

Remarque : La valeur de n_k^* donnée par l'équation (4.6) dépend de $f^*(t_0)$ et de $t*$ dont nous ne connaissons pas *a priori* la valeur. Nous devrons donc dans la pratique y substituer des valeurs calculables à l'aide des constantes du problème que nous connaissons. Comme f^* est lipschitzienne de constante 1, $f^*(t_0) \leq t^* - t_0$. Le problème revient donc à trouver un majorant $\overline{t}$ de t_*, ce que nous avons déjà fait au paragraphe 4.2.4. Nous utiliserons donc en pratique la valeur de n_k^* suivante

$$n_k^* = \left(\frac{1-\kappa}{\kappa}\right)^2 \frac{(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha))^2}{(\overline{t}-t_0)^2}(2(1-\kappa))^{2k}. \tag{4.7}$$

Maintenant que nous avons donné des bornes pour n_k, il nous reste à décider comment nous allons fixer sa valeur dans la pratique puisque le nombre de pas n_k à effectuer dans l'algorithme 2.5 doit être connu à l'avance. L'idée est de choisir une valeur de n_k qui soit faible (afin de limiter

autant que faire se peut le temps de calcul de l'algorithme), de calculer $f^+(t_k)$ et $f^-(t_k)$ à l'aide de l'algorithme 2.5 et, si la condition $f^+(t_k)(1-\kappa) \leq f^-(t_k)$ n'est pas satisfaite, de recommencer en augmentant la valeur de n_k. Plus précisément, la valeur initiale de n_k sera fixée à $\min\{N_0, n_k^*\}$ et tant que des valeurs adéquates de $f^+(t_k)$ et $f^-(t_k)$ n'auront pas été trouvées nous doublerons la valeur courante de n_k jusqu'à ce qu'elle atteigne (ou dépasse) la valeur $n(\epsilon, \alpha)$ en actualisant les valeurs de $f^+(t_k)$ et $f^-(t_k)$ à chaque étape à l'aide de l'algorithme 2.5.

4.3.4 Algorithme : complexité et précision

Donnons maintenant l'algorithme que nous proposons pour résoudre le problème (4.1). Le schéma général de l'algorithme reprend comme annoncé celui de l'algorithme 4.1, le calcul des valeurs de f^+ et f^- s'effectue dans une boucle interne à l'aide de l'algorithme 2.5 et la politique de choix des valeurs pour n_k est celle que nous venons de décrire.

Algorithme 4.3

- *Initialisation : choisir* $t_0 < t^*$, $\kappa \in \left]0, \frac{1}{2}\right[$, $\alpha > 0$ *et* $\epsilon > 0$.
- *Pas* k *(*$k \geq 0$*) :*
  1. $n_k = \max(N_0, n_k^*)$
  2. *Utiliser l'Algorithme 2.5 pour calculer* $x^{(n_k)}$, $\hat{S}_{n_k}$

$$\begin{aligned} f^-(t_k) &= \hat{S}_{n_k} - \frac{\overline{\lambda}(\alpha) + \overline{\lambda}'(\alpha)}{\sqrt{n_k}} \\ f^+(t_k) &= \hat{S}_{n_k} + \frac{\overline{\lambda}'(\alpha)}{\sqrt{n_k}} \end{aligned}$$

  3. *Alors*
     - *si* $n_k \geq n(\epsilon, \alpha)$ *ou* $f^+(t_k) \leq \epsilon$, *fin de l'algorithme , renvoyer* $x^{(k)}$, t_k.
     - *si* $f^-(t_k) < (1-\kappa) f^+(t_k)$ *faire la mise à jour* $n_k \leftarrow \min(2n_k, n(\epsilon, \alpha))$ *et aller à 2.*
     - *sinon* $t_{k+1} = \begin{cases} t_0 + f^-(t_0) & \text{si } k = 0, \\ t_k + f^-(t_k) \frac{t_k - t_{k-1}}{f^+(t_{k-1}) - f^-(t_k)} & \text{sinon.} \end{cases}$

Nous allons maintenant étudier la complexité et la précision de notre algorithme. Pour ce qui est de la complexité de l'algorithme 4.3, nous allons la mesurer en comptant le nombre total $T(\alpha, \epsilon)$ d'appels à l'oracle au cours de son exécution. La précision de l'algorithme, quant à elle, sera mesurée par $f^*(t_{sol}) - f^*(t^*)$ où t_{sol} est la valeur de t renvoyée par l'algorithme 4.3 : si aucune erreur, au sens utilisé dans le paragraphe 4.2.4 n'a lieu au cours de l'exécution de l'algorithme, cette quantité sera positive. La quantité $|t_{sol} - t^*|$ pourrait elle aussi être intéressante à évaluer en vue d'étudier la précision de notre algorithme, malheureusement l'algorithme 4.1 que nous avons choisi ne permet pas d'obtenir de bornes précises pour cette quantité.

Commençons par le premier point, à savoir la complexité de notre algorithme. D'après le corollaire 4.3, il faut d'une part effectuer au maximum $N(\epsilon)$ fois la boucle principale où $N(\epsilon) = \frac{1}{\ln(2(1-\kappa))} \ln \frac{\sqrt{(t^*-t_0)f^*(t_0)}}{(1-\kappa)\epsilon}$. D'autre part, au plus $\max\{3n(\epsilon,\alpha), N_0\}$ appels à l'oracle sont nécessaires pour construire les valeurs $f^+(t_k)$ et $f^-(t_k)$ lorsque t_k est donné. En effet,

- si $N_0 \geq n(\epsilon,\alpha)$ alors N_0 appels à l'oracle sont effectués pour calculer $f^+(t_k)$ et $f^-(t_k)$,
- si $N_0 < n(\epsilon,\alpha)$ la stratégie que nous avons adoptée pour les choix successifs de n_k nous garantit qu'au plus $3n(\epsilon,\alpha)$ appels à l'oracle seront nécessaires.

Nous avons donc obtenu

Théorème 4.2 *Le nombre d'appels à l'oracle $T(\alpha,\epsilon)$ dans l'algorithme 4.3est majoré par :*

$$T(\alpha,\epsilon) \leq N(\epsilon) \max\{3n(\epsilon,\alpha), N_0\}.$$

En particulier, si $N_0 < n(\epsilon,\alpha)$, on a

$$T(\alpha,\epsilon) \leq \frac{1}{\ln(2(1-\kappa))} \ln\left(\frac{\sqrt{(t^*-t_0)f^*(t_0)}}{(1-\kappa)\epsilon}\right) \frac{\left(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)\right)^2}{\kappa^2\epsilon^2}, \tag{4.8}$$

et ainsi $T(\alpha,\epsilon) = \mathcal{O}\left(\frac{\ln(\alpha^{-1})\ln(\epsilon^{-1})}{\epsilon^2}\right)$.

Remarque : L'expression donnée par (4.8) est seulement une borne supérieure et les valeurs réelles de $T(\alpha,\epsilon)$ pourront être bien plus petites. En effet :

- le nombre de pas à effectuer dans la boucle principale peut être plus petit que $N(\epsilon)$,
- si $N_0 < n(\epsilon,\alpha)$, alors la valeur finale de n_k est au plus une fois égale à $n(\epsilon,\alpha)$ (dans la dernière boucle principale), dans la grande majorité des boucles, ce nombre peut être nettement plus petit (en particulier dans les toutes premières boucles où les intervalles de confiance à calculer n'ont pas besoin d'être de faible amplitude).
- Afin de simplifier les calculs, nous nous sommes donnés une valeur $\overline{\lambda}(\alpha)$ qui ne dépendait pas du nombre n de pas à exécuter dans la boucle interne de l'algorithme 4.3, dans la pratique, nous pouvons utiliser les valeurs $\lambda^*(n,\alpha)$ données dans la proposition 4.6. Comme $\lambda^*(n,\alpha)$ décroît avec n et est plus petit que $\overline{\lambda}(\alpha)$, il est possible de réduire ainsi le nombre d'appels à l'oracle à effectuer.

Ainsi la complexité de l'algorithme 4.3 peut être inférieure à la borne donnée dans la proposition 4.2. Pour illustrer plus concrètement la remarque précédente, supposons que nous ayons $N(\epsilon)$ boucles principales, n_k^* pas dans la $k^{ème}$ boucle et oublions N_0 pour cet exemple. Alors

$$\begin{aligned}
T(\alpha,\epsilon) = \sum_{k=0}^{N(\epsilon)} n_k^* &= \left(\frac{1-\kappa}{\kappa}\right)^2 \frac{\left(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)\right)^2}{f^*(t_0)(t^*-t_0)} \sum_{k=0}^{N(\epsilon)} (2(1-\kappa))^{2k}, \\
&\leq \left(\frac{1-\kappa}{\kappa}\right)^2 \frac{\left(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)\right)^2}{f^*(t_0)(t^*-t_0)} \frac{(2(1-\kappa))^{2(N(\epsilon)+1)}}{(2(1-\kappa))^2 - 1}, \\
&\leq \frac{4\left(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha)\right)^2 (1-\kappa)^2}{\kappa^2(1-2\kappa)(3-2\kappa)\epsilon^2},
\end{aligned}$$

par conséquent,la complexité de l'algorithme devient $T(\alpha,\epsilon) = \mathcal{O}\left(\frac{\ln(\alpha^{-1})}{\epsilon^2}\right)$. Remarquons que la suite $(n_k)_k$ n'est pas nécessairement croissante, c'est-à-dire que l'amplitude de l'intervalle de confiance pour t_k n'est pas toujours une fonction décroissante de l'indice k.

Nous abordons maintenant le problème de la précision de l'algorithme que nous avons proposé. Nous allons nous intéresser à la quantité $\varepsilon = f^*(t_{sol}) - f^*(t^*)$ où t_{sol} est la valeur de t renvoyée par l'algorithme à la fin de son exécution. Deux cas peuvent se produire :

- la vraie valeur de f^* est toujours dans notre intervalle de confiance (nous bornerons la probabilité de cet événement par 1 dans la suite) auquel cas nous avons $\varepsilon \leq \epsilon$;
- au moins un des intervalles de confiance construits ne contient pas le point $f^*(t_k)$ qu'il était supposé contenir (ce qui se produit avec une probabilité inférieure à $\alpha N(\epsilon)$) ; dans ce (mauvais) cas nous devons nous contenter de choisir $t_{sol} = t_0$ (nous n'avons pas réduit l'erreur initiale et nous avons donc $\varepsilon \leq t^* - t_0$ (puisque f^* est continûment lipschitzienne de constante 1 d'après la proposition 4.1).

Par conséquent, nous avons obtenu

$$\begin{aligned}\mathbb{E}[\varepsilon] &\leq \epsilon + \alpha N(\epsilon)(t^* - t_0),\\ &\leq \epsilon + \frac{\alpha(t^* - t_0)}{\ln(2(1-\kappa))} \ln \frac{\sqrt{(t^* - t_0) f^*(t_0)}}{(1-\kappa)\epsilon}.\end{aligned}$$

4.4 Exemple : problème de Markowitz

Introduction

Le problème auquel nous nous intéressons maintenant est celui de la gestion de portefeuille boursier. Le modèle que nous considérons ici est le modèle moyenne variance de Markowitz (voir [19]).

Nous souhaitons investir une certaine somme S dans n actifs risqués. Nous supposons que le temps est discrétisé et qu'à chaque pas de temps k nous pouvons réajuster le portefeuille, c'est-à dire modifier le montant $x_k^{(j)}$ de chacun des n actifs dans le portefeuille, les coûts liés à ces transactions étant supposés nuls. Nous notons $r_k^{(j)}$ le rendement de l'actif j au temps k. Le rendement du portefeuille au temps k est alors donné par

$$R(k) = \sum_{j=1}^{n} x_k^{(j)} r_k^{(j)}.$$

Les actifs considérés étant supposés risqués, le rendement de chacun des actifs peut s'écrire sous la forme

$$r_k^{(j)} = \rho_k^{(j)} + \eta_k^{(j)},$$

où les variables aléatoires sont centrées. Notons $r_k = (r_k^{(j)})_{j=1\cdots n}$, $\rho_k = (\rho_k^{(j)})_{j=1\cdots n}$ et $\eta_k = (\eta_k^{(j)})_{j=1\cdots n}$. Alors $\mathbb{E}[r_k] = \rho_k$ et nous noterons $Q_k = \mathbb{E}[\eta_k \eta_k^T]$ la matrice de covariance des rendements des actifs considérés. Par construction, la matrice Q_k est semi-définie positive. Nous nous plaçons ici dans le cas stationnaire, c'est-à-dire que ρ_k et Q_k ne dépendent pas de k et nous les noterons ρ et Q dans la suite.

Notre objectif est de minimiser les risques liés aux investissements, c'est-à-dire la variance $x^T Q x$ du

portefeuille, tout en garantissant un rendement minimal ℓ pour le portefeuille. Autrement dit nous allons chercher à résoudre le problème suivant :

$$\min_{x\in\Delta_S} \frac{1}{2}x^TQx \text{ sous contrainte } \rho^Tx \geq \ell. \tag{4.9}$$

Paramètres du problème

Pour résoudre ce problème, nous allons utiliser l'algorithme 2.5 et l'algorithme 4.1 comme décrit dans ce chapitre. Dans ce but, commençons par préciser les différents paramètres du problème que nous étudions. Les ensembles X et Y sont donnés par

$$X = \Delta_S^n = \left\{ x \in \mathbb{R}^n | X \geq 0, \sum_{j=1}^n x^{(j)} = S \right\} \text{ et } Y = \Delta_1^2 = \left\{ y \in \mathbb{R}^2 | Y \geq 0, y^{(1)} + y^{(2)} = 1 \right\}.$$

Les normes que nous considérons sur ces ensembles sont $\| \cdot \|_X = \| \cdot \|_1$, $\| \cdot \|_{X,*} = \| \cdot \|_\infty$, $\| \cdot \|_Y = \| \cdot \|_1$ et $\| \cdot \|_{Y,*} = \| \cdot \|_\infty$.
Lorsque le paramètre t est fixé, la fonction objectif est donnée par

$$L_t(x,y) = y^{(1)} \left(\frac{1}{2}x^TQx - t \right) + y^{(2)} \left(\rho^Tx - \ell \right),$$

et ses dérivées partielles valent respectivement

$$\frac{\partial L_t}{\partial x}(x,y) = y^{(1)}Qx + y^{(2)}\rho \text{ et } \frac{\partial L_t}{\partial y}(x,y) = \left(\frac{1}{2}x^TQx - t, \rho^Tx - \ell \right)^T,$$

de sorte que

$$\begin{aligned} \| \frac{\partial L_t}{\partial x}(x,y) \|_\infty &\leq S \| Q \|_\infty + \| \rho \|_\infty = M_X, \\ \| \frac{\partial L_t}{\partial y}(x,y) \|_\infty &\leq \frac{S^2}{2} \| Q \|_\infty + |t| + S \| \rho \|_\infty = M_Y(t). \end{aligned}$$

Les versions stochastiques de ces dérivées sont données par

$$\Phi(x,y) = \frac{\partial L_t}{\partial x}(x,y) + \eta_X(x,y) \text{ et } \Psi(x,y) = \frac{\partial L_t}{\partial y}(x,y) + \eta_Y(x,y),$$

où

$$\eta_X(x,y) = y^{(2)}\eta \text{ et } \eta_Y(x,y) = (0, \eta^Tx)^T.$$

Par conséquent, l'hypothèse 2.10 est vérifiée avec $\sigma_X^2 = n \| Q \|_\infty$ et $\sigma_Y^2 = S \| Q \|_\infty$.
Il ne nous reste plus, afin de mettre en œuvre notre algorithme qu'à préciser quelles sont les fonctions proximales à utiliser et à donner leur β-conjuguées respectives :

$$V_X(x) = S \ln\left(\frac{n}{S}\right) + \sum_{j=1}^n x^{(j)} \ln(x^{(j)}) \quad \text{et} \quad W_{X,\beta}(\zeta) = S\beta \ln\left(\frac{1}{n} \sum_{j=1}^n e^{\frac{\zeta^{(j)}}{\beta}} \right),$$

$$V_Y(y) = \ln(2) + \sum_{j=1}^2 y^{(j)} \ln(y^{(j)}) \quad \text{et} \quad W_{Y,\beta}(\xi) = \beta \ln\left(\frac{1}{2} \sum_{j=1}^2 e^{\frac{\xi^{(j)}}{\beta}} \right).$$

La fonction V_X est fortement convexe sur X de paramètre $\alpha_{X,0} = S^{-1}$ et son maximum sur X vaut $\overline{V}_X = S\ln(n)$. De même, la fonction V_Y est fortement convexe sur Y de paramètre $\alpha_{Y,0} = 1$ et son maximum sur Y vaut $\overline{V}_Y = \ln(2)$.

Algorithme

Donnons maintenant la version de l'algorithme 2.5 à laquelle nous allons faire appel au cours de l'exécution de l'algorithme 4.3 lorsque t est fixé. Conformément à la proposition 2.9, les suites $(\gamma_i)_i$, $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$ sont choisies constantes et telles que $\gamma_i \equiv 1$, $\beta_{X,i} \equiv \frac{(M_X+\sigma_X)\sqrt{N}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} \equiv \frac{(M_Y+\sigma_Y)\sqrt{N}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}$, soit

$$\gamma_i \equiv 1,\ \beta_{X,i} \equiv \frac{(S \parallel Q \parallel_\infty + \parallel \rho \parallel_\infty + \sqrt{n \parallel Q \parallel_\infty})\sqrt{N}}{\sqrt{2\ln(n)}},\ \beta_{Y,i} \equiv= \frac{(\frac{3S^2}{2} \parallel Q \parallel_\infty + |t| + S \parallel \rho \parallel_\infty)\sqrt{N}}{\sqrt{2\ln(2)}}.$$

Algorithme 4.4

- *Initialisation* $x_0 = (\frac{S}{n}, \cdots, \frac{S}{n})^T$, $y_0 = (\frac{1}{2}, \frac{1}{2})$ *et* $(\zeta_0, \xi_0) = 0 \in E_X^* \times E_Y^*$.
- *Pour* $i = 1, \cdots, N$ *faire la mise à jour*

$$\begin{aligned}
\zeta_i &= \zeta_{i-1} - \gamma_i y_{i-1}^{(1)} Q x_{i-1} + y_{i-1}^{(2)} r, \\
\xi_i &= \xi_{i-1} + \gamma_i \left(\frac{1}{2} x_{i-1}^T Q x_{i-1} - t, r^T x_{i-1} - \ell \right)^T, \\
x_i &= \nabla W_{X,\beta_{X,i}}(\zeta_i) = S \left(e^{\frac{\zeta_i^{(1)}}{\beta_X}}, \cdots, e^{\frac{\zeta_i^{(n)}}{\beta_X}} \right)^T \left(\sum_{j=1}^n e^{\frac{\zeta_i^{(j)}}{\beta_X}} \right)^{-1}, \\
y_i &= \nabla W_{Y,\beta_{Y,i}}(\xi_i) = \left(e^{\frac{\xi_i^{(1)}}{\beta}}, e^{\frac{\xi_i^{(2)}}{\beta}} \right)^T \left(e^{\frac{\xi_i^{(1)}}{\beta}} + e^{\frac{\xi_i^{(2)}}{\beta}} \right)^{-1}.
\end{aligned}$$

- *Sortie*

$$x^N = \frac{\sum_{i=1}^N \gamma_i\, x_{i-1}}{\sum_{i=1}^N \gamma_i},\ y^N = \frac{\sum_{i=1}^N \gamma_i\, y_{i-1}}{\sum_{i=1}^N \gamma_i}.$$

Nous supposons que l'hypothèse 2.12 est vérifiée pour un certain $\varkappa > 0$ (c'est par exemple le cas si le vecteur aléatoire η est gaussien). Nous pouvons alors résoudre le problème (4.9) en utilisant l'algorithme 4.3 avec $N_0 = \frac{4}{\varkappa}\ln(8\alpha^{-1})$, $\overline{\lambda}(\alpha) = \sup_{p \geq N_0} \lambda(p, \alpha)$, $\overline{\lambda}'(\alpha) = \sup_{p \geq N_0} \lambda'(p, \alpha)$ où $\lambda(p, \alpha)$ et $\lambda'(p, \alpha)$ sont donnés par le théorème 4.1 *mutatis mutandis*, n_k^* est donné par l'équation (4.7) et $n(\epsilon, \alpha) = \frac{1}{\varkappa^2 \epsilon^2}(\overline{\lambda}(\alpha) + 2\overline{\lambda}'(\alpha))^2$.

Chapitre 5

Conclusion et perspectives

Nous avons étudié dans ce mémoire plusieurs problèmes généraux relevant de l'optimisation stochastique en grande dimension : les problèmes de minimisation sans contrainte, les problèmes d'inégalités variationnelles à opérateur monotone et les problèmes de point selle. Les algorithmes, adaptatifs ou non, que nous avons développés pour les résoudre offrent de très bonnes propriétés de convergence théoriques et les simulations numériques que nous avons pu effectué jusqu'ici sont elles aussi porteuses de résultats prometteurs. Nous avons aussi mis au point un algorithme pour résoudre les problèmes d'optimisation sous contraintes qui jouit aussi de bonnes performances théoriques.

Les perspectives de poursuite de travaux pour cette thèse sont les suivantes.
Sur le plan théorique elles concernent essentiellement les problèmes d'optimisation sous contraintes et sont :

- mettre au point un algorithme de résolution pour ce type de problèmes qui s'appuie sur l'algorithme d'extrapolation dual, puis des versions adaptatives pour pouvoir gérer de manière plus efficace le choix des paramètres n_k,
- lorsque les contraintes sont de deux types, déterministes et stochastiques, envisager des stratégies qui permettent de les traiter séparément.

Sur le plan pratique, il s'agit de

- développer d'avantage les applications, notamment en statistiques et en gestion de production,
- effectuer une comparaison en pratique des différents algorithmes obtenus : d'une part comparer les versions "simples" et adaptatives des deux algorithmes de descente en miroir et d'extrapolation duale et d'autre part comparer les performances de ces deux algorithmes,
- il apparaît dans les simulations que "gonfler" les valeurs de β (par rapport aux valeurs qui minimisent l'espérance de l'erreur d'approximation) améliore les performances des différents algorithmes, il peut être intéressant de savoir en pratique à quel point il convient de surestimer les valeurs de β.

Annexe A

Eléments d'analyse convexe, transformée de Legendre-Fenchel

Nous donnons dans ce chapitre les résultats d'analyse convexe nécessaires la lecture de ce document. Les résultats présentés dans la première partie de ce chapitre sont classiques et seront énoncés sans preuve. Le lecteur intéressé trouvera leur démonstration dans tout ouvrage traitant de l'analyse convexe, par exemple [7] et [10]. Les résultats de la seconde partie sont empruntés à [3] et au chapitre 7 de [10].

Dans toute la suite, E est un $\mathbb{R}$-espace vectoriel de dimension finie muni d'une norme (dite *primale*) $\| \cdot \|$, E^* est l'espace dual associé et $\| \cdot \|_*$ la norme duale définie de manière usuelle pour tout $z \in E^*$ par :

$$\| z \|_* = \max_{\|u\| \leq 1} \langle z, u \rangle,$$

et Θ est un convexe compact inclus dans E.

A.1 Fonctions convexes régulières et fonctions fortement convexes

Définition A.1 *Soit f une fonction sur E à valeur réelles. On dit que f est convexe si pour tout x, $y \in E$ et tout $\lambda \in [0,1]$, on a*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

A.1.1 Fonctions convexes régulières

Notons $C_L^{1,1}(E)$ où $L \in \mathbb{R}_+^*$ l'ensemble des fonctions de classe C^1 sur E à valeurs réelles dont la dérivée soit L-lipschitzienne, c'est-à-dire que pour tous $x, y \in E$,

$$\| f'(x) - f'(y) \|_* \leq L \| x - y \| .$$

Lemme A.1 *Soit $f \in C_L^{1,1}(E)$. Alors pour tous x, $y \in E$, on a :*

$$|f(y) - f(x) - \langle f'(x),\ y - x \rangle| \leq \frac{L}{2} \| y - x \|_*^2 .$$

Proposition A.1 *Une fonction continûment différentiable $f(x)$ est convexe sur E (notation $f \in \mathcal{F}^1(E)$) si et seulement si pour tous x, $y \in E$, on a : $f(y) \geq f(x) + \langle f'(x),\ y - x \rangle$.*

Proposition A.2 *Une fonction $f \in \mathcal{F}^1(E)$ si et seulement si elle est continûment différentiable et pour tous x, $y \in E$ on a : $\langle f'(x) - f'(y), x - y \rangle \geq 0$.*

Notons $\mathcal{F}_L^{1,1}(E)$ la classe des fonctions convexes dont la dérivée est L-lipschitzienne.

Proposition A.3 *Soit $f \in \mathcal{F}^1$. Les assertions suivantes sont équivalentes :*

1. $f \in \mathcal{F}_L^{1,1}(E)$,
2. $\forall x,\ y \in E,\ 0 \leq f(y) - f(x) - \langle f'(x), y - x \rangle \leq \frac{L}{2} \| x - y \|^2$,
3. $\forall x,\ y \in E,\ f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2L} \| f'(x) - f'(y) \|_*^2$,
4. $\forall x,\ y \in \mathbb{R}^n,\ \langle f'(x) - f'(y), x - y \rangle \geq \frac{1}{L} \| f'(x) - f'(y) \|_*^2$.

A.1.2 Fonctions fortement convexes

Définition A.2 *Soit $\alpha_0 > 0$. Une fonction continue $d(\cdot)$ sur Θ est dite fortement convexe de paramètre α_0 pour la norme $\| \cdot \|$ si*

$$d(\lambda x + (1-\lambda)y) \leq \lambda d(x) + (1-\lambda)d(y) - \frac{\alpha_0}{2}\lambda(1-\lambda) \parallel x - y \parallel^2$$

pour tous x, $y \in \Theta$ et tout $\lambda \in [0,1]$.

Lemme A.2 *Si $d(\cdot)$ est α_0-fortement convexe sur Θ alors l'ensemble*

$$\operatorname*{argmin}_{x\in\Theta} d(x)$$

est un singleton x^* et pour tout $x \in \Theta$ on a

$$d(x) \geq d(x^*) + \frac{\alpha_0}{2} \parallel x - x^* \parallel^2 .$$

Définition A.3 *Soit $d(\cdot)$ une fonction α_0-fortement convexe sur Θ et $x_0 \in \Theta$ tel que $d(\cdot)$ soit différentiable en x_0. On appelle fonction proximale la fonction $V(\cdot, x_0)$ définie sur Θ par*

$$V(x, x_0) = d(x) - d(x_0) + \langle d'(x_0), x_0 - x\rangle.$$

Cette fonction est fortement convexe de paramètre α_0 sur Θ et vérifie pour tout $x \in \Theta$,

$$V(x, x_0) \geq \frac{\alpha_0}{2} \parallel x - x_0 \parallel^2 .$$

A.2 Transformée de Legendre d'une fonction convexe

A.2.1 Fonctions semi-continues inférieurement

Définition A.4 *Soit f une fonction définie sur E et à valeurs dans $\mathbb{R} \cup +\infty$. On dit que f est semi-continue inférieurement en $x \in E$ si pour toute suite $(x_i)_{i\in\mathbb{N}}$ convergeant vers x on a $f(x) \leq \liminf_{i\to+\infty} f(x_i)$. On dira que f est semi-continue inférieurement sur E si elle est semi-continue inférieurement en tout point de E.*

Proposition A.4 *Une fonction f définie sur $\mathbb{R}^n$ et à valeurs dans $\mathbb{R} \cup \{+\infty\}$ est semi-continue inférieurement si et seulement si son épigraphe est fermé.*

Définition A.5 *Une fonction convexe sera dite propre si son épigraphe est fermé.*

Proposition A.5 *Une fonction convexe f propre est la borne supérieure de tous ses minorants affines. De plus, si x est un point de l'intérieur du domaine de f, cette borne supérieure est un maximum : il existe un minorant affine de f sur E qui vaut $f(x)$ en x (c'est-à-dire qu'il existe un hyperplan affine qui sépare proprement le point $(x, f(x))$ et l'intérieur de l'épigraphe de f).*

Définition A.6 *Soit f une fonction convexe . On appelle clôture (convexe) de f et on note cl f la fonction (convexe) dont l'épigraphe est l'adhérence de l'épigraphe de f.*

A.2.2 Sousgradient, sousdifférentiel

Soit f une fonction convexe sur E et $x \in Dom(f)$. On suppose qu'il existe un minorant affine de f sur E qui soit exact en x, c'est-à-dire qu'il existe un vecteur $z \in E^*$ tel que $f(y) \geq f(x) + \langle z, y - x\rangle \; \forall y \in E$.

Définition A.7 *Tout vecteur z vérifiant la propriété précédente est appelé sousgradient de f en x. On appelle sousdifférentiel de f en x et on note $\partial f(x)$ l'ensemble de tous les sousgradients de f en x.*

Proposition A.6 *Soit f une fonction convexe et $x \in Dom(f)$. Alors :*

1. *$\partial f(x)$ est un ensemble convexe fermé qui est non vide dès que x appartient à l' intérieur du domaine de f.*
2. *Si $x \in int\, Dom(f)$ et est différentiable en x alors $\partial f(x)$ est un singleton et son unique élément est $\nabla f(x)$.*

A.2.3 Transformée de Legendre d'une fonction convexe

Définition A.8 *Soit f une fonction convexe sur E. On appelle transformée de Legendre de f et on note f^* la fonction définie pour tout $z \in E^*$ par $f^*(z) = \sup_{x\in E}[\langle z, x\rangle - f(x)]$.*

Proposition A.7 *Soit f une fonction convexe sur E. Alors :*

1. *f^* est convexe et semi-continue inférieurement ;*
2. *$(f^*)^* = cl\, f$. En particulier, si f est propre, $(f^*)^* = f$.*

Preuve

1. f^* est le supremum des fonctions affines $(g_x)_{x\in E}$ où $g_x(z) = \langle z, x\rangle - f(x)$ pour tout $z \in E^*$. Elle est donc convexe et semi-continue inférieurement par construction.
2. Pour tout $x \in E$, on a

$$(f^*)^*(x) = \sup_{z\in E^*}[\langle z, x\rangle - f^*(z)] = \sup_{z\in E^*,\, a\geq f^*(z)}[\langle z, x\rangle - a]. \tag{A.1}$$

   Or on a

$$(a \geq f^*(z)) \Leftrightarrow (\forall x \in X \; a \geq \langle z, x\rangle - f(x)) \Leftrightarrow (\forall x \in X \; f(x) \geq \langle z, x\rangle - a).$$

   Par conséquent, le second supremum dans (A.1) est la borne supérieure de tous les minorants affines de f, c'est donc la clôture de f et $(f^*)^* = cl(f)$.
   Dans le cas où f est propre, $f = cl(f)$ et donc $(f^*)^* = f$. □

Proposition A.8 *Soit f une fonction convexe. Alors :*

$$(z \in \partial f(x)) \Leftrightarrow (f^*(z) + f(x) - \langle z, x\rangle = 0),$$

et si f est propre :

$$(z \in \partial f(x)) \Leftrightarrow (x \in \partial f^*(z)).$$

Preuve
D'une part, l'inégalité suivante, dite inégalité de Fenchel, découle de la définition de f^* :

$$\forall (z,x) \in E^* \times E,\ f^*(z) + f(x) \geq \langle z, x\rangle.$$

D'autre part, $z \in E^*$ est un sousgradient de f en $x \in E$ si et seulement si

$$\forall y \in E\ f(y) \geq f(x) + \langle z, y - x\rangle,$$

ce qui est équivalent à $f^*(z) + f(x) \leq \langle z, x\rangle$. Si bien que $(z \in \partial f(x)) \Leftrightarrow (f^*(z) + f(x) - \langle z, x\rangle = 0)$. Si f est propre, $(f^*)^* = f$ et en appliquant ce qui vient d'être démontré à f puis à f^*, on obtient

$$(z \in \partial f(x)) \Leftrightarrow (f^*(z) + f(x) - \langle z, x\rangle = 0) \Leftrightarrow (x \in \partial f^*(z)).$$

□

A.2.4 Transformée de Legendre d'une fonction fortement convexe

Définition A.9 *Soient $\beta > 0$ et $V : \Theta \to \mathbb{R}$ une fonction convexe. On appelle β-conjuguée de V la fonction W_β donnée par*

$$\forall z \in E^*,\ W_\beta(z) = \sup_{\theta \in \Theta}\{\langle z, \theta\rangle - \beta V(\theta)\}.$$

Notons que si V est à valeurs positives et si $\beta_2 \geq \beta_1 \geq 0$ alors pour tout $z \in E^*$,

$$W_{\beta_2}(z) \leq W_{\beta_1}(z).$$

Théorème A.1 *Soient $V : \Theta \to \mathbb{R}$ une fonction convexe continue et $\beta > 0$. Alors la $\beta-$conjuguée W_β de V possède les propriétés suivantes :*

1. *La fonction $W_\beta : E^* \to \mathbb{R}$ est convexe et sa conjuguée est βV, c'est-à-dire*

$$\forall \theta \in \Theta,\ \beta V(\theta) = \sup_{z \in E^*}\{\langle z, \theta\rangle - W_\beta(z)\};$$

2. *Si V est α_0-fortement convexe pour la norme $\|\cdot\|$, alors :*
   (a) *W_β est continûment différentiable et son gradient W'_β vérifie*

$$\operatorname*{argmax}_{\theta \in \Theta}\{\langle z, \theta\rangle - \beta V(\theta)\} = W'_\beta(z) \in \Theta;$$

   (b)

$$\| W'_\beta(z) - W'_\beta(\tilde{z}) \| \leq \frac{1}{\alpha_0 \beta} \| z - \tilde{z} \|_*,\ \forall z, \tilde{z} \in E^*,$$

   et par conséquent

$$W_\beta(z) \leq W_\beta(\tilde{z}) + \langle W'_\beta(z), z - \tilde{z}\rangle + \frac{1}{2\alpha_0 \beta} \| z - \tilde{z} \|_*^2\ \ \forall z, \tilde{z} \in E^*.$$

Preuve

1. Preuve analogue à celle de la proposition A.7 en remarquant que la fonction V est propre puisqu'elle est continue.
2.
   (a) Comme W_β est convexe, il suffit de vérifier que pour tout $z \in E^*$, $\partial W_\beta(z)$ est un singleton. La proposition A.8 et la compactité de Θ impliquent que $\partial W_\beta(z) = \text{argmax}_{\theta\in\Theta}\{\langle z, \theta\rangle - \beta V(\theta)\}$. La convexité forte de V a pour conséquence que ce dernier ensemble est réduit à un point, et donc $\partial W_\beta(z)$ est un singleton.

   (b) Soient $z, \tilde{z} \in E^*$. On pose $x = W'_\beta(z)$ et $y = W'_\beta(\tilde{z})$. Comme V est fortement convexe de paramètre α_0,
   $$\langle \frac{1}{\beta}(z - \tilde{z}), x - y\rangle \geq \alpha_0 \parallel x - y \parallel^2 .$$
   Et puisque $\langle z - \tilde{z}, x - y\rangle \leq \parallel x - y \parallel \cdot \parallel z - \tilde{z} \parallel_*$, il vient
   $$\parallel x - y \parallel = \parallel W'_\beta(z) - W'_\beta(\tilde{z}) \parallel \leq \frac{1}{\alpha_0\beta} \parallel z - \tilde{z} \parallel_*,$$
   ce qui prouve que $W_\beta \in \mathcal{F}^{1,1}_{\frac{1}{\alpha_0\beta}}(E^*)$. La dernière égalité à démontrer est alors donnée par la proposition A.3.

Annexe B

Inégalités de déviation

B.1 Introduction

Le but de ce chapitre est de donner des résultats généraux permettant d'étudier des bornes pour les probabilité de déviation des erreurs commises par les différents algorithmes étudiés. L'intérêt de ce type de résultat est double :

- ces bornes en déviation viennent, en premier lieu, en complément des bornes en espérance obtenues pour les algorithmes que nous avons étudiés afin de nous permettre de mieux appréhender leur précision ;
- en second lieu, elles nous sont aussi indispensables pour aborder les problèmes d'optimisation sous contraintes du chapitre 4.

B.1.1 Hypothèses

Commençons par nous donner un cadre général pour cette étude. On considère une suite $(X_i)_i$ de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $\mathcal{F} = (\mathcal{F}_i)_i$. On fera dans la suite l'hypothèse suivante :

Hypothèse B.1 – *$(X_i)_i$ est une différence de martingale dans la filtration $(\mathcal{F}_i)_i$:*

$$\mathbb{E}[X_i|\mathcal{F}_{i-1}] = 0 \text{ p.s. pour tout } i > 0. \tag{B.1}$$

– *Il existe des constantes $(\sigma_i)_i$ et σ strictement positives telles que pour tout i, $\sigma_i \leq \sigma \leq \infty$ et*

$$\mathbb{E}[X_i^2|\mathcal{F}_{i-1}] \leq \sigma_i^2 \text{ p.s..} \tag{B.2}$$

Cette hypothèse est simplement une reformulation des hypothèses effectuées lors de l'étude des bornes en espérance pour la précision des différents algorithmes. Pour obtenir des bornes en déviations, nous allons faire une hypothèse supplémentaire et supposer que les variables aléatoires $(X_i)_i$ ont un moment exponentiel. Plus précisément, on supposera que l'une des hypothèses suivantes est vérifiée :

Hypothèse B.2 *Il existe une constantes $\varkappa > 0$ telle que*

$$\mathbb{E}[\exp(\varkappa|X_i|\sigma_i^{-1}|\mathcal{F}_{i-1}] \leq e \text{ p.s. pour tout } i > 0. \tag{B.3}$$

Hypothèse B.3 *Il existe une constante $\varkappa > 0$ telle que*

$$\mathbb{E}[\exp(\varkappa|X_i|^2\sigma_i^{-2}|\mathcal{F}_{i-1}] \leq e \text{ p.s. pour tout } i > 0. \tag{B.4}$$

Nous voulons établir des bornes pour les probabilités de déviations modérées de $S_N^{(1)}$ et $S_N^{(2)}$ où

$$S_N^{(1)} = \sum_{i=1}^N X_i \text{ et } S_N^{(2)} = \sum_{i=1}^N (X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}]),$$

quand l'hypothèse B.1 et l'une des 2 hypothèses précédentes sont vérifiées.

B.1.2 Lemmes calculatoires

On rappelle ici quelques résultats qui seront utilisés dans la suite.

Lemme B.1 *Soient a, b et c des réels positifs.Alors*

$$\sup_{x\geq 0} x^a e^{-cx^b} = \left(\frac{a}{c\,b\,e}\right)^{\frac{a}{b}} ;$$

et le maximum est atteint pour $x = \left(\frac{a}{bc}\right)^{\frac{1}{b}}$.

Corollaire B.1 *Si Y est une variable aléatoire qui satisfait $\mathbb{E}[e^{|Y|}] \leq e$, alors pour tout $k \in \mathbb{N}$, on a $\mathbb{E}[|Y|^k] \leq (\frac{k}{e})^k e$.*

Preuve

$$\mathbb{E}[|Y|^k] = \mathbb{E}[|Y|^k e^{-|Y|} e^{|Y|}] \leq \sup_{x>0} x^k\, e^{-x} \mathbb{E}[e^{|Y|}] \leq (\frac{k}{e})^k e.$$

□

Lemme B.2 *Pour tout $x \in \mathbb{R}$ et tout $k \in \mathbb{N}$*

$$\left| e^x - \sum_{i=0}^{k-1} \frac{x^i}{i!} \right| \leq \frac{|x^k|}{k!} e^{|x|}.$$

Lemme B.3 *(Formule de Stirling) Pour tout $k \geq 1$*

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \leq k! \leq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12k}}.$$

Corollaire B.2 *Soit Y une variable aléatoire telle que $\mathbb{E}[e^{|Y|}] \leq e$. Pour tout $0 \leq s < 1$, on a*

$$\mathbb{E}\left[\exp(sY)\right] \leq \mathbb{E}\left[sY\right] + \frac{s^2}{2}\mathbb{E}\left[Y^2\right] + \frac{es^3}{\sqrt{6\pi}}\frac{1}{1-s}.$$

Preuve
A l'aide du corollaire B.1 et de la formule de Stirling, on obtient

$$\begin{aligned}
\mathbb{E}\left[\exp(sY)\right] &\leq 1 + \mathbb{E}\left[sY\right] + \frac{s^2}{2}\mathbb{E}\left[Y^2\right] + \sum_{k=3}^{\infty}\frac{s^k}{k!}\mathbb{E}\left[Y^k\right], \\
&\leq 1 + \mathbb{E}\left[sY\right] + \frac{s^2}{2}\mathbb{E}\left[Y^2\right] + e\sum_{k=3}^{\infty}\frac{s^k}{\sqrt{2\pi k}}, \\
&\leq 1 + \mathbb{E}\left[sY\right] + \frac{s^2}{2}\mathbb{E}\left[Y^2\right] + \frac{es^3}{\sqrt{6\pi}}\frac{1}{1-s}.
\end{aligned}$$

□

B.2 Borne avec un moment exponentiel d'ordre 1

On commence par le cas où les hypothèses B.1 et B.2 sont satisfaites. On cherche dans un premier temps une borne pour les déviations modérées de $S_N^{(1)} = \sum_{i=1}^N X_i$.

Proposition B.1 *Notons*

$$\lambda^* = \frac{\varkappa}{\sigma}\left(\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}\right)\sqrt{\sum_{i=1}^N \sigma_i^2}.$$

Alors pour tout $0 < \lambda \leq \lambda^*$, *on a*

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{1}{4}\left(\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}\right)^{-1}\lambda^2\right),$$

et pour tout $\lambda > \lambda^*$,

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{\varkappa}{2\sigma}\sqrt{\sum_{i=1}^N \sigma_i^2}\left(\lambda - \frac{\lambda^*}{2}\right)\right).$$

Preuve
Soit $0 < t \leq \frac{\varkappa}{2\sigma}$. Pour tout i, on déduit du corollaire B.2 avec $Y = \varkappa\frac{X_i}{\sigma_i}$ et $s = \frac{t\sigma_i}{\varkappa}$ que

$$\mathbb{E}\left[\exp(tX_i)|\mathcal{F}_{i-1}\right] \leq 1 + \frac{t^2}{2}\sigma_i^2 + \frac{et^2\sigma_i^2}{\varkappa^2\sqrt{6\pi}}. \tag{B.5}$$

On a alors, par inégalité de Markov, pour tout $\lambda > 0$

$$\begin{aligned}\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) &\leq \mathbb{P}\left(tS_N^{(1)} \geq t\lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right), \\ &\leq \mathbb{E}\left[\exp(tS_N^{(1)})\right]\exp\left(-\lambda t\sqrt{\sum_{i=1}^N \sigma_i^2}\right).\end{aligned}$$

En utilisant (B.5) et des conditionnements successifs, on obtient

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(t^2\sum_{i=1}^N \sigma_i^2\left(\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}\right) - \lambda t\sqrt{\sum_{i=1}^N \sigma_i^2}\right). \tag{B.6}$$

Notons

$$\lambda^* = \frac{\varkappa}{\sigma}\left(\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}\right)\sqrt{\sum_{i=1}^N \sigma_i^2}.$$

Si $0 < \lambda \leq \lambda^*$, on peut choisir

$$t = \frac{\lambda}{2}\left(\frac{1}{2} + \frac{e}{\varkappa^2\sqrt{6\pi}}\right)^{-1}\left(\sum_{i=1}^N \sigma_i^2\right)^{-\frac{1}{2}}$$

dans (B.6) et on a

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{1}{4}\left(\frac{1}{2}+\frac{e}{\varkappa^2\sqrt{6\pi}}\right)^{-1}\lambda^2\right);$$

sinon on choisit $t = \frac{\varkappa}{2\sigma}$ et on obtient

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{\varkappa}{2\sigma}\sqrt{\sum_{i=1}^N \sigma_i^2}\left(\lambda-\frac{\lambda^*}{2}\right)\right).$$

□

La proposition suivante se déduit immédiatement de la proposition précédente

Proposition B.2 *Notons*

$$\lambda^* = \left(\frac{\varkappa}{2}+\frac{e}{\varkappa\sqrt{6\pi}}\right)\sqrt{N}.$$

Alors pour tout $0 < \lambda \leq \lambda^*$, *on a*

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sigma\sqrt{N}\right) \leq \exp\left(-\frac{1}{4}\left(\frac{1}{2}+\frac{e}{\varkappa^2\sqrt{6\pi}}\right)^{-1}\lambda^2\right),$$

et pour tout $\lambda > \lambda^*$,

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sigma\sqrt{N}\right) \leq \exp\left(-\frac{\varkappa}{2}\sqrt{N}\left(\lambda-\frac{\lambda^*}{2}\right)\right).$$

En particulier, pour tout niveau de confiance α *donné, on a* $\mathbb{P}\left(S_N^{(1)} \geq \lambda\sigma\sqrt{N}\right) \leq \alpha$ *dés que*

$$\lambda \geq 2\sqrt{\left(\frac{1}{2}+\frac{e}{\varkappa^2\sqrt{6\pi}}\right)}\sqrt{\ln(\alpha^{-1})}.$$

On s'intéresse maintenant aux déviations modérées de $S_N^{(2)} = \sum_{i=1}^N (X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}])$.

Proposition B.3 *Soit* $N \in \mathbb{N}^*$, $K > 0$ *et* $t > 0$. *Supposons que* $t \leq \frac{\varkappa}{2\sigma}K^{-1}$ *et* $K \geq \frac{2}{\varkappa}\sigma$. *Alors, pour tout* $\tau > 0$

$$\mathbb{P}\left(\sum_{i=1}^N (X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}]) > \tau\right) \leq N\, e^{1-\varkappa\frac{K}{\sigma}} + \exp\left(-t\tau + tNK^2 e^{-\varkappa\frac{K}{\sigma}}e + \frac{Nt^2}{2}e\left(\frac{8}{\varkappa e}\right)^4\sigma^4\right).$$

Preuve

Dans cette preuve, on note $Y_i = X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}]$. Montrons d'abord les inégalités suivantes

Lemme B.4 *Soit* $K \geq \frac{2}{\varkappa}\sigma$, *alors*

$$|\mathbb{E}[Y_i \mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}]| \leq K^2 e^{-\varkappa\frac{K}{\sigma}}e, \tag{B.7}$$

$$\mathbb{E}[Y_i^2 e^{\frac{\varkappa}{2}\left|\frac{X_i}{\sigma}\right|}] \leq e\left(\frac{8}{\varkappa e}\right)^4\sigma^4. \tag{B.8}$$

Preuve
Remarquons que

$$X_i^2\mathbf{1}_{\{|X_i|\leq K\}} - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}] \leq Y_i\mathbf{1}_{\{|X_i|\leq K\}} \leq X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}]\mathbf{1}_{\{|X_i|\leq K\}}$$

et

$$|\mathbb{E}[Y_i\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}]| = \mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|>K\}}|\mathcal{F}_{i-1}] \leq \mathbb{E}[e^{\varkappa\left|\frac{X_i}{\sigma}\right|}]\sup_{x>K} x^2e^{-\varkappa\frac{x}{\sigma}}.$$

Alors pour $x \geq \frac{2\sigma}{u}$, la fonction $x \to x^2e^{-u\frac{x}{\sigma}}$ est décroissante. Par conséquent, pour $K \geq \frac{2}{\varkappa}\sigma$

$$|\mathbb{E}[Y_i\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}]| \leq K^2e^{-\varkappa\frac{K}{\sigma}}e.$$

On déduit (B.8) du lemme B.1 :

$$\begin{aligned}
\mathbb{E}[Y_i^2e^{\frac{\varkappa}{2}\left|\frac{X_i}{\sigma}\right|}] &\leq \mathbb{E}[X_i^4e^{\frac{\varkappa}{2}\left|\frac{X_i}{\sigma}\right|}],\\
&\leq \mathbb{E}[e^{\varkappa\left|\frac{X_i}{\sigma}\right|}]\sup_{x>0} x^4e^{-\frac{\varkappa}{2}\frac{x}{\sigma}}\\
&\leq e\left(\frac{8}{\varkappa e}\right)^4\sigma^4.
\end{aligned}$$

□

Soit t tel que $0 < t \leq \frac{\varkappa}{2\sigma}K^{-1}$. Sur l'événement $\{X_i \leq K\}$, on a

$$tY_i \leq tX_i^2 \leq \frac{\varkappa}{2\sigma}|X_i|,$$

et on déduit des lemmes B.1 et B.2 que

$$\begin{aligned}
\mathbb{E}\left[e^{tY_i}\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}\right] &\leq 1 + t\mathbb{E}[Y_i\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}] + \frac{t^2}{2}\mathbb{E}\left[Y_i^2e^{t|Y_i|}\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}\right],\\
&\leq 1 + tK^2e^{1-\varkappa\frac{K}{\sigma}} + \frac{t^2}{2}\mathbb{E}\left[Y_i^2e^{\frac{\varkappa}{2\sigma}|X_i|}\mathbf{1}_{\{|X_i|\leq K\}}|\mathcal{F}_{i-1}\right],\\
&\leq 1 + tK^2e^{1-\varkappa\frac{K}{\sigma}} + \frac{t^2}{2}e\left(\frac{8}{\varkappa e}\right)^4\sigma^4.\\
&\leq \exp\left(tK^2e^{1-\varkappa(\frac{K}{\sigma})} + \frac{t^2}{2}e\left(\frac{8}{\varkappa e}\right)^4\sigma^4\right).
\end{aligned}$$

Notons $\nu_K = \min\{i \geq 1 : |X_i| > K\}$. On déduit de l'inégalité de Markov pour $\tau > 0$ que

$$\mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > \tau\right) \leq \exp\left(-t\,\tau + NtK^2e^{1-\varkappa\frac{K}{\sigma}} + \frac{Nt^2}{2}e\left(\frac{8}{\varkappa e}\right)^4\sigma^4\right).$$

Maintenant remarquons que

$$\mathbb{P}(\nu_K \leq N) \leq N\sup_i \mathbb{P}(|X_i| > K) \leq Ne^{1-\varkappa\frac{K}{\sigma}},$$

ce qui permet de conclure la preuve. □

Étudions maintenant les déviations modérées de $S_N^{(2)}$.

Proposition B.4 *Pour tout $\lambda > 0$ tel que $\lambda \geq \frac{37}{\varkappa^2 N^{\frac{1}{3}}}$, on a*

$$\mathbb{P}\left(S_N^{(2)} > N\lambda\sigma^2\right) \leq (Ne+1)\exp\left(-\frac{\varkappa}{2}\sqrt{N\lambda}\right).$$

En particulier pour tout $1\alpha > 0$, et tout $\lambda > 0$ tel que

$$\lambda \geq \max\left\{\frac{37}{\varkappa^2 N^{\frac{1}{3}}}, \frac{4}{\varkappa^2 N}\left[\ln\left((Ne+1)\alpha^{-1}\right)\right]^2\right\},$$

on a

$$\mathbb{P}\left(S_N^{(2)} > N\lambda\sigma^2\right) \leq \alpha.$$

Preuve
Grâce à la proposition B.3, on a pour tout $\tau > 0$:

$$\begin{aligned}\mathbb{P}\left(S_N^{(2)} > \tau\right) &\leq \mathbb{P}(\nu_K \leq N) + \mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > \tau\right), \\ &\leq N\, e^{1-\varkappa(\frac{K}{\sigma})} + \exp\left(-t\tau + tNK^2 e^{1-\varkappa(\frac{K}{\sigma})} + \frac{Nt^2 e}{2}\left(\frac{8\sigma}{\varkappa e}\right)^4\right),\end{aligned} \tag{B.9}$$

avec $t \leq \frac{\varkappa}{2\sigma K}$ et $K \geq \frac{2\sigma}{\varkappa}$. On choisit τ de la forme $N\lambda\sigma^2$ et $t = \frac{\varkappa}{2\sigma K}$. Alors

$$\begin{aligned}\mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > N\lambda\sigma^2\right) &\leq \exp\left(-tN\lambda\sigma^2 + tNK^2 e^{1-\varkappa(\frac{K}{\sigma})} + \frac{Nt^2 e}{2}\left(\frac{8\sigma}{\varkappa e}\right)^4\right), \\ &\leq \exp\left(-\frac{\varkappa\sigma\lambda}{2K}N + \frac{\varkappa e}{2\sigma}NKe^{-\varkappa(\frac{K}{\sigma})} + \frac{8Ne}{K^2}\left(\frac{8\sigma}{\varkappa e^2}\right)^2\right).\end{aligned}$$

Prenons maintenant $K = \frac{\sigma}{2}\sqrt{N\lambda}$, de telle sorte que

$$\mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > N\lambda(\alpha)\sigma^2\right) \leq \exp\left(-\varkappa\sqrt{N\lambda} + \frac{\varkappa e}{4}N\sqrt{N\lambda}e^{-\frac{\varkappa}{2}\sqrt{N\lambda}} + \frac{(8e^{-1})^4}{2\lambda\varkappa^2}e\right). \tag{B.10}$$

Par le lemme B.1, on a

$$\begin{aligned}\frac{\varkappa e}{4}N\sqrt{N\lambda}e^{-\frac{\varkappa}{2}\sqrt{N\lambda}} &\leq \frac{\varkappa e}{4\lambda}(N\lambda)^{\frac{3}{2}}e^{-\frac{\varkappa}{2}\sqrt{N\lambda}}, \\ &\leq \frac{\varkappa e}{4\lambda}\left(\frac{6}{\varkappa e}\right)^3 = \frac{54}{\varkappa^2 e^2\lambda},\end{aligned}$$

si bien que (B.10) permet d'obtenir

$$\mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > N\lambda(\alpha)\sigma^2\right) \leq \exp\left(-\varkappa\sqrt{N\lambda} + \frac{110}{\varkappa^2\lambda}\right).$$

Quand $\lambda \geq \frac{37}{\varkappa^2 N^{\frac{1}{3}}}$, on obtient $\frac{110}{\varkappa\lambda} \leq \frac{\varkappa}{2}\sqrt{N\lambda}$ et on déduit que

$$\mathbb{P}\left(S_N^{(2)}\mathbf{1}_{\{N<\nu_K\}} > N\lambda\sigma^2\right) \leq \exp\left(-\frac{\varkappa}{2}\sqrt{N\lambda}\right).$$

On déduit alors de (B.9)

$$\mathbb{P}\left(S_N^{(2)} > N\lambda\sigma^2\right) \leq (Ne+1)\exp\left(-\frac{\varkappa}{2}\sqrt{N\lambda}\right).$$

Un calcul simple permet alors de vérifier que si λ vérifie les conditions données dans la proposition, on a

$$\mathbb{P}\left(S_N^{(2)} > N\lambda\sigma^2\right) \leq \alpha.$$

□

B.3 Borne avec un moment exponentiel d'ordre 2

On veut maintenant montrer des inégalités de déviations similaires quand les hypothèses B.1 et B.3 sont vérifiées. On commence par donner une borne pour les déviations de $S_N^{(1)}$.

Proposition B.5 *Pour tout* $\lambda > 0$,

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{\varkappa\lambda^2}{4}\right), \tag{B.11}$$

Notons que (B.11) implique en particulier que pour tout $\lambda > 0$,

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sigma\sqrt{N}\right) \leq \exp\left(-\frac{\varkappa\lambda^2}{4}\right).$$

Preuve
Remarquons d'abord que $x \in \mathbb{R}$, $e^x \leq x + e^{x^2}$. On conclut que pour $0 < t \leq \frac{\sqrt{\varkappa}}{\sigma_i}$,

$$\mathbb{E}[\exp(tX_i)|\mathcal{F}_{i-1}] \leq \mathbb{E}[\exp(t^2X_i^2)|\mathcal{F}_{i-1}] \leq \exp(t^2\sigma_i^2).$$

D'autre part,

$$tX_i \leq \frac{t^2\sigma_i^2}{\varkappa 2} + \frac{\varkappa X_i}{2\sigma_i^2}, \text{ et } \mathbb{E}[\exp(tX_i)|\mathcal{F}_{i-1}] \leq \exp\left(\frac{t^2\sigma_i^2}{2\varkappa} + \frac{1}{2}\right) \leq \exp\left(\frac{t^2\sigma_i^2}{\varkappa}\right),$$

lorsque $t \geq \frac{\sqrt{\varkappa}}{\sigma_i}$. On obtient alors par conditionnements successifs pour tout $t > 0$

$$\mathbb{E}\left[\exp\left(t\sum_{i=1}^N X_i\right)\right] \leq \exp\left(\frac{t^2\sum_{i=1}^N \sigma_i^2}{\varkappa}\right). \tag{B.12}$$

Si bien que

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \mathbb{E}\left[\exp\left(tS_N^{(1)} - t\lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right)\right].$$

En particulier pour

$$t = \frac{\varkappa\lambda}{2\sqrt{\sum_{i=1}^N \sigma_i^2}}$$

on a

$$\mathbb{P}\left(S_N^{(1)} \geq \lambda\sqrt{\sum_{i=1}^N \sigma_i^2}\right) \leq \exp\left(-\frac{\varkappa\lambda^2}{4}\right).$$

□

On donne maintenant une borne pour les déviations de $S_N^{(2)}$

Proposition B.6 *Soit* $\lambda > 0$

$$\mathbb{P}\left(S_N^{(2)} \geq \lambda\sigma^2 N\right) \leq \begin{cases} \exp\left(-\frac{5\varkappa^2\lambda^2 N}{28}\right) & \textit{pour} \quad 0 < \lambda \leq \frac{7}{5\varkappa}; \\ \exp\left(-\frac{\varkappa N}{2}(\lambda - \frac{7}{10\varkappa})\right) & \textit{pour} \quad \lambda > \frac{7}{5\varkappa}. \end{cases}$$

En particulier pour tout $\alpha \geq \exp\left(-\frac{\varkappa N}{4}\right)$ *et pour tout* $\lambda \geq \sqrt{\frac{28}{5\varkappa^2 N}\ln(\alpha^{-1})}$, *on a*

$$\mathbb{P}\left(S_N^{(2)} \geq \lambda\sum_{i=1}^N \sigma_i^2\right) \leq \alpha.$$

Preuve
Soit $t > 0$. Nous allons appliquer le corollaire B.2 avec $Y = \frac{\varkappa X_i^2}{\sigma_i^2}$ et $s = \frac{t\sigma_i^2}{\varkappa}$. Pour cela, il nous faut calculer une borne pour $E[X_i^4|\mathcal{F}_{i-1}]$. On déduit du corollaire B.1

$$\mathbb{E}[X_i^4|\mathcal{F}_{i-1}] \leq \left(\frac{4}{e}\right)\frac{\sigma_i^4}{\varkappa^2}.$$

On obtient alors pour tout $0 < t < \frac{\varkappa}{2\sigma_i^2}$

$$\begin{aligned} \mathbb{E}\left[\exp(tX_i^2)|\mathcal{F}_{i-1}\right] &\leq 1 + t\sigma_i^2 + \frac{2\sigma_i^4}{e}t^2 + \frac{2e\,\sigma_i^6 t^3}{\sqrt{6\pi}} \leq 1 + \sigma_i^2 t + \frac{2\sigma_i^4}{\varkappa^2 e}t^2 + \frac{e\,\sigma_i^4 t^2}{\varkappa^2\sqrt{6\pi}}, \\ &\leq 1 + \sigma_i^2 t + \frac{7\sigma_i^4}{5\varkappa^2}t^2 \leq \exp\left(\sigma_i^2 t + \frac{7\sigma_i^4}{5\varkappa^2}t^2\right). \end{aligned}$$

On obtient alors par conditionnements successifs que pour $0 < t \leq \frac{\varkappa}{2\sigma^2}$

$$\mathbb{E}\left[\exp\left(t\sum_{i=1}^N X_i^2\right)\right] \leq \exp\left(t\sum_{i=1}^N \sigma_i^2 + \frac{7t^2}{5\varkappa^2}\sum_{i=1}^N \sigma_i^4\right),$$

et par conséquent que

$$\mathbb{E}\left[\exp\left(tS_N^{(2)}\right)\right] \leq \exp\left(\frac{7t^2}{5\varkappa^2}\sum_{i=1}^N \sigma_i^4\right),$$

Tout est maintenant en place pour finir la preuve : en effet, par inégalité de Markov, pour $\lambda > 0$ et $0 < t \leq \frac{\varkappa}{2\sigma^2}$ on a :

$$\begin{aligned} \mathbb{P}\left(S_N^{(2)} \geq \lambda \sum_{i=1}^N \sigma_i^2\right) &\leq \mathbb{E}\left[\exp\left(tS_N^{(2)} - t\lambda \sum_{i=1}^N \sigma_i^2\right)\right], \\ &\leq \exp\left(-t\lambda \sum_{i=1}^N \sigma_i^2 + \frac{7t^2}{5\varkappa^2} \sum_{i=1}^N \sigma_i^4\right). \end{aligned} \tag{B.13}$$

Notons

$$\lambda^* = \frac{7\sum_{i=1}^N \sigma_i^4}{5\varkappa\sigma^2 \sum_{i=1}^N \sigma_i^2}.$$

Pour $0 < \lambda \leq \lambda^*$, on choisit

$$t = \frac{5\varkappa^2 \lambda \sum_{i=1}^N \sigma_i^2}{14\sum_{i=1}^N \sigma_i^4}$$

dans (B.13) et on obtient

$$\mathbb{P}\left(S_N^{(2)} \geq \lambda \sum_{i=1}^N \sigma_i^2\right) \leq \exp\left(-\frac{5\varkappa^2\lambda^2 \left(\sum_{i=1}^N \sigma_i^2\right)^2}{28\sum_{i=1}^N \sigma_i^4}\right).$$

Pour $\lambda \geq \lambda^*$, on choisit $t = \frac{\varkappa}{2\sigma^2}$ et on obtient

$$\mathbb{P}\left(S_N^{(2)} \geq \lambda \sum_{i=1}^N \sigma_i^2\right) \leq \exp\left(-\frac{\varkappa}{2\sigma^2} \sum_{i=1}^N \sigma_i^2 \left(\lambda - \frac{\lambda^*}{2}\right)\right).$$

Le résultat annoncé dans la proposition s'en déduit. □

Annexe C

Preuves diverses

C.1 Preuve des bornes pour l'algorithme de descente en miroir

C.1.1 Lemme technique préliminaire

On se place sur un ensemble convexe compact Q inclus dans un espace vectoriel de dimension finie $(E, \|\cdot\|)$ de dual $(E^*, \|\cdot\|_*)$. On se donne deux suites de variables aléatoire $(x_i)_{i\geq 0}$ et $(\eta_i)_{i\geq 0}$ à valeurs dans Q et E^* respectivement où x_0 est une constante fixée et x_{i+1} est une fonction de η_i. On définit une famille de tribu $\mathcal{F}_i$ par $\mathcal{F}_{-1} = \{\varnothing, \Omega\}$, et pour $i \geq 0$, $\mathcal{F}_i$ est la tribu engendrée par $\mathcal{F}_{i-1}$ et η_i de sorte que x_i est $\mathcal{F}_{i-1}$-mesurable. On suppose connues une fonction fortement convexe V sur Q de paramètre α_0 dont le minimum est atteint en x_0 et vaut 0 ainsi que ses β-transformées W_β données pour tout $z \in E^*$ par

$$W_\beta(z) = \max_{x\in Q} \{\langle z, x\rangle - \beta V(x)\}.$$

On suppose que pour tout i, $\mathbb{E}[\eta_i|\mathcal{F}_{i-1}] = 0$ et $\mathbb{E}[\| \eta_i \|_*^2 |\mathcal{F}_{i-1}] \leq \sigma^2$ p.s. On note $\overline{V} = \max_{x\in Q} V(x)$.

Lemme C.1 *Soient $\beta > 0$ et $(\gamma_i)_i$ une suite de réels strictement positifs. Alors*

$$\begin{aligned}
\max_{x\in Q} \sum_{i=1}^N \gamma_i \langle \eta_i, x\rangle &\leq \beta\overline{V} + \sum_{i=1}^N \gamma_i \left\langle \eta_i, W'_\beta\left(\sum_{k=1}^{i-1} \gamma_k \eta_k\right)\right\rangle + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha_0\beta} \| \eta_i \|_*^2, \\
\mathbb{E}\left[\max_{x\in Q} \sum_{i=1}^N \gamma_i \langle \eta_i, x\rangle\right] &\leq 2\sigma\sqrt{\frac{\overline{V}\sum_{i=1}^N \gamma_i^2}{2\alpha_0}}.
\end{aligned}$$

Preuve
On a

$$\begin{aligned}
W_\beta\left(\sum_{i=1}^N \gamma_i\eta_i\right) &= \max_{x\in Q}\left(\sum_{i=1}^N \gamma_i\langle\eta_i, x\rangle - \beta V(x)\right), \\
&\geq \max_{x\in Q}\left(\sum_{i=1}^N \gamma_i\langle\eta_i, x\rangle\right) - \beta \max_{x\in Q} V(x).
\end{aligned}$$

De plus

$$\begin{aligned}
W_\beta\left(\sum_{i=1}^N \gamma_i\eta_i\right) &\leq W_\beta\left(\sum_{i=1}^{N-1} \gamma_i\eta_i\right) + \gamma_N \left\langle \eta_N, W'_\beta\left(\sum_{k=1}^{N-1} \gamma_k\eta_k\right)\right\rangle + \frac{\gamma_N^2}{2\alpha_0\beta} \| \eta_N \|_*^2, \\
&\leq W_\beta(0) + \sum_{i=1}^N \gamma_i \left\langle \eta_i, W'_\beta\left(\sum_{k=1}^{i-1} \gamma_k\eta_k\right)\right\rangle + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha_0\beta} \| \eta_i \|_*^2.
\end{aligned}$$

Comme $W_\beta(0) = 0$, il vient

$$\max_{x\in Q} \sum_{i=1}^N \gamma_i\langle\eta_i, x\rangle \leq \beta\overline{V} + \sum_{i=1}^N \gamma_i \left\langle \eta_i, W'_\beta\left(\sum_{k=1}^{i-1} \gamma_k\eta_k\right)\right\rangle + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha_0\beta} \| \eta_i \|_*^2.$$

Comme $W'_\beta\left(\sum_{k=1}^{i-1}\gamma_k\eta_k\right)$ est $\mathcal{F}_{i-1}$-mesurable, on obtient par conditionnements successifs :

$$\begin{aligned}\mathbb{E}\left[\max_{x\in Q}\sum_{i=1}^{N}\gamma_i\langle\eta_i,x\rangle\right] &\leq \beta\overline{V}+\sum_{i=1}^{N}\frac{\gamma_i^2\sigma^2}{2\alpha_0\beta},\\ &\leq 2\sigma\sqrt{\frac{\overline{V}\sum_{i=1}^{N}\gamma_i^2}{2\alpha_0}},\end{aligned}$$

si on choisit $\beta=\sigma\sqrt{\frac{\sum_{i=1}^{N}\gamma_i^2}{2\alpha_0\overline{V}}}$. □

C.1.2 Cas des inégalités variationelles à opérateur monotone

Preuve de la proposition 2.4

On veut borner l'erreur $\epsilon(x^N)=\max_{x\in Q}\{\langle g(x),x^N-x\rangle\}$. Montrons d'abord que ϵ est une 'bonne' fonction d'erreur.

Lemme C.2 *La fonction ϵ est bien définie et convexe sur Q. Pour tout $x\in Q$, $\epsilon(x)\geq 0$ avec égalité si et seulement si x est solution du problème variationel.*

Preuve
ϵ est bien définie puisque Q est supposé borné, ϵ est clairement positive ($\epsilon(x)\geq\langle g(x),x-x\rangle=0$ et $\epsilon(x)=0$ est équivalent à (2.1). □

On a par monotonie de g,

$$\begin{aligned}\epsilon(x^N) &= \max_{x\in Q}\langle g(x),x^N-x\rangle,\\ &= \left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\max_{x\in Q}\sum_{i=1}^{N}\gamma_i\langle g(x),x_i-x\rangle,\\ &\leq \left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\max_{x\in Q}\sum_{i=1}^{N}\gamma_i\langle g(x_i),x_i-x\rangle.\end{aligned}$$

Pour tout $x\in Q$ on a alors

$$\sum_{i=1}^{N}\gamma_i\langle g(x_i),x_i-x\rangle=\sum_{i=1}^{N}\gamma_i\langle\Psi(x_i),x_i-x\rangle-\sum_{i=1}^{N}\gamma_i\langle\eta(x_i),x_i-x\rangle$$

D'autre part on a pour tout i,

$$\begin{aligned}W_{\beta_i}(\zeta_i)-W_{\beta_{i-1}}(\zeta_{i-1}) &\leq W_{\beta_{i-1}}(\zeta_i)-W_{\beta_{i-1}}(\zeta_{i-1}),\\ &\leq \langle W'_{\beta_{i-1}}(\zeta_{i-1}),\zeta_i-\zeta_{i-1}\rangle+\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\zeta_i-\zeta_{i-1}\parallel_*^2,\\ &\leq -\gamma_i\langle\Psi(x_{i-1}),x_{i-1}\rangle+\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2.\end{aligned}$$

On en déduit

$$\begin{aligned}
\sum_{i=1}^{N}\gamma_i\langle g(x_i),x_i-x\rangle &\leq -\sum_{i=1}^{N}\gamma_i\langle\eta(x_i),x_i-x\rangle-\sum_{i=1}^{N}\gamma_i\langle\Psi(x_{i-1}),x_{i-1}\rangle\\
&\quad+\sum_{i=1}^{N}W_{\beta_i}(\zeta_i)-W_{\beta_{i-1}}(\zeta_{i-1})+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2,\\
&\leq -\sum_{i=1}^{N}\gamma_i\langle\eta(x_i),x_i-x\rangle+\langle x,\zeta_N\rangle+W_{\beta_0}(\zeta_0)-W_{\beta_N}(\zeta_N)+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2,\\
&\leq \beta_N V(x)-\sum_{i=1}^{N}\gamma_i\langle\eta(x_i),x_i-x\rangle+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2,
\end{aligned}$$

puisque $W_{\beta_0}(\zeta_0)=0$ et $\langle x,\zeta_N\rangle-W_{\beta_N}(\zeta_N)\leq\beta_N V(x)$. On a donc obtenu la borne suivante :

$$\left(\sum_{i=1}^{N}\gamma_i\right)\epsilon(x^N)\leq\max_{x\in Q}\beta_N V(x)+\max_{x\in Q}\sum_{i=1}^{N}\gamma_i\langle\eta(x_i),x-x_i\rangle+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2.$$

Preuve de la proposition 2.5

D'après le lemme C.1 , on a

$$\mathbb{E}[\epsilon(x^N)]\leq\left(\sum_{i=1}^{N}\gamma_i\right)^{-1}\left[\beta_N\overline{V}+2\sigma\sqrt{\frac{\overline{V}\sum_{i=1}^{N}\gamma_i^2}{2\alpha_0}}+\sum_{i=1}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_{i-1}}(M+\sigma)^2\right]. \tag{C.1}$$

On cherche maintenant les suites $(\gamma_i)_i$ et $(\beta_i)_i$ qui minimisent notre majorant de $\mathbb{E}[\epsilon(x^N)]$, sachant qu'elles doivent être strictement positives et que la seconde doit être croissante. Le troisième terme de (C.1) étant maximal si β_{i-1} est le plus grand possible, on prendra donc $(\beta_i)_i$ constante. Ensuite, si $\sum_{i=1}^{N}\gamma_i$ est fixée, le second et le troisième terme de C.1 sont minimaux quand la suite $(\gamma_i)_i$ est constante.
On a donc

$$\mathbb{E}[\epsilon(x^N)]\leq\frac{\beta_N\overline{V}}{\gamma_N N}+2\sigma\sqrt{\frac{\overline{V}}{2\alpha_0 N}}+\frac{\gamma_N(M+\sigma)^2}{2\beta_N\alpha_0}.$$

En choisissant

$$\frac{\beta_N}{\gamma_N}=\frac{(M+\sigma)\sqrt{N}}{\sqrt{2\alpha_0\overline{V}}},$$

on obtient

$$\mathbb{E}[\epsilon(x^N)]\leq\frac{2\sqrt{\overline{V}}(M+\sigma)}{\sqrt{2\alpha_0 N}}+2\sigma\sqrt{\frac{\overline{V}}{2\alpha_0 N}}.$$

Si on ne se donne pas à l'avance le nombre de pas que l'on va effectuer, on peut choisir $\gamma_i\equiv 1$ et $\beta_i=\frac{(M+\sigma)\sqrt{i+1}}{\alpha_0\overline{V}}$. On a alors

$$\mathbb{E}[\epsilon(x^N)]\leq\frac{2(M+\sigma)\sqrt{N+1}}{\sqrt{\alpha_0\overline{V}}N}+2\sigma\sqrt{\frac{\overline{V}}{2\alpha_0 N}}.$$

Preuve de la proposition 2.6

D'après la proposition 2.4 et le lemme C.1, on a

$$\begin{aligned}\epsilon(x^N) &\leq \frac{1}{N}\left(\beta_N\overline{V} + \max_{x\in Q}\sum_{i=1}^{N}\langle\eta(x_{i-1}),\, x - x_{i-1}\rangle + \sum_{i=1}^{N}\frac{1}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2\right),\\ &\leq \frac{1}{N}\left((\beta_N+\beta)\overline{V} + \sum_{i=1}^{N}\langle\eta(x_{i-1}),\, \check{x}_{i-1} - x_{i-1}\rangle + \sum_{i=1}^{N}\left(\frac{1}{2\alpha_0\beta_{i-1}} \parallel \Psi(x_{i-1}) \parallel_*^2 + \frac{1}{2\alpha_0\beta} \parallel \eta(x_{i-1}) \parallel_*^2\right)\right) \quad (C.2)\end{aligned}$$

où $\check{x}_i = W'_\beta\left(\sum_{k=1}^{i}\eta(x_{k-1})\right)$ et $\beta = \sigma\sqrt{\frac{N}{2\alpha_0\overline{V}}}$. Il vient alors

$$\begin{aligned}\epsilon(x^N) &\leq \frac{1}{N}\left((M+2\sigma)\sqrt{\frac{N\overline{V}}{2\alpha_0}} + \sum_{i=1}^{N}\langle\eta(x_{i-1}),\, \check{x}_{i-1} - x_{i-1}\rangle\right.\\ &\qquad\left. + \sum_{i=1}^{N}\sqrt{\frac{\overline{V}}{2\alpha_0 N}}\left(\frac{2M^2}{M+\sigma} + \frac{2 \parallel \eta(x_i-1) \parallel_*^2}{M+\sigma} + \frac{2 \parallel \eta(x_i-1) \parallel_*^2}{\sigma}\right)\right),\\ &\leq \frac{M+2\sigma}{\sqrt{N}}\sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{1}{N}\sum_{i=1}^{N}\langle\eta(x_{i-1}),\, \check{x}_{i-1} - x_{i-1}\rangle + \frac{2M^2}{(M+\sigma)\sqrt{N}}\sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{3}{N\sigma}\sqrt{\frac{\overline{V}}{2\alpha_0 N}} \parallel \eta(x_i-1) \parallel_*^2 .\end{aligned}$$

Il suffit alors appliquer la proposition B.2 à la suite de variables aléatoires $\frac{1}{D}\langle\eta(x_{i-1}),\, \check{x}_{i-1} - x_{i-1}\rangle$ et la proposition B.4 à la suite $\eta(x_{i-1})$ pour obtenir le résultat annoncé.

La proposition 2.7 se démontre de manière analogue.

C.1.3 Cas des problèmes de point selle

Preuve de la proposition 2.8

Compte tenu du schéma de l'algorithme, on a pour tout $i \geq 1$

$$\begin{aligned}W_{X,\beta_{i-1}}(\zeta_i) &\leq W_{X,\beta_{i-1}}(\zeta_{i-1}) + \langle\nabla W_{X,\beta_{i-1}}(\zeta_{i-1}), \zeta_i - \zeta_{i-1}\rangle + \frac{1}{2\alpha_{X,0}\beta_{X,i-1}} \parallel \zeta_i - \zeta_{i-1} \parallel_*^2,\\ &\leq W_{X,\beta_{i-1}}(\zeta_{i-1}) - \gamma_i\langle x_{i-1}, \Phi(x_{i-1}, y_{i-1})\rangle + \frac{\gamma_i^2}{2\alpha_{X,0}\beta_{X,i-1}} \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_*^2 .\end{aligned}$$

En utilisant la dernière inégalité, le fait que $(\beta_{X,i})_{i\geq 0}$ est une suite décroissante et que, pour tout z fixé, $\beta \to W_\beta(z)$ es une fonction décroissante, il vient

$$W_{X,\beta_i}(\zeta_i) - W_{X,\beta_{i-1}}(\zeta_{i-1}) \leq -\gamma_i\langle x_{i-1}, \Phi(x_{i-1}, y_{i-1})\rangle + \frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}}.$$

Puisque $\Phi(x_i, y_i) = L_X(x, y) + \eta_X(x, y)$,on a pour tout $x\ inX$

$$\begin{aligned}\gamma_i\langle(x_{i-1} - x), L_X(x_{i-1}, y_{i-1})\rangle &\leq -\gamma_i\langle x, \Phi(x_{i-1}, y_{i-1})\rangle - \gamma_i\langle(x_{i-1} - x), \eta_X(x_{i-1}, y_{i-1})\rangle\\ &\quad + W_{X,\beta_{X,i-1}}(\zeta_{i-1}) - W_{X,\beta_{X,i}}(\zeta_i) + \frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}}.\end{aligned}$$

En sommant et en utilisant que $\sum_{i=1}^{N} -\gamma_i \Phi(x_{i-1}, y_{i-1}) = \zeta_N$ et $W_{X,\beta_0}(\zeta_0) = 0$, on obtient

$$\sum_{i=1}^{N} \gamma_i \langle (x_{i-1} - x), L'_X(x_{i-1}, y_{i-1}) \rangle \leq \langle x, \zeta_N \rangle - W_{X,\beta_{X,N}}(\zeta_N) - \sum_{i=1}^{N} \gamma_i \langle (x_{i-1} - x), \eta_X(x_{i-1}, y_{i-1}) \rangle + \sum_{i=1}^{N} \frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}}. \quad \text{(C.3)}$$

Comme $\beta_{X,N} V_X(\cdot, x_0)$ et $W_{X,\beta_{X,N}}(\cdot)$ sont conjuguées l'une de l'autre, $\forall x \in X$, $\langle x, \zeta_N \rangle - W_{X,\beta_{X,N}}(\zeta_N) \leq \beta_{X,N} V_X(x, x_0)$. De plus, pour tout $y \in Y$, $x \to L(x, y)$ est convexe si bien que $\langle (x_{i-1} - x), L'_X(x_{i-1}, y_{i-1}) \rangle \geq L(x_{i-1}, y_{i-1}) - L(x, y_{i-1})$. D'où

$$\sum_{i=1}^{N} \gamma_i [L(x_{i-1}, y_{i-1}) - L(x, y_{i-1})] \leq \beta_{X,N} V_X(x, x_0) - \sum_{i=1}^{N} \gamma_i \langle (x_{i-1} - x), \eta_X(x_{i-1}, y_{i-1}) \rangle + \sum_{i=1}^{N} \frac{\gamma_i^2 \parallel \Phi(x_{i-1}, y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}}. \quad \text{(C.4)}$$

De manière analogue à ce qui précède, on obtient

$$\sum_{i=1}^{N} -\gamma_i \langle y_i - y, L_Y(x_i, y_i) \rangle \leq \langle y, \xi_N \rangle - W_{Y,\beta_{Y,N}}(\xi_N) + \sum_{i=1}^{N} \gamma_i \langle y_i - y, \eta_Y(x_i, y_i) \rangle + \sum_{i=1}^{N} \frac{\gamma_i^2 \parallel \Psi(x_i, y_i) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}}.$$

Et finalement puisque pour tout $x \in X$, $y \to L(x, y)$ est concave, on a

$$\sum_{i=1}^{N} \gamma_i [L(x_i, y) - L(x_i, y_i)] \leq \beta_{Y,N} V_Y(y, y_0) + \sum_{i=1}^{N} \gamma_i \langle y_i - y, \eta_Y(x_i, y_i) \rangle + \sum_{i=1}^{N} \frac{\gamma_i^2 \parallel \Psi(x_i, y_i) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}}. \quad \text{(C.5)}$$

Il suffit alors d'additionner (C.4) et (C.5) pour achever la preuve.

Preuve de la proposition 2.9

Avec le lemme C.1 et l'hypothèse 2.9 et (2.8), on obtient

$$\mathbb{E}[\Delta_N] \leq \frac{\beta_{X,N}\overline{V}_X + \beta_{Y,N}\overline{V}_Y}{\sum_{i=1}^{N} \gamma_i} + \frac{\left(\sum_{i=1}^{N} \gamma_i^2\right)^{\frac{1}{2}}}{\sum_{i=1}^{N} \gamma_i} \left(\sigma_X \sqrt{\frac{2\overline{V}_X}{\alpha_{X,0}}} + \sigma_Y \sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}} \right) + \frac{1}{\sum_{i=1}^{N} \gamma_i} \sum_{i=1}^{N} \frac{\gamma_i^2}{2} \left[\frac{(M_X + \sigma_X)^2}{\alpha_{X,0}\beta_{X,i-1}} + \frac{(M_Y + \sigma_Y)^2}{\alpha_{Y,0}\beta_{Y,i-1}} \right].$$

En choisissant $\gamma_i \equiv 1$, $\beta_{X,i} \equiv \frac{(M_X + \sigma_X)\sqrt{N}}{\sqrt{2\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} \equiv \frac{(M_Y + \sigma_Y)\sqrt{N}}{\sqrt{2\alpha_{Y,0}\overline{V}_Y}}$, on obtient

$$\mathbb{E}[\Delta_N] \leq \frac{2\sqrt{\overline{V}_X}(M_X + \sigma_X)}{\sqrt{2\alpha_{X,0}N}} + \frac{2\sqrt{\overline{V}_Y}(M_Y + \sigma_Y)}{\sqrt{2\alpha_{Y,0}N}} + 2\sigma_X \sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}N}} + 2\sigma_Y \sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}N}}.$$

Si le nombre de pas à effectuer dans l'algorithme n'est pas connu à l'avance, en posant $\gamma_i \equiv 1$, $\beta_{X,i} = \frac{(M_X+\sigma_X)\sqrt{i+1}}{\sqrt{\alpha_{X,0}\overline{V}_X}}$ et $\beta_{Y,i} \equiv \frac{(M_Y+\sigma_Y)\sqrt{i+1}}{\sqrt{\alpha_{Y,0}\overline{V}_Y}}$, on obtient de manière analogue

$$\mathbb{E}[\Delta_N] \leq \frac{2(M_X+\sigma_X)\sqrt{N+1}}{\sqrt{\alpha_{X,0}\overline{V}_X}N} + \frac{2(M_Y+\sigma_Y)\sqrt{N+1}}{\sqrt{\alpha_{Y,0}\overline{V}_Y}N} + 2\sigma_X\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}N}} + 2\sigma_Y\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}N}}.$$

Preuve de la proposition 2.10

On a d'après (2.8) et le lemme C.1

$$N\Delta_N \leq (\beta_{X,N}+\beta_X)\overline{V}_X+(\beta_{Y,N}+\beta_Y)\overline{V}_Y-\sum_{i=1}^{N}\langle x_{i-1}-\check{x}_{i-1}, \eta_X(x_{i-1},y_{i-1})\rangle+\sum_{i=1}^{N}\langle y_{i-1}-\check{y}_{i-1}, \eta_Y(x_{i-1},y_{i-1})\rangle$$
$$+\sum_{i=1}^{N}\left(\frac{\parallel \Phi(x_{i-1},y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}} + \frac{\parallel \eta_X(x_{i-1},y_{i-1}) \parallel_{X,*}^2}{2\alpha_{X,0}\beta_X} + \frac{\parallel \Psi(x_{i-1},y_{i-1}) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}} + \frac{\parallel \eta_Y(x_{i-1},y_{i-1}) \parallel_{Y,*}^2}{2\alpha_{Y,0}\beta_Y}\right),$$

où $\check{x}_i = W'_{\beta_X}\left(\sum_{k=1}^{i}\eta_X(x_{k-1},y_{k-1})\right)$, $\check{y}_i = W'_{\beta_Y}\left(-\sum_{k=1}^{i}\eta_Y(x_{k-1},y_{k-1})\right)$, $\beta_X = \sigma_X\sqrt{\frac{N}{2\alpha_{X,0}\overline{V}_X}}$ et $\beta_Y = \sigma_Y\sqrt{\frac{N}{2\alpha_{Y,0}\overline{V}_Y}}$.
D'où

$$N\Delta_N \leq (M_X+2\sigma_X)\sqrt{\frac{N\overline{V}_X}{2\alpha_{X,0}}} + (M_Y+2\sigma_Y)\sqrt{\frac{N\overline{V}_Y}{2\alpha_{Y,0}}} - D_X\sum_{i=1}^{N}\langle\frac{x_{i-1}-\check{x}_{i-1}}{D_X}, \eta_X(x_{i-1},y_{i-1})\rangle$$
$$+ D_Y\sum_{i=1}^{N}\langle\frac{y_{i-1}-\check{y}_{i-1}}{D_Y}, \eta_Y(x_{i-1},y_{i-1})\rangle + \frac{2M_X^2}{M_X+\sigma_X}\sqrt{\frac{N\overline{V}_X}{2\alpha_{X,0}}} + \frac{2M_Y^2}{M_Y+\sigma_Y}\sqrt{\frac{N\overline{V}_Y}{2\alpha_{Y,0}}}$$
$$+\sum_{i=1}^{N}\left(\frac{3}{\sigma_X}\sqrt{\frac{\overline{V}_X}{2N\alpha_{X,0}}} \parallel \eta_X(x_{i-1},y_{i-1}) \parallel_{X,*}^2 + \frac{3}{\sigma_Y}\sqrt{\frac{\overline{V}_Y}{2N\alpha_{Y,0}}} \parallel \eta_Y(x_{i-1},y_{i-1}) \parallel_{Y,*}^2\right).$$

Il suffit alors appliquer la proposition B.2 aux suites de variables aléatoires de terme général $\frac{-1}{D_X}\langle x_{i-1}-\check{x}_{i-1}, \eta_X(x_{i-1},y_{i-1})\rangle$ et $\frac{1}{D_Y}\langle y_{i-1}-\check{y}_{i-1}, \eta_Y(x_{i-1},y_{i-1})\rangle$ et la proposition B.4 aux suites de terme général $\eta_X(x_{i-1},y_{i-1})$ et $\eta_Y(x_{i-1},y_{i-1})$ pour obtenir le résultat annoncé.

La proposition 2.11 se démontre de manière analogue.

C.2 Preuve des bornes pour l'algorithme de descente en miroir à gain adaptatif

C.2.1 Lemme préliminaire

Lemme C.3 *Soit $(u_i)_i$ une suite de réels strictement positifs et $\beta_0 > 0$. Considérons les deux suites $(\beta_i^{(1)})$ et $(\beta_i^{(2)})$ données par $\beta_0^{(1)} = \beta_0^{(2)} = \beta_0$ et pour tout $i \geq 0$,*

$$\begin{aligned} \beta_{i+1}^{(1)} &= \beta_i^{(1)} + \frac{u_{i+1}}{\beta_i^{(1)}} \\ \beta_{i+1}^{(2)} &= \beta_i^{(2)} + \frac{u_i}{\beta_i^{(2)}}. \end{aligned}$$

Alors, pour tout $N > 0$,

$$\begin{aligned} \beta_N^{(1)\,2} &\leq \beta_0^2 + 2\sum_{i=1}^{N} u_i + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0}\sum_{i=1}^{N} u_i\right), \\ \beta_{N+1}^{(2)2} &\leq \beta_0^2 + 2\sum_{i=0}^{N} u_i + \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0}\sum_{i=0}^{N} u_i\right), \end{aligned}$$

où $u_N^* = \max_{i=0\cdots N} u_i$.

Preuve
On déduit immédiatement de l'équation de construction de $(\beta_i^{(1)})$ que pour tout i, $\beta_{i+1}^{(1)\,2} = \beta_i^{(1)\,2} + 2u_{i+1} + u_{i+1}^2 \beta_i^{(1)\,-2}$. En sommant ces inégalités, il vient

$$\beta_N^{(1)\,2} = \beta_0^2 + 2\sum_{i=1}^{N} u_i + \sum_{i=0}^{N-1} \frac{u_{i+1}^2}{\beta_i^{(1)\,2}}. \tag{C.6}$$

Remarquons maintenant que

$$0 \leq \sum_{i=0}^{N-1}\left(\frac{u_{i+1}^2}{\beta_i^{(1)\,2}} - \frac{u_{i+1}^2}{\beta_{i+1}^{(1)\,2}}\right) \leq u_N^{*\,2}\sum_{i=0}^{N-1}\left(\frac{1}{\beta_i^{(1)\,2}} - \frac{1}{\beta_{i+1}^{(1)\,2}}\right) \leq \frac{u_N^{*\,2}}{\beta_0^2},$$

$$\frac{2u_{i+1}}{\beta_{i+1}^{(1)\,2}} \leq \frac{2u_{i+1}}{\beta_0^2 + 2\sum_{k=1}^{i+1} u_k} \leq \ln\left(\beta_0^2 + 2\sum_{k=1}^{i+1} u_k\right) - \ln\left(\beta_0^2 + 2\sum_{k=1}^{i} u_k\right),$$

d'où

$$\sum_{i=0}^{N-1} \frac{u_{i+1}^2}{\beta_{i+1}^{(1)\,2}} \leq \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0^2}\sum_{i=1}^{N} u_i\right).$$

Par conséquent

$$\beta_N^{(1)\,2} \leq \beta_0^2 + 2\sum_{i=1}^{N} u_i + \frac{u_N^{*\,2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1 + \frac{2}{\beta_0^2}\sum_{i=1}^{N} u_i\right).$$

La preuve de la seconde inégalité utilise les mêmes arguments. □

C.2.2 Cas des problèmes d'inégalités variationnelles

Preuve de la proposition 2.16

D'après l'équation (C.2) on a

$$N\epsilon(x^N) \leq (\beta_N+\beta)\overline{V}+\sum_{i=1}^{N}\langle\eta(x_{i-1}),\check{x}_{i-1}-x_{i-1}\rangle+\sum_{i=1}^{N}\left(\frac{1}{2\alpha_0\beta_{i-1}}\parallel\Psi(x_{i-1})\parallel_*^2+\frac{1}{2\alpha_0\beta}\parallel\eta(x_{i-1})\parallel_*^2\right),$$

où $\beta=\frac{\sigma\sqrt{N}}{\sqrt{2\alpha_0\overline{V}}}$ et $\check{x}_i=W'_\beta\left(\sum_{k=1}^{i}\eta(x_{k-1})\right)$. Compte tenu de la définition de la suite $(\beta_i)_i$, il vient alors

$$N\epsilon(x^N) \leq (2\beta_N+\beta)\overline{V}+\sum_{i=1}^{N}\langle\eta(x_{i-1}),\check{x}_{i-1}-x_{i-1}\rangle+\sum_{i=1}^{N}\frac{1}{2\alpha_0\beta}\parallel\eta(x_{i-1})\parallel_*^2.$$

D'autre part, d'après le lemme C.3 et en notant $u_N^*=\max i=1\cdots Nu_i$, on a

$$\beta_N^2\leq\beta_0^2+2\sum_{i=1}^{N}u_i+\frac{u_N^{*2}}{\beta_0^2}+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right),$$

si bien que

$$\begin{aligned}N\epsilon(x^N) \leq \beta\overline{V}+\sum_{i=1}^{N}\langle\eta(x_{i-1}),\check{x}_{i-1}-x_{i-1}\rangle+\sum_{i=1}^{N}\frac{1}{2\alpha_0\beta}\parallel\eta(x_{i-1})\parallel_*^2\\ +2\overline{V}\sqrt{\beta_0^2+2\sum_{i=1}^{N}u_i+\frac{u_N^{*2}}{\beta_0^2}+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)}.\end{aligned}\quad\text{(C.7)}$$

Il nous faut maintenant majorer l'espérance des termes sous le radical.

Lemme C.4 *Pour tout $N\geq 3$*

$$\begin{aligned}\mathbb{E}[u_N^{*2}] &\leq \frac{4\sigma^4}{\alpha_0^2\overline{V}^2\varkappa^2}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln(N)\right)^2,\\ \mathbb{E}\left[u_N^*\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^{N}u_i\right)\right] &\leq \frac{2\sigma^2}{\alpha_0\overline{V}\varkappa}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{4\sigma^2}{\alpha_0\overline{V}\varkappa\beta_0^2}\right)+\ln(N)\right)^2.\end{aligned}$$

Preuve
Voir la preuve du lemme 2.1. □

On obtient alors

$$\begin{aligned}\mathbb{E}[\epsilon(x^N)] \leq \sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{2\sigma}{\sqrt{N}}+\frac{2\overline{V}\beta_0}{N}+\frac{2\overline{V}}{N}\Bigg[\frac{2N(M+\sigma)^2}{\alpha_0\overline{V}}+\frac{4\sigma^2}{\alpha_0^2\beta_0^2\overline{V}^2\varkappa^2}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln(N)\right)^2\\ +\frac{\sigma^2}{\alpha_0\overline{V}\varkappa}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{4\sigma^2}{\alpha_0\overline{V}\varkappa\beta_0^2}\right)+\ln(N)\right)^2\Bigg]^{\frac{1}{2}}.\end{aligned}$$

On utilise alors le fait que pour tout x, $y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$ pour obtenir

$$\mathbb{E}[\epsilon(x^N)] \leq \sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{(6\sigma+4M)}{\sqrt{N}} + \frac{2\overline{V}\beta_0}{N} + \sqrt{\frac{\alpha_0\overline{V}}{2N^3(M+\sigma^2)}}\left[\frac{4\sigma^2}{\alpha_0^2\beta_0^2\overline{V}\varkappa^2}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln(N)\right)^2\right.$$
$$\left.+\frac{\sigma^2}{\alpha_0\varkappa}\left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{4\sigma^2}{\alpha_0\overline{V}\varkappa\beta_0^2}\right)+\ln(N)\right)^2\right].$$

On en déduit que

$$\mathbb{E}[\epsilon(x^N)] \leq \frac{2(3\sigma+2M)}{\sqrt{N}}\sqrt{\frac{\overline{V}}{2\alpha_0}} + \frac{2\overline{V}\beta_0}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}}{2\alpha_0}}\frac{2\sigma^2}{\varkappa(M+\sigma)}\left(1+\frac{4\sigma^2}{\alpha_0\beta_0^2\overline{V}\varkappa}\right) + \frac{C}{N^{\frac{3}{2}}},$$

où

$$C = \frac{2\sigma^2}{\varkappa(M+\sigma)}\sqrt{\frac{\overline{V}}{2\alpha_0}}\left[\frac{4\sigma^2}{\alpha_0\beta_0^2\overline{V}\varkappa}\left(\varkappa\frac{M^2}{\sigma^2}+1\right)^2 + \left(\varkappa\frac{M^2}{\sigma^2}+1+\ln\left(1+\frac{4\sigma^2}{\alpha_0\overline{V}\varkappa\beta_0^2}\right)\right)^2\right].$$

Preuve de la proposition 2.17

D'après,

$$N\epsilon(x^N) \leq \beta\overline{V} + \sum_{i=1}^N \langle \eta(x_{i-1}), \check{x}_{i-1} - x_{i-1}\rangle + \sum_{i=1}^N \frac{1}{2\alpha_0\beta} \parallel \eta(x_{i-1}) \parallel_*^2$$
$$+ 2\overline{V}\sqrt{\beta_0^2 + 2\sum_{i=1}^N u_i + \frac{u_N^{*2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)},$$

où $\beta = \frac{\sigma\sqrt{N}}{\sqrt{2\alpha_0\overline{V}}}$ et $\check{x}_i = W'_\beta\left(\sum_{k=1}^i \eta(x_{k-1})\right)$, soit

$$\epsilon(x^N) \leq \frac{\sigma}{\sqrt{N}}\sqrt{\frac{2\overline{V}}{\alpha_0}} + \frac{D}{\sqrt{N}}\sum_{i=1}^N \langle \eta(x_{i-1}), \frac{\check{x}_{i-1}-x_{i-1}}{D}\rangle + \frac{1}{\sigma N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}}{2\alpha_0}}\sum_{i=1}^N\left(\parallel \eta(x_{i-1}) \parallel_*^2 - \sigma^2\right)$$
$$+ \frac{2\overline{V}}{N}\sqrt{\beta_0^2 + \frac{4(M^2+\sigma^2)N}{\alpha_0\overline{V}} + \frac{4}{\alpha_0\overline{V}}\sum_{i=1}^N(\parallel \eta(x_{i-1}) \parallel_*^2 - \sigma^2) + \frac{u_N^{*2}}{\beta_0^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right)}, \quad \text{(C.8)}$$

Les propositions B.5 et B.6 permettent de contrôler les déviations des termes $\sum_{i=1}^N \langle \eta(x_{i-1}), \frac{\check{x}_{i-1}-x_{i-1}}{D}\rangle$ et $\sum_{i=1}^N(\parallel \eta(x_{i-1}) \parallel_*^2 - \sigma^2)$. Il reste donc à borner les déviations des deux autres termes non constants de l'expression C.8, ce que nous permet de faire le lemme suivant

Lemme C.5 *On suppose que les hypothèses 2.5 et 2.8 sont vérifiées. Alors pour tout N et tout $\alpha > 0$, on a*

$$\mathbb{P}\left(u_N^{*2} \geq \left\{\frac{2\sigma^2}{\varkappa\overline{V}\alpha_0}\ln\left(\exp(\varkappa M^2\sigma^{-2})\frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha,$$

$$\mathbb{P}\left(u_N^* \ln\left(1+\frac{2}{\beta_0^2}\sum_{i=1}^N u_i\right) \geq \frac{2\sigma^2}{\varkappa\overline{V}\alpha_0}\left\{\ln\left(\left(1+\frac{4\sigma^2}{\varkappa\alpha_0\overline{V}\beta_0^2}\right)\exp(\varkappa M^2\sigma^{-2})\frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha.$$

Preuve
Voir la preuve du lemme 2.2. □

La preuve de la proposition 2.17 se déduit alors de (C.8), des propositions B.5 et B.6 et du lemme C.5.

C.2.3 Cas des problèmes de point selle

Preuve de la proposition 2.18

D'après l'équation (2.8) et le lemme C.1, on a

$$\begin{aligned}
N\Delta_N \leq (\beta_{X,N}+\beta_X)\overline{V}_X - \sum_{i=1}^N \langle \eta_X(x_{i-1},y_{i-1}), x_{i-1}-\check{x}_{i-1}\rangle \\
+ (\beta_{Y,N}+\beta_Y)\overline{V}_Y + \sum_{i=1}^N \langle \eta_Y(x_{i-1},y_{i-1}), y_{i-1}-\check{y}_{i-1}\rangle \\
+ \sum_{i=1}^N \left(\frac{\| \Phi(x_{i-1},y_{i-1}) \|_{X,*}^2}{2\alpha_{X,0}\beta_{X,i-1}} + \frac{\| \eta_X(x_{i-1},y_{i-1}) \|_{X,*}^2}{2\alpha_{X,0}\beta_X}\right) \\
+ \sum_{i=1}^N \left(\frac{\| \Psi(x_{i-1},y_{i-1}) \|_{Y,*}^2}{2\alpha_{Y,0}\beta_{Y,i-1}} + \frac{\| \eta_Y(x_{i-1},y_{i-1}) \|_{Y,*}^2}{2\alpha_{Y,0}\beta_Y}\right),
\end{aligned}$$

où

$$\begin{aligned}
\beta_X &= \sigma_X\sqrt{\tfrac{N}{2\alpha_{X,0}\overline{V}_X}}, \quad \check{x}_i = W'_{\beta_X}\left(\textstyle\sum_{k=1}^i \eta_X(x_{k-1},y_{k-1})\right), \\
\beta_Y &= \sigma_Y\sqrt{\tfrac{N}{2\alpha_{Y,0}\overline{V}_Y}}, \quad \check{y}_i = W'_{\beta_Y}\left(\textstyle\sum_{k=1}^i \eta_Y(x_{k-1},y_{k-1})\right).
\end{aligned}$$

Compte tenu de la définition de $(\beta_{X,i})_i$ et $(\beta_{Y,i})_i$, l'inégalité précédente devient

$$\begin{aligned}
N\Delta_N \leq (2\beta_{X,N}+\beta_X)\overline{V}_X - \sum_{i=1}^N \langle \eta_X(x_{i-1},y_{i-1}), x_{i-1}-\check{x}_{i-1}\rangle \\
+ (2\beta_{Y,N}+\beta_Y)\overline{V}_Y + \sum_{i=1}^N \langle \eta_Y(x_{i-1},y_{i-1}), y_{i-1}-\check{y}_{i-1}\rangle \\
+ \sum_{i=1}^N \left(\frac{\| \eta_X(x_{i-1},y_{i-1}) \|_{X,*}^2}{2\alpha_{X,0}\beta_X} + \frac{\| \eta_Y(x_{i-1},y_{i-1}) \|_{Y,*}^2}{2\alpha_{Y,0}\beta_Y}\right),
\end{aligned}$$

On note $u_N^* = \sup_{1\leq i\leq N} u_i$ et $v_N^* = \sup_{1\leq i\leq N} v_i$. D'après le lemme C.3, pour tout $\beta_{X,0} > 0$ et $\beta_{Y,0} > 0$

$$\begin{aligned}
\beta_{X,N}^2 &\leq \beta_{X,0}^2 + 2\sum_{i=1}^N u_i + \frac{{u_N^*}^2}{\beta_{X,0}^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right),\\
\beta_{Y,N}^2 &\leq \beta_{Y,0}^2 + 2\sum_{i=1}^N v_i + \frac{{v_N^*}^2}{\beta_{Y,0}^2} + \frac{v_N^*}{2}\ln\left(1+\frac{2}{\beta_{Y,0}^2}\sum_{i=1}^N v_i\right).
\end{aligned}$$

On en déduit que

$$\begin{aligned}
N\Delta_N \leq\ & \sigma_X\sqrt{\frac{N\overline{V}_X}{2\alpha_{X,0}}} + \sigma_Y\sqrt{\frac{N\overline{V}_Y}{2\alpha_{Y,0}}} - \sum_{i=1}^N\langle\eta_X(x_{i-1},y_{i-1}), x_{i-1}-\check{x}_{i-1}\rangle + \sum_{i=1}^N\langle\eta_Y(x_{i-1},y_{i-1}), y_{i-1}-\check{y}_{i-1}\rangle\\
&+\sum_{i=1}^N\left(\frac{1}{\sigma_X}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}N}}\parallel\eta_X(x_{i-1},y_{i-1})\parallel_{X,*}^2 + \frac{1}{\sigma_Y}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}N}}\parallel\eta_Y(x_{i-1},y_{i-1})\parallel_{Y,*}^2\right)\\
&+2\overline{V}_X\sqrt{\beta_{X,0}^2 + 2\sum_{i=1}^N u_i + \frac{{u_N^*}^2}{\beta_{X,0}^2} + \frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right)}\\
&+2\overline{V}_Y\sqrt{\beta_{Y,0}^2 + 2\sum_{i=1}^N v_i + \frac{{v_N^*}^2}{\beta_{Y,0}^2} + \frac{v_N^*}{2}\ln\left(1+\frac{2}{\beta_{Y,0}^2}\sum_{i=1}^N v_i\right)}.
\end{aligned} \tag{C.9}$$

D'autre part,

Lemme C.6 *Pour tout $N \geq 3$*

$$\begin{aligned}
\mathbb{E}[{u_N^*}^2] &\leq \frac{4\sigma_X^4}{\alpha_{X,0}^2\overline{V}_X^2\varkappa^2}\left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1 + \ln(N)\right)^2,\\
\mathbb{E}\left[u_N^*\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right)\right] &\leq \frac{2\sigma_X^2}{\alpha_{X,0}\overline{V}_X\varkappa}\left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1 + \ln\left(1+\frac{4\sigma_X^2}{\alpha_{X,0}\overline{V}_X\varkappa\beta_{X,0}^2}\right) + \ln(N)\right)^2,\\
\mathbb{E}[{v_N^*}^2] &\leq \frac{4\sigma_Y^4}{\alpha_{Y,0}^2\overline{V}_Y^2\varkappa^2}\left(\varkappa\frac{M_Y^2}{\sigma_Y^2} + 1 + \ln(N)\right)^2,\\
\mathbb{E}\left[v_N^*\ln\left(1+\frac{2}{\beta_{Y,0}^2}\sum_{i=1}^N v_i\right)\right] &\leq \frac{2\sigma_Y^2}{\alpha_{Y,0}\overline{V}_Y\varkappa}\left(\varkappa\frac{M_Y^2}{\sigma_Y^2} + 1 + \ln\left(1+\frac{4\sigma_Y^2}{\alpha_{Y,0}\overline{V}_Y\varkappa\beta_{Y,0}^2}\right) + \ln(N)\right)^2.
\end{aligned}$$

Preuve
Voir la preuve du lemme 2.1. □

D'où

$$
\begin{aligned}
\mathbb{E}\left[\Delta_N\right] \leq & \frac{2\sigma_X}{\sqrt{N}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + \frac{2\sigma_Y}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} \\
& + \frac{2\overline{V}_X}{N}\left[\beta_{X,0}^2 + \frac{2N(M_X+\sigma_X)^2}{\alpha_{X,0}\overline{V}_X} + \frac{4\sigma_X^4}{\alpha_{X,0}^2\overline{V}_X^2\varkappa^2\beta_{X,0}^2}\left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1 + \ln(N)\right)^2\right. \\
& \left. + \frac{\sigma_X^2}{\alpha_{X,0}\overline{V}_X\varkappa}\left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1 + \ln\left(1 + \frac{4\sigma_X^2}{\alpha_{X,0}\overline{V}_X\varkappa\beta_{X,0}^2}\right) + \ln(N)\right)^2\right]^{\frac{1}{2}} \\
& + \frac{2\overline{V}_Y}{N}\left[\beta_{Y,0}^2 + \frac{2N(M_Y+\sigma_Y)^2}{\alpha_{Y,0}\overline{V}_Y} + \frac{4\sigma_Y^4}{\alpha_{Y,0}^2\overline{V}_Y^2\varkappa^2\beta_{Y,0}^2}\left(\varkappa\frac{M_Y^2}{\sigma_Y^2} + 1 + \ln(N)\right)^2\right. \\
& \left. + \frac{\sigma_Y^2}{\alpha_{Y,0}\overline{V}_Y\varkappa}\left(\varkappa\frac{M_Y^2}{\sigma_X^2} + 1 + \ln\left(1 + \frac{4\sigma_Y^2}{\alpha_{Y,0}\overline{V}_Y\varkappa\beta_{Y,0}^2}\right) + \ln(N)\right)^2\right]^{\frac{1}{2}}.
\end{aligned}
$$

On utilise alors le fait que pour tout x, $y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \frac{y}{2\sqrt{x}}$ pour obtenir

$$
\begin{aligned}
\mathbb{E}[\Delta_N] \leq & \frac{2(3\sigma_X + 2M_X)}{\sqrt{N}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}} + \frac{2\overline{V}_X\beta_{X,0}}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\frac{2\sigma_X^2}{\varkappa(M_X+\sigma_X)}\left(1 + \frac{4\sigma_X^2}{\alpha_{X,0}\beta_{X,0}^2\overline{V}_X\varkappa}\right) \\
& + \frac{2(3\sigma_Y + 2M_Y)}{\sqrt{N}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}} + \frac{2\overline{V}_Y\beta_{Y,0}}{N} + \frac{\ln(N)^2}{N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\frac{2\sigma_Y^2}{\varkappa(M_Y+\sigma_Y)}\left(1 + \frac{4\sigma_Y^2}{\alpha_{Y,0}\beta_{Y,0}^2\overline{V}_Y\varkappa}\right) + \frac{C}{N^{\frac{3}{2}}},
\end{aligned}
$$

où

$$
\begin{aligned}
C = & \frac{2\sigma_X^2}{\varkappa(M_X+\sigma_X)}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\left[\frac{4\sigma_X^2}{\alpha_{X,0}\beta_{X,0}^2\overline{V}_X\varkappa}\left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1\right)^2 + \left(\varkappa\frac{M_X^2}{\sigma_X^2} + 1 + \ln\left(1 + \frac{4\sigma_X^2}{\alpha_{X,0}\overline{V}_X\varkappa\beta_{X,0}^2}\right)\right)^2\right] \\
& + \frac{2\sigma_Y^2}{\varkappa(M_Y+\sigma_Y)}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\left[\frac{4\sigma_Y^2}{\alpha_{Y,0}\beta_{Y,0}^2\overline{V}_Y\varkappa}\left(\varkappa\frac{M_Y^2}{\sigma_Y^2} + 1\right)^2 + \left(\varkappa\frac{M_Y^2}{\sigma_Y^2} + 1 + \ln\left(1 + \frac{4\sigma_Y^2}{\alpha_{Y,0}\overline{V}_Y\varkappa\beta_{Y,0}^2}\right)\right)^2\right].
\end{aligned}
$$

Preuve de la proposition 2.19

D'après (C.9), on a

$$\begin{aligned}
N\Delta_N \quad \leq \quad & \sigma_X\sqrt{\frac{N\overline{V}_X}{2\alpha_{X,0}}}+\sigma_Y\sqrt{\frac{N\overline{V}_Y}{2\alpha_{Y,0}}}-\sum_{i=1}^N\langle\eta_X(x_{i-1},y_{i-1}),x_{i-1}-\check{x}_{i-1}\rangle+\sum_{i=1}^N\langle\eta_Y(x_{i-1},y_{i-1}),y_{i-1}-\check{y}_{i-1}\rangle\\
&+\sum_{i=1}^N\left(\frac{1}{\sigma_X}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}N}}\parallel\eta_X(x_{i-1},y_{i-1})\parallel_{X,*}^2+\frac{1}{\sigma_Y}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}N}}\parallel\eta_Y(x_{i-1},y_{i-1})\parallel_{Y,*}^2\right)\\
&+2\overline{V}_X\sqrt{\beta_{X,0}^2+2\sum_{i=1}^N u_i+\frac{{u_N^*}^2}{\beta_{X,0}^2}+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right)}\\
&+2\overline{V}_Y\sqrt{\beta_{Y,0}^2+2\sum_{i=1}^N v_i+\frac{{v_N^*}^2}{\beta_{Y,0}^2}+\frac{v_N^*}{2}\ln\left(1+\frac{2}{\beta_{Y,0}^2}\sum_{i=1}^N v_i\right)},
\end{aligned}$$

où

$$\begin{aligned}
\beta_X=\sigma_X\sqrt{\tfrac{N}{2\alpha_{X,0}\overline{V}_X}},\quad \check{x}_i=W'_{\beta_X}\left(\textstyle\sum_{k=1}^i\eta_X(x_{k-1},y_{k-1})\right),\\
\beta_Y=\sigma_Y\sqrt{\tfrac{N}{2\alpha_{Y,0}\overline{V}_Y}},\quad \check{y}_i=W'_{\beta_Y}\left(\textstyle\sum_{k=1}^i\eta_Y(x_{k-1},y_{k-1})\right).
\end{aligned}$$

Soit

$$\begin{aligned}
\Delta_N\leq{}&\frac{\sigma_X}{\sqrt{N}}\sqrt{\frac{2\overline{V}_X}{\alpha_{X,0}}}+\frac{D_X}{\sqrt{N}}\sum_{i=1}^N\langle\eta_X(x_{i-1},y_{i-1}),\frac{\check{x}_{i-1}-x_{i-1}}{D_X}\rangle+\frac{1}{\sigma_X N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_X}{2\alpha_{X,0}}}\sum_{i=1}^N\left(\parallel\eta_X(x_{i-1},y_{i-1})\parallel_{X,*}^2-\sigma_X^2\right)\\
&+\frac{\sigma_Y}{\sqrt{N}}\sqrt{\frac{2\overline{V}_Y}{\alpha_{Y,0}}}-\frac{D_Y}{\sqrt{N}}\sum_{i=1}^N\langle\eta_Y(x_{i-1},y_{i-1}),\frac{\check{y}_{i-1}-y_{i-1}}{D_Y}\rangle+\frac{1}{\sigma_Y N^{\frac{3}{2}}}\sqrt{\frac{\overline{V}_Y}{2\alpha_{Y,0}}}\sum_{i=1}^N\left(\parallel\eta_Y(x_{i-1},y_{i-1})\parallel_{Y,*}^2-\sigma_Y^2\right)\\
&+\frac{2\overline{V}_X}{N}\sqrt{\beta_{X,0}^2+\frac{4(M_X^2+\sigma_X^2)N}{\alpha_{X,0}\overline{V}_X}+\frac{4}{\alpha_{X,0}\overline{V}_X}\sum_{i=1}^N(\parallel\eta_X(x_{i-1},y_{i_1})\parallel_{X,*}^2-\sigma_X^2)+\frac{{u_N^*}^2}{\beta_{X,0}^2}+\frac{u_N^*}{2}\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right)}\\
&+\frac{2\overline{V}_Y}{N}\sqrt{\beta_{Y,0}^2+\frac{4(M_Y^2+\sigma_Y^2)N}{\alpha_{Y,0}\overline{V}_Y}+\frac{4}{\alpha_{Y,0}\overline{V}_Y}\sum_{i=1}^N(\parallel\eta_Y(x_{i-1},y_{i_1})\parallel_{Y,*}^2-\sigma_Y^2)+\frac{{v_N^*}^2}{\beta_{Y,0}^2}+\frac{v_N^*}{2}\ln\left(1+\frac{2}{\beta_{Y,0}^2}\sum_{i=1}^N uvi\right)}.
\end{aligned}\tag{C.10}$$

De plus,

Lemme C.7 *On suppose que les hypothèses 2.9 et 2.12 sont vérifiées. Alors pour tout N et tout $\alpha>0$, on a*

$$\mathbb{P}\left({u_N^*}^2\geq\left\{\frac{2\sigma_X^2}{\varkappa\overline{V}_X\alpha_{X,0}}\ln\left(\exp(\varkappa M_X^2\sigma_X^{-2})\frac{N}{e\alpha}\right)\right\}^2\right)\leq\alpha,$$

$$\mathbb{P}\left(u_N^*\ln\left(1+\frac{2}{\beta_{X,0}^2}\sum_{i=1}^N u_i\right)\geq\frac{2\sigma_X^2}{\varkappa\overline{V}_X\alpha_{X,0}}\left\{\ln\left(\left(1+\frac{4\sigma_X^2}{\varkappa\alpha_{X,0}\overline{V}_X\beta_{X,0}^2}\right)\exp(\varkappa M_X^2\sigma_X^{-2})\frac{N}{e\alpha}\right)\right\}^2\right)\leq\alpha,$$

$$\mathbb{P}\left(v_N^{*2} \geq \left\{\frac{2\sigma_Y^2}{\varkappa \overline{V}_Y \alpha_{Y,0}} \ln\left(\exp(\varkappa M_Y^2 \sigma_Y^{-2}) \frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha,$$

$$\mathbb{P}\left(v_N^* \ln\left(1 + \frac{2}{\beta_{Y,0}^2} \sum_{i=1}^N v_i\right) \geq \frac{2\sigma_Y^2}{\varkappa \overline{V}_Y \alpha_{Y,0}} \left\{\ln\left(\left(1 + \frac{4\sigma_Y^2}{\varkappa \alpha_{Y,0} \overline{V}_Y \beta_{Y,0}^2}\right) \exp(\varkappa M_Y^2 \sigma_Y^{-2}) \frac{N}{e\alpha}\right)\right\}^2\right) \leq \alpha.$$

Preuve
Voir la preuve du lemme 2.2. □

La proposition 2.19 se déduit alors de l'équation (C.10), du lemme C.7 et des propositions B.5 et B.6.

C.3 Preuve des bornes pour l'algorithme d'extrapolation dual

C.3.1 Cas des problèmes d'inégalités variationnelles

Lemmes préliminaires

Lemme C.8 *Soient s, $\delta \in E^*$ et $z \in Q$. Alors*

$$W_\beta(z, s+\delta) \leq W_\beta(z, s) + \langle \delta, T_\beta(z, s) - z\rangle + W_\beta(T_\beta(z, s), \delta).$$

Preuve
Voir la preuve du lemme 3.1. □

Lemme C.9 *Soit $s \in E^*$ et $(x, y, s_+) = \mathcal{E}_{\beta,\gamma}(s)$. Alors*

$$\gamma\langle \Psi(y), y - \overline{x}\rangle + W_\beta(\overline{x}, s_+) \leq W_\beta(\overline{x}, s) + \frac{\gamma^2}{2\alpha_0\beta} \parallel \Psi(x) - \Psi(y) \parallel_*^2 - \frac{\alpha_0\beta}{2} \parallel x - y \parallel^2 .$$

Preuve
Voir la preuve du lemme 3.2. □

Preuve de la proposition 3.6

Par monotonie de g, on a

$$\begin{aligned}
\epsilon_D(y^N) &= \max\{\langle g(x), y^N - x\rangle \,|V(\overline{x}, x) \leq D\}, \\
&\leq \frac{1}{\sum_{i=0}^N \gamma_i} \max\{\sum_{i=0}^N \gamma_i \langle g(y_i), y_i - x\rangle \,|V(\overline{x}, y) \leq D\}, \\
&\leq \frac{1}{\sum_{i=0}^N \gamma_i} \max\left\{\sum_{i=0}^N \gamma_i \left(\langle \Psi(y_i), y_i - \overline{x}\rangle - \langle \eta(y_i), y_i - \overline{x}\rangle + \langle \Psi(y_i), \overline{x} - x\rangle - \langle \eta(y_i), \overline{x} - x\rangle\right)\right\}, \\
&\leq \frac{1}{\sum_{i=0}^N \gamma_i} \left\{\sum_{i=0}^N \gamma_i(\langle \Psi(y_i), y_i - \overline{x}\rangle - \langle \eta(y_i), y_i - \overline{x}\rangle) + \xi_D\left(-\sum_{i=0}^N \gamma_i \Psi(y_i)\right) + \xi_D\left(\sum_{i=0}^N \gamma_i \eta(y_i)\right)\right\}.
\end{aligned}$$

Le lemme C.9 nous permet alors d'obtenir

$$\begin{aligned}\sum_{i=0}^{N}\gamma_i\langle\Psi(y_i),y_i-\overline{x}\rangle+\xi_D\left(-\sum_{i=0}^{N}\gamma_i\Psi(y_i)\right) &\leq \sum_{i=0}^{N}\gamma_i\langle\Psi(y_i),y_i-\overline{x}\rangle+\beta_N D+W_{\beta_N}(\overline{x},s_N),\\ &\leq \beta_N D+\sum_{i=1}^{N}\left(\frac{\gamma_i^2}{2\alpha_0\beta_i}\parallel\Psi(x_i)-\Psi(y_i)\parallel_*^2-\frac{\alpha_0\beta_i}{2}\parallel x_i-y_i\parallel^2\right).\end{aligned}$$

De plus, grâce au lemme C.8, pour tout $\beta>0$,

$$\begin{aligned}\xi_D\left(\sum_{i=0}^{N}\gamma_i\eta(y_i)\right) &\leq \beta D+W_\beta\left(\overline{x},\sum_{i=0}^{N}\gamma_i\eta(y_i)\right)-W_\beta(\overline{x},0),\\ &\leq \beta D+\sum_{i=0}^{N}\left[W_\beta\left(\overline{x},\sum_{j=0}^{i}\gamma_j\eta(y_j)\right)-W_\beta\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta(y_j)\right)\right]\\ &\leq \beta D+\sum_{i=0}^{N}\left\langle\gamma_i\eta(y_i),T_\beta\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta(y_j)\right)-\overline{x}\right\rangle+\sum_{i=0}^{N}W_\beta\left(T_\beta\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta(y_j)\right),\gamma_i\eta(y_i)\right),\\ &\leq \beta D+\sum_{i=0}^{N}\langle\gamma_i\eta(y_i),\check{y}_i-\overline{x}\rangle+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_0\beta}\parallel\eta(y_i)\parallel_*^2,\end{aligned}$$

où

$$\check{y}_i=T_\beta\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta(y_j)\right).$$

Si bien que

$$\begin{aligned}\left(\sum_{i=0}^{N}\gamma_i\right)\epsilon_D(y^N)\leq(\beta+\beta_N)D+\sum_{i=0}^{N}\gamma_i\langle\eta(y_i),\check{y}_i-y_i\rangle+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_0\beta}\parallel\eta(y_i)\parallel_*^2\\ +\sum_{i=0}^{N}\left(\frac{\gamma_i^2}{2\alpha_0\beta_i}\parallel\Psi(x_i)-\Psi(y_i)\parallel_*^2-\frac{\alpha_0\beta_i}{2}\parallel x_i-y_i\parallel^2\right).\end{aligned}$$

Preuve de la proposition 3.7

Comme l'hypothèse 3.4 est vérifiée, on a pour tout i

$$\parallel\Psi(x_i)-\Psi(y_i)\parallel_*\leq M+\parallel\eta(x_i)\parallel_*+\parallel\eta(y_i)\parallel_*,$$

si bien qu'en passant à l'espérance dans la borne obtenue à la proposition 3.6, on obtient

$$\left(\sum_{i=0}^{N}\gamma_i\right)\mathbb{E}[\epsilon_D(y^N)]\leq(\beta+\beta_N)D+\frac{\sigma^2}{2\alpha_0\beta}\sum_{i=0}^{N}\gamma_i^2+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_0\beta_i}(M+2\sigma)^2.$$

Cette borne est minimale quand les suites $(\gamma_i)_i$ et $(\beta_i)_i$ sont constantes et lorsque

$$\gamma_i=1,\ \beta_i=(M+2\sigma)\sqrt{\frac{N+1}{2\alpha_0 D}}\text{ et }\beta=\sigma\sqrt{\frac{N+1}{2\alpha_0 D}}.$$

On a alors

$$\mathbb{E}[\epsilon_D(y^N)] \leq \sqrt{2}(M+3\sigma)\sqrt{\frac{D}{\alpha_0(N+1)}}.$$

Preuve de la proposition 3.8

Comme l'hypothèse 3.5 est vérifiée, on a pour tout i

$$\| \Psi(x_i) - \Psi(y_i) \|_* \leq L \| x_i - y_i \| + \| \eta(x_i) \|_* + \| \eta(y_i) \|_*,$$

et

$$\left(\sum_{i=0}^{N} \gamma_i\right) \epsilon_D(y^N) \leq (\beta+\beta_N)D + \sum_{i=0}^{N} \gamma_i \langle \eta(y_i), \check{y}_i - y_i \rangle + \sum_{i=0}^{N} \frac{\gamma_i^2}{2\alpha_0\beta} \| \eta(y_i) \|_*^2$$
$$+ \sum_{i=0}^{N} \frac{3\gamma_i^2}{2\alpha_0\beta_i} \left(\| \eta(x_i) \|_*^2 + \| \eta(y_i) \|_*^2\right) + \sum_{i=0}^{N} \left(\frac{3\gamma_i^2 L^2}{2\alpha_0\beta_i} - \frac{\alpha_0\beta_i}{2}\right) \| x_i - y_i \|^2 .$$

On obtient alors en passant à l'espérance lorsque $\frac{3\gamma_i^2 L^2}{\alpha_0\beta_i} - \alpha_0\beta_i \leq 0$,

$$\left(\sum_{i=0}^{N} \gamma_i\right) \mathbb{E}[\epsilon_D(y^N)] \leq (\beta+\beta_N)D + \frac{\sigma^2}{2\alpha_0\beta} \sum_{i=0}^{N} \gamma_i^2 + \sum_{i=0}^{N} \frac{3\gamma_i^2}{\alpha_0\beta_i}\sigma^2.$$

On choisit alors les suites $(\gamma_i)_i$ et $(\beta_i)_i$ constantes avec $\gamma_i = 1$, $\beta = \sigma\sqrt{\frac{N+1}{2\alpha_0 D}}$ et

- $\beta_i = \sigma\sqrt{\frac{3(N+1)}{\alpha_0 D}}$ si $N+1 \geq \frac{L^2 D}{\sigma^2\alpha_0}$,
- $\beta_i = \sqrt{3}L\alpha_0^{-1}$ sinon.

Les bornes annoncées dans la proposition s'en déduisent en remarquant que dans le second cas $\beta_i^{-1} \leq \sigma^{-1}\sqrt{\frac{\alpha_0 D}{3(N+1)}}$.

C.3.2 Cas des problèmes de point selle

Préliminaires

Remarque : Pour tous $\beta > 0$, $s \in E_X^*$

$$\xi_{D_X}(s) \leq \beta D_X + W_{X,\beta}(\overline{x}, s).$$

Lemme C.10 *soient s, $\delta \in E_X^*$ et $z \in X$. Alors*

$$W_{X,\beta}(z, s+\delta) \leq W_{X,\beta}(z,s) + \langle \delta, T_\beta^X(z,s) - z \rangle + W_{X,\beta}(T_\beta^X(z,s), \delta).$$

De plus $\forall z \in X$, $s \in E_X^*$, on a

$$W_{X,\beta}(z,s) \leq \frac{\| s \|_{X,*}^2}{2\alpha_{X,0}\beta}$$

Preuve
Analogue à celle du lemme 3.1. □

Lemme C.11 *Soit* $s \in E_X^*$ *et* $(x, u, s_+) = \mathcal{E}^X_{\beta,\gamma}(s, y, v)$. *Alors*

$$\gamma\langle\Phi(u,v), u-\overline{x}\rangle + W_{X,\beta}(\overline{x}, s_+) \leq W_{X,\beta}(\overline{x}, s) + \frac{\gamma^2}{2\alpha_{X,0}\beta} \parallel \Phi(x,y) - \Phi(u,v) \parallel^2_{X,*} - \frac{\alpha_{X,0}\beta}{2} \parallel x-u \parallel^2_X .$$

Preuve
On pose $z = \overline{x}$ et $\delta = -\gamma\Phi(u)$ dans le lemme C.10. On a

$$W_{X,\beta}(\overline{x}, s - \gamma\Phi(u,v)) \leq W_{X,\beta}(\overline{x}, s) + \langle -\gamma\Phi(u,v), T^X_\beta(\overline{x}, s) - \overline{x}\rangle + W_{X,\beta}(T^X_\beta(\overline{x}, s), -\gamma\Phi(u,v)).$$

C'est-à-dire

$$W_{X,\beta}(\overline{x}, s_+) \leq W_{X,\beta}(\overline{x}, s) + \langle -\gamma\Phi(u,v), x - \overline{x}\rangle + W_{X,\beta}(x, -\gamma\Phi(u,v)). \tag{C.11}$$

On choisit maintenant $z = x$, $s = -\gamma\Phi(x,y)$, $\delta = \gamma(\Phi(x,y) - \Phi(u,v))$ dans le lemme C.10. On a

$$\begin{aligned} W_{X,\beta}(x, -\gamma\Phi(u,v)) \leq W_{X,\beta}(x, -\gamma\Phi(x,y)) &+ \langle \gamma(\Phi(x,y) - \Phi(u,v)), T^X_\beta(x, -\gamma\Phi(x,y)) - x\rangle \\ &+ W_{X,\beta}(T_\beta(x, -\gamma\Phi(x,y)), \gamma(\Phi(x,y) - \Phi(u,v))). \end{aligned}$$

Soit

$$\begin{aligned} W_{X,\beta}(x, -\gamma\Phi(u,v)) &\leq W_{X,\beta}(x, -\gamma\Phi(x,y)) + \langle \gamma(\Phi(x,y) - \Phi(u,v)), u - x\rangle + W_{X,\beta}(u, \gamma(\Phi(x,y) - \Phi(u,v))), \\ &\leq W_{X,\beta}(x, -\gamma\Phi(x,y)) + \langle \gamma(\Phi(x,y) - \Phi(u,v)), u - x\rangle + \frac{\gamma^2}{2\alpha_{X,0}\beta} \parallel \Phi(x, y - \Phi(u,v) \parallel^2_{X,*} . \end{aligned}$$

On a choisit u tel que $W_{X,\beta}(x, -\gamma\Phi(x,y)) = \langle -\gamma\Phi(x,y), u - x\rangle - \beta V_X(x,u)$, et on a $V_X(x,u) \geq \frac{1}{2}\alpha_{X,0} \parallel x - u \parallel^2_X$. On déduit donc des deux inégalités précédentes que

$$\begin{aligned} \gamma\langle\Phi(u,v), u-\overline{x}\rangle + W_{X,\beta}(\overline{x}, s_+) - W_{X,\beta}(\overline{x}, s) &\leq \gamma\langle\Phi(u,v), u-\overline{x}\rangle - \gamma\langle\Phi(u,v), x-\overline{x}\rangle + W_{X,\beta}(x, -\gamma\Phi(u,v)), \\ &\leq \gamma\langle\Phi(u,v), u-x\rangle + W_{X,\beta}(x, -\gamma\Phi(x,y)) \\ &\quad + \gamma\langle\Phi(x,y) - \Phi(u,v), u-x\rangle \\ &\quad + \frac{\gamma^2}{2\alpha_{X,0}\beta} \parallel \Phi(x,y) - \Phi(u,v) \parallel^2_{X,*}, \\ &\leq \gamma\langle\Phi(x,y), u-x\rangle - \gamma\langle\Phi(x,y), u-x\rangle - \beta V_X(x,u) \\ &\quad + \frac{\gamma^2}{2\alpha_{X,0}\beta} \parallel \Phi(x,y) - \Phi(u,v) \parallel^2_{X,*}, \\ &\leq \frac{\gamma^2}{2\alpha_{X,0}\beta} \parallel \Phi(x,y) - \Phi(u,v) \parallel^2_{X,*} - \frac{\alpha_{X,0}\beta}{2} \parallel x-u \parallel^2_X . \end{aligned}$$

□

Lemme C.12 *Soit* $t \in E_Y^*$ *et* $(y, v, t_+) = \mathcal{E}^Y_{\beta,\gamma}(t, x, u)$. *Alors*

$$-\gamma\langle\Psi(u,v), v-\overline{y}\rangle + W_{Y,\beta}(\overline{y}, t_+) - W_{Y,\beta}(\overline{y}, t) \leq \frac{\gamma^2}{2\alpha_{Y,0}\beta} \parallel \Psi(x,y) - \Psi(u,v) \parallel^2_{Y,*} - \frac{\alpha_{Y,0}\beta}{2} \parallel y-v \parallel^2_Y .$$

Preuve
Analogue à celle du lemme précédent. □

Preuve de la proposition 3.11

On a

$$\begin{aligned}
\left(\sum_{i=0}^{N}\gamma_i\right)\varepsilon_N &= \left(\sum_{i=0}^{N}\gamma_i\right)\sup_{x\in\mathcal{F}_{D_X},y\in\mathcal{F}_{D_Y}}[L(u^N,y)-L(y,v^N)],\\
&\leq \sup_{y\in\mathcal{F}_{D_Y}}\sum_{i=0}^{N}\gamma_i[L(u_i,y)-L(u_i,v_i)]+\sup_{x\in\mathcal{F}_{D_X}}\sum_{i=0}^{N}\gamma_i[L(u_i,v_i)-L(x,v_i)],\\
&\leq \sup_{x\in\mathcal{F}_{D_X},y\in\mathcal{F}_{D_Y}}\sum_{i=0}^{N}\gamma_i[\langle L'_X(u_i,v_i),u_i-x\rangle+\langle L'_Y(u_i,v_i),y-v_i\rangle],
\end{aligned}$$

puisque L est convexe-concave. Majorons maintenant chacune de ces deux sommes.

$$\begin{aligned}
\sup_{x\in\mathcal{F}_{D_X}}\sum_{i=0}^{N}\gamma_i\langle L'_X(u_i,v_i),u_i-x\rangle &= \sup_{x\in\mathcal{F}_{D_X}}\sum_{i=0}^{N}\gamma_i(\langle\Phi(u_i,v_i),\overline{x}-x\rangle-\eta_X(u_i,v_i),\overline{x}-x\rangle)\\
&\quad+\sum_{i=0}^{N}\gamma_i(\langle\Phi(u_i,v_i),u_i-\overline{x}\rangle-\eta_X(u_i,v_i),u_i\overline{x}\rangle),\\
&\leq \xi_{D_X}\left(-\sum_{i=0}^{N}\gamma_i\Phi(u_i,v_i)\right)+\xi_{D_X}\left(\sum_{i=0}^{N}\gamma_i\eta_X(u_i,v_i)\right)\\
&\quad+\sum_{i=0}^{N}\gamma_i(\langle\Phi(u_i,v_i),u_i-\overline{x}\rangle-\eta_X(u_i,v_i),u_i\overline{x}\rangle).
\end{aligned}$$

Grâce au lemme C.12,

$$\begin{aligned}
\sum_{i=0}^{N}\gamma_i\langle\Phi(u_i,v_i),u_i-\overline{x}\rangle+\xi_{D_X}\left(-\sum_{i=0}^{N}\gamma_i\Phi(u_i,v_i)\right) &\leq \sum_{i=0}^{N}\gamma_i\langle\Phi(u_i,v_i),u_i-\overline{x}\rangle+\beta_{X,N}D_X+W_{X,\beta_{X,N}}(\overline{x},s_N),\\
&\leq \beta_{X,N}D_X+\sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}\parallel\Phi(x_i,y_i)-\Phi(u_i,v_i)\parallel_{X,*}^2\\
&\quad-\sum_{i=0}^{N}\frac{\alpha_{X,0}\beta_{X,i}}{2}\parallel x_i-u_i\parallel_X^2 .
\end{aligned}$$

D'autre part d'après le lemme C.10 pour tout $\beta_X > 0$

$$\begin{aligned}
\xi_{D_X}\left(\sum_{i=0}^{N}\gamma_i\eta_X(u_i,v_i)\right) &\leq \beta_X D_X + W_{X,\beta_X}\left(\overline{x},\sum_{i=0}^{N}\gamma_i\eta_X(u_i,v_i)\right) - W_{X,\beta_X}(\overline{x},0),\\
&\leq \beta_X D_X + \sum_{i=0}^{N} W_{X,\beta_X}\left(\overline{x},\sum_{j=0}^{i}\gamma_j\eta_X(u_j,v_j)\right) - \sum_{i=0}^{N} W_{X,\beta_X}\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta_X(u_j,v_j)\right),\\
&\leq \beta_X D_X + \sum_{i=0}^{N}\gamma_i\left\langle \eta_X(u_i,v_i), T^X_{\beta_X}\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta_X(u_j,v_j)\right) - \overline{x}\right\rangle\\
&\quad + \sum_{i=0}^{N} W_{X,\beta_X}\left(T^X_{\beta_X}\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta_X(u_j,v_j)\right),\gamma_i\eta_X(u_i,v_i)\right),\\
&\leq \beta_X D_X + \sum_{i=0}^{N}\gamma_i\langle\eta_X(u_i,v_i),\check{u}_i-\overline{x}\rangle + \sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{X,0}\beta_X}\parallel \eta_X(u_i,v_i)\parallel^2_{X,*},
\end{aligned}$$

où $\check{u}_i = T^X_{\beta_X}\left(\overline{x},\sum_{j=0}^{i-1}\gamma_j\eta_X(u_j,v_j)\right)$. Si bien que

$$\begin{aligned}
\sup_{x\in\mathcal{F}_{D_X}}\sum_{i=0}^{N}\gamma_i\langle L'_X(u_i,v_i),u_i-x\rangle &\leq (\beta_X+\beta_{X,N})D_X + \sum_{i=0}^{N}\gamma_i\langle\eta_X(u_i,v_i),\check{u}_i-u_i\rangle + \sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{X,0}\beta_X}\parallel\eta_X(u_i,v_i)\parallel^2_{X,*}\\
&\quad + \sum_{i=0}^{N}\left(\frac{\gamma_i^2}{2\alpha_{X,0}\beta_{X,i}}\parallel\Phi(x_i,y_i)-\Phi(u_i,v_i)\parallel^2_{X,*} - \frac{\alpha_{X,0}\beta_{X,i}}{2}\parallel x_i-u_i\parallel^2_X\right). \quad \text{(C.12)}
\end{aligned}$$

On montre de la même manière que pour tout $\beta_Y > 0$

$$\begin{aligned}
\sup_{y\in\mathcal{F}_{D_Y}}\sum_{i=0}^{N}\gamma_i\langle L'_Y(u_i,v_i),x-v_i\rangle &\leq (\beta_Y+\beta_{Y,N})D_Y - \sum_{i=0}^{N}\gamma_i\langle\eta_Y(u_i,v_i),\check{v}_i-v_i\rangle + \sum_{i=0}^{N}\frac{\gamma_i^2}{2\alpha_{Y,0}\beta_Y}\parallel\eta_Y(u_i,v_i)\parallel^2_{Y,*}\\
&\quad + \sum_{i=0}^{N}\left(\frac{\gamma_i^2}{2\alpha_{Y,0}\beta_{Y,i}}\parallel\Psi(x_i,y_i)-\Psi(u_i,v_i)\parallel^2_{Y,*} - \frac{\alpha_{Y,0}\beta_{Y,i}}{2}\parallel y_i-v_i\parallel^2_Y\right), \quad \text{(C.13)}
\end{aligned}$$

où $\check{v}_i = T^Y_{\beta_Y}\left(\overline{y},-\sum_{j=0}^{i-1}\gamma_j\eta_Y(u_j,v_j)\right)$. Le résultat annoncé se déduit alors de (C.12) et (C.13). □

Preuve de la proposition 3.10

D'après la proposition 3.6, on a pour tout $\beta > 0$

$$\begin{aligned}
(N+1)\epsilon_{\overline{V}}(y^N) &\leq (\beta+\beta_N)\overline{V} + \sum_{i=0}^{N}\langle\eta(y_i),\check{y}_i-y_i\rangle + \sum_{i=0}^{N}\frac{1}{2\alpha_0\beta}\parallel\eta(y_i)\parallel^2_*\\
&\quad + \sum_{i=0}^{N}\left(\frac{1}{2\alpha_0\beta_i}\parallel\Psi(x_i)_\Psi(y_i)\parallel^2_* - \frac{\alpha_0\beta_i}{2}\parallel x_i-y_i\parallel^2\right),
\end{aligned}$$

où

$$\check{y}_i = T_\beta\left(\overline{x}, \sum_{j=0}^{i-1}\eta(y_i)\right) - \overline{x} \text{ et } \beta = \sigma\sqrt{\frac{N+1}{2\alpha_0\overline{V}}}.$$

Comme g est lipschitzien sur Q, il vient

$$\begin{aligned}(N+1)\epsilon_{\overline{V}}(y^N) &\leq \sigma\sqrt{\frac{\overline{V}(N+1))}{2\alpha_0}} + \beta_N\overline{V} + \sum_{i=0}^{N}\langle\eta(y_i), \check{y}_i - y_i\rangle + \sum_{i=0}^{N}\frac{\sqrt{\overline{V}}}{\sqrt{2\alpha_0(N+1)}\sigma} \parallel \eta(y_i) \parallel_*^2 \\ &+ \sum_{i=0}^{N}\left(\frac{3}{2\alpha_0\beta_i}\left(\parallel \eta(x_i) \parallel_*^2 + \parallel \eta(y_i) \parallel_*^2 + L^2 \parallel x_i - y_i \parallel^2\right) - \frac{\alpha_0\beta_i}{2} \parallel x_i - y_i \parallel^2\right).\end{aligned}$$

En choisissant $\beta_i = \max\left(\sigma\sqrt{\frac{3(N+1)}{\alpha_0\overline{V}}}, \frac{\sqrt{3}L}{\alpha_0}\right)$, on obtient

$$\begin{aligned}\epsilon_{\overline{V}}(y^N) &\leq (\sqrt{2}+\sqrt{3})\frac{\sigma}{\sqrt{N+1}}\sqrt{\frac{\overline{V}}{\alpha_0}} + \frac{\overline{V}}{N+1}\max\left\{\sigma\sqrt{\frac{3(N+1)}{\alpha_0\overline{V}}}, \frac{\sqrt{3}L}{\alpha_0}\right\} \\ &+ \frac{2}{N+1}\sqrt{\frac{2\overline{V}}{\alpha_0}}\sum_{i=0}^{N}\langle\eta(y_i), \sqrt{\frac{\alpha_0}{2\overline{V}}}\frac{(\check{y}_i - y_i)}{2}\rangle \\ &+ \sqrt{\frac{\overline{V}}{\alpha_0}}\frac{(\sqrt{2}+\sqrt{3})}{2\sigma(N+1)^{\frac{3}{2}}}\sum_{i=0}^{N}(\parallel \eta(y_i) \parallel_*^2 - \sigma^2) + \sqrt{\frac{\overline{V}}{\alpha_0}}\frac{\sqrt{3}}{2\sigma(N+1)^{\frac{3}{2}}}\sum_{i=0}^{N}(\parallel \eta(x_i) \parallel_*^2 - \sigma^2).\end{aligned}$$

Il ne reste plus alors qu'à appliquer les propositions B.5 et B.6 pour achever la preuve.

C.4 Preuve des résultats du chapitre 4

Preuve de la proposition 4.3

Cette preuve est empruntée à [28].

Lemme C.13

$$f^+(t_k) \leq \frac{t^* - t_0}{1-\kappa}\left[\frac{1}{2(1-\kappa)}\right]^k.$$

Preuve
On note $\beta = \frac{1}{2(1-\kappa)}(< 1)$ et $\delta_k = \frac{f^+(t_k)}{\sqrt{t_{k+1}-t_k}}$. Si $k \geq 1$, alors les trois points $(t_{k-1}, f^+(t_{k-1}))$, $(t_k, f^-(t_k))$ et $(t_{k+1}, 0)$ sont situés sur une même droite (par construction), et par conséquent

$$\begin{aligned}f^+(t_{k-1}) &= f^-(t_k) + \frac{t_k - t_{k-1}}{t_{k+1} - t_k}(f^-(t_k) - 0) \\ &\geq \left(1 + \frac{t_k - t_{k-1}}{t_{k+1} - t_k}\right)(1-\kappa)f^+(t_k) \\ &\geq 2\sqrt{\frac{t_k - t_{k-1}}{t_{k+1} - t_k}}(1-\kappa)f^+(t_k).\end{aligned}$$

Donc $\delta_k \leq \beta\delta_{k-1}$ et il s'ensuit que

$$\begin{aligned} f^+(t_k) &= \delta_k\sqrt{t_{k+1}-t_k} \\ &\leq \beta^k\delta_0\sqrt{t_{k+1}-t_k} \\ &\leq \beta^k f^+(t_0)\sqrt{\frac{t_{k+1}-t_k}{t_1-t_0}} \\ &\leq \beta^k f^+(t_0)\sqrt{\frac{t_{k+1}-t_k}{f^-(t_0)}}. \end{aligned}$$

Alors, puisque $f^-(t_k) \geq (1-\kappa)f^+(t_k)$,

$$\begin{aligned} f^+(t_k) &\leq \frac{\beta^k}{1-\kappa}\sqrt{f^-(t_0)(t_{k+1}-t_k)} \\ &\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t^*-t_0)} \\ &\leq \frac{\beta^k}{1-\kappa}(t^*-t_0). \end{aligned}$$

□

Annexe D

Index des hypothèses

Hypothèses utilisées dans le chapitre 2

Hypothèse 2.1
f' est borné sur Q :

$$\sup_{x\in Q} \| f'(x) \|_* = M < +\infty.$$

Hypothèse 2.2
Pour tout i,

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0 \text{ p.s., et } \mathbb{E}[\| \eta(x_i) \|_*^2 |\mathcal{F}_i] \leq \sigma^2 < +\infty \text{ p.s.}$$

Hypothèse 2.3
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*}{\sigma_i}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.4
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*^2}{\sigma_i^2}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.5
L'opérateur g est borné sur Q, il existe une constante $M > 0$ telle que

$$\forall x \in Q \quad \| g(x) \|_* \leq M.$$

Hypothèse 2.6
Pour tout i,

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_{i-1}] = 0 \text{ et } \mathbb{E}[\| \eta(x_i) \|_*^2 |\mathcal{F}_{i-1}] \leq \sigma^2 \text{ p.s.}$$

Hypothèse 2.7
Il existe $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*}{\sigma_i}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.8
Il existe $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*^2}{\sigma_i^2}\right)\middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.9
Les sous-/super-différentiels de L sont bornés : pour tout $x \in X$ et tout $y \in Y$, on a

$$\begin{aligned} \| L'_X(x,y) \|_{X,*} &\leq M_X, \\ \| L'_X(x,y) \|_{X,*} &\leq M_Y. \end{aligned}$$

Hypothèse 2.10
Pour tout i,

$$\mathbb{E}[\eta_X(x_{i-1}, y_{i-1})|\mathcal{F}_{i-1}] = 0, \ \mathbb{E}[\eta_Y(x_{i-1}, y_{i-1})|\mathcal{F}_{i-1}] = 0,$$

$$\mathbb{E}[\| \eta_X(x_{i-1}, y_{i-1}) \|_*^2 |\mathcal{F}_{i-1}] \leq \sigma_X^2 \text{ et } \mathbb{E}[\| \eta_Y(x_{i-1}, y_{i-1}) \|_*^2 |\mathcal{F}_{i-1}] \leq \sigma_Y^2 \text{ p.s.}$$

Hypothèse 2.11
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_X(x_i, y_i) \|_{X,*}}{\sigma_X}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s. et } \mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_Y(x_i, y_i) \|_{Y,*}}{\sigma_Y}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.12
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_X(x_i, y_i) \|_{X,*}^2}{\sigma_X^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s. et } \mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta_Y(x_i, y_i) \|_{Y,*}^2}{\sigma_Y^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 2.13
f' est lipschitzien sur Q : il existe une constante positive $L < \infty$ telle que pour tous $x, x' \in Q$,

$$\| f'(x) - f'(x') \|_* \leq L \| x - x' \| .$$

Hypothèses utilisées dans le chapitre 3

Hypothèse 3.1
f' est à variations bornées sur Q. Il existe une constante $M > 0$ telle que

$$\forall x, y \in Q \ \ \| f'(x) - f'(y) \|_* \leq M.$$

Hypothèse 3.2
Il existe $0 < \sigma^2 < +\infty$ telle que pour tout i,

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0 \text{ p.s., } \mathbb{E}[\eta(y_i)|\mathcal{F}_i] = 0 \text{ p.s.,}$$

$$\mathbb{E}[\| \eta(x_i) \|_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s. et } \mathbb{E}[\| \eta(y_i) \|_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s.}$$

Hypothèse 3.3
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$,

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*^2}{\sigma_i^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s. et } \mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(y_i) \|_*^2}{\sigma_i^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 3.4
L'opérateur g est à variations bornées sur Q. Il existe une constante $M > 0$ telle que

$$\forall x, y \in Q \ \ \| g(x) - g(y) \|_* \leq M.$$

Hypothèse 3.5 L'opérateur g est lipschitzien sur Q, il existe une constante $L > 0$ telle que

$$\| g(x) - g(y) \|_* \leq L \| x - y \| \ \ \forall x, y \in Q.$$

Hypothèse 3.6
Il existe $0 < \sigma^2 < +\infty$ telle que pour tout i,

$$\mathbb{E}[\eta(x_i)|\mathcal{F}_i] = 0 \text{ p.s., } \mathbb{E}[\eta(y_i)|\mathcal{F}_i] = 0 \text{ p.s.,}$$

$$\mathbb{E}[\| \eta(x_i) \|_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s. et } \mathbb{E}[\| \eta(y_i) \|_*^2 |\mathcal{F}_i] \leq \sigma^2 \text{ p.s.}$$

Hypothèse 3.7
Il existe une constante $0 < \varkappa < \infty$ telle que pour tout $i \geq 0$,

$$\mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(x_i) \|_*^2}{\sigma_i^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s. et } \mathbb{E}\left[\exp\left(\varkappa \frac{\| \eta(y_i) \|_*^2}{\sigma_i^2}\right) \middle| \mathcal{F}_i\right] \leq e \text{ p.s.}$$

Hypothèse 3.8
L'_X et L'_Y sont à variations bornées sur $Q = X \times Y$, c'est-à-dire que

$$\forall (x,y),(w,z) \in Q, \left\{ \begin{array}{rcl} \| L'_X(x,y) - L'_X(w,z) \|_{X,*} & \leq & M_X, \\ \| L'_Y(x,y) - L'_Y(w,z) \|_{Y,*} & \leq & M_Y, \end{array} \right.$$

où M_X et M_Y sont deux constantes strictement positives.

Hypothèse 3.9
L'_X et L'_Y sont lipschitziens sur $Q = X \times Y$, c'est-à-dire que

$$\forall (x,y),(w,z) \in Q, \left\{ \begin{array}{rcl} \| L'_X(x,y) - L'_X(w,z) \|_{X,*} & \leq & M_{X,X} \| x - w \|_X + M_{X,Y} \| y - z \|_Y, \\ \| L'_Y(x,y) - L'_Y(w,z) \|_{Y,*} & \leq & M_{Y,X} \| x - w \|_X + M_{Y,Y} \| y - z \|_Y \end{array} \right.$$

où $M_{X,X}$, $M_{X,Y}$, $M_{Y,X}$ et $M_{Y,Y}$ sont quatre constantes strictement positives.

Hypothèse 3.10
Il existe deux constantes strictement positives σ_X et σ_Y telles que pour tout $i \geq 0$ les relations suivantes sont vraies presque sûrement

$$\begin{array}{ll} \mathbb{E}[\eta_X(u_i,v_i)|\mathcal{F}_i] = 0, & \mathbb{E}[\eta_X(x_i,y_i)|\mathcal{F}_i] = 0, \\ \mathbb{E}[\eta_Y(u_i,v_i)|\mathcal{F}_i] = 0, & \mathbb{E}[\eta_Y(x_i,y_i)|\mathcal{F}_i] = 0, \\ \mathbb{E}[\| \eta_X(u_i,v_i) \|_{X,*}^2 |\mathcal{F}_i] \leq \sigma_X^2, & \mathbb{E}[\| \eta_X(x_i,y_i) \|_{X,*} |\mathcal{F}_i] \leq \sigma_X^2, \\ \mathbb{E}[\| \eta_Y(u_i,v_i) \|_{Y,*}^2 |\mathcal{F}_i] \leq \sigma_Y^2, & \mathbb{E}[\| \eta_Y(x_i,y_i) \|_{Y,*} |\mathcal{F}_i] \leq \sigma_Y^2. \end{array}$$

Bibliographie

[1] K. Barty, J.S. Roy, and C. Strugarek. Hilbert-Valued Perturbed Subgradient Algorithms. *Mathematics of Operations Research*, 32(3) :551, 2007.

[2] A. Ben-Tal, T. Margalit, and A. Nemirovski. The Ordered Subsets Mirror Descent Optimization Method with Applications to Tomography. *SIAM JOURNAL OF OPTIMIZATION*, 12(1) :79–108, 2001.

[3] A. Ben-Tal and A. Nemirovski. The conjugate barrier mirror descent method for non-smooth convex optimization. *MINERVA Optimization Center Report, Technion Institute of Technology. Available at http ://iew3. technion. ac. il/Labs/Opt/opt/Pap/CP MD. pdf*, 1999.

[4] A. Ben-Tal and A.S. Nemirovskii. *Lectures on modern convex optimization*. Society for Industrial and Applied Mathematics Philadelphia, PA, 2001.

[5] A. Benveniste, P. Priouret, and M. Métivier. *Adaptive algorithms and stochastic approximations*. 1990.

[6] O. Berman, Z. Drezner, J. Wang, and G. Wesolowsky. The minimax and maximin location problems on a network with uniform distributed weights. *IIE Transactions*, 35, November 2003.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[8] KL Chung. On a stochastic approximation method. *Ann. Math. Statist*, 25(3) :463–483, 1954.

[9] M. Duflo. *Random Iterative Models*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1997.

[10] J.-B. Hiriart-Urruty. *Optimisation et analyse convexe*. Presses Universitaires de France, 1998.

[11] A. Juditsky, G. Lan, A. Nemirovski, and A. Shapiro. Stochastic approximation approach to stochastic programming. *URL http ://http ://www2. isye. gatech. edu/nemirovs*, 2007.

[12] A. Juditsky and A.S. Nemirovski. Large Deviations of Vector-valued Martingales in 2-Smooth Normed Spaces. *Arxiv preprint arXiv :0809.0813*, 2008.

[13] A. Juditsky, A.S. Nemirovskii, and C. Tauvel. Solving variational inequalities with Stochastic Mirror-Prox algorithm. *Arxiv preprint arXiv :0809.0815*, 2008.

[14] A.B. Juditsky, A.V. Nazin, A.B. Tsybakov, and N. Vayatis. Recursive Aggregation of Estimators by the Mirror Descent Algorithm with Averaging. *Problems of Information Transmission*, 41(4) :368–384, 2005.

[15] Anatoli Juditsky and Arkadii Nemirovski. Nonparametric estimation via convex programming. Rapport de recherche, Optimization Online, 2007.

[16] G. Korpelevich. Extrapolation gradient methods and relation to Modified Lagrangeans. *Ekonomika i Matematicheskie Metody*, 19 :694–703, 1983.

[17] R. Liptser and V. Spokoiny. Deviation probability bound for martingales with applications to statistical estimation. *Statistics and Probability Letters*, 46(4) :347–357, 2000.

[18] Z. Lu, A. Nemirovski, and R.D.C. Monteiro. Large-scale semidefinite programming via a saddle point Mirror-Prox algorithm. *Mathematical Programming*, 109(2) :211–237, 2007.

[19] H.M. Markowitz. Portfolio selection. *Journal of Finance*, 1952.

[20] A. Nemirovski. Prox-method with rate of convergence o (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle-point problems. *SIAM Journal on Optimization*, 15(1) :229–251, 2004.

[21] A. Nemirovski and R. Rubinstein. An efficient stochastic approximation algorithm for stochastic saddle point problems,.

[22] A. Nemirovski and B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley and Sons, 1983.

[23] A. Nemirovski and D. Yudin. Cesaro convergence of the gradient method for approximation of saddle points of convex-concave functions. In *Soviet Math. Dokl*, volume 19, pages 482–486, 1978.

[24] Y. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems, CORE DP 2003/68. *Mathematical Programming*.

[25] Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, pages 1–39.

[26] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1) :127–152, 2005.

[27] Yu. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Core Discussion Paper*, (2003/69).

[28] Yu. Nesterov. *Introductory Lectures on Convex Optimization : a Basic Course*. Kiuwer Academic Publisher, 2004.

[29] B. Polyak. New stochastic approximation type procedures. *Automat. i Telemekh*, 51 :98–107, 1990.

[30] B. Polyak and A. Juditsky. Acceleration of Stochastic Approximation by Averaging. *SIAM Journal on Control and Optimization*, 30 :838, 1992.

[31] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist*, 22(3) :400–407, 1951.

[32] J. Sacks. Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist*, 29(2) :373–405, 1958.