Approches variationnelles en optimisation combinatoire

et autres sujets en optimisation

Jérôme MALICK

Chercheur CNRS, Laboratoire Jean Kuntzmann, Grenoble

26 janvier 2017

Soutenance pour l'habilitation à diriger les recherches

Plan de la présentation

1. Introduction
   - Un travail en optimisation mathématique...
   - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
   - Optimisation discrète : analyse des fonctions de coupes
   - Optimisation semidéfinie : problèmes binaires quadratiques
   - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
   - Optimisation avec données incertaines
   - Optimisation avec données distribuées
   - Optimisation avec données structurées

Plan de la présentation

1. Introduction
    - Un travail en optimisation mathématique...
    - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
    - Optimisation discrète : analyse des fonctions de coupes
    - Optimisation semidéfinie : problèmes binaires quadratiques
    - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
    - Optimisation avec données incertaines
    - Optimisation avec données distribuées
    - Optimisation avec données structurées

Plan de la présentation

1. Introduction
   - Un travail en optimisation mathématique...
   - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
   - Optimisation discrète : analyse des fonctions de coupes
   - Optimisation semidéfinie : problèmes binaires quadratiques
   - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
   - Optimisation avec données incertaines
   - Optimisation avec données distribuées
   - Optimisation avec données structurées

Optimisation mathématique : des maths utiles

L'optimisation est le domaine des maths appliquées qui s'intéresse à

$$\left\{\begin{array}{lll} \text{minimiser} \quad f(x) & & \text{(objectif)} \\ \qquad g(x) \leqslant 0 & & \text{(contraintes)} \\ \qquad x \in X & & \text{(domaine)} \end{array}\right. \quad \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{(ensemble réalisable)}$$

Des applications partout !

Optimisation mathématique : des maths utiles

L'optimisation est le domaine des maths appliquées qui s'intéresse à

$$\left\{\begin{array}{lll} \text{minimiser} \quad f(x) & \text{(objectif)} & \\ \quad g(x) \leqslant 0 & \text{(contraintes)} & \left.\right\} \text{(ensemble réalisable)} \\ \quad x \in X & \text{(domaine)} & \end{array}\right.$$

Des applications partout !

Exemple : problème industriel en production électrique

			
plannings de production	$x_1 \in X_1$	$x_2 \in X_2$	$x_3 \in X_3$
coût de production	${c_1}^\top x_1$	${c_2}^\top x_2$	${c_3}^\top x_3$

Optimisation mathématique : des maths utiles

L'optimisation est le domaine des maths appliquées qui s'intéresse à

$$\left\{\begin{array}{lll} \text{minimiser} & f(x) & \text{(objectif)} \\ & g(x) \leqslant 0 & \text{(contraintes)} \\ & x \in X & \text{(domaine)} \end{array}\right. \quad \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{(ensemble réalisable)}$$

Des applications partout !

Exemple : problème industriel en production électrique



plannings de production	$x_1 \in X_1$	$x_2 \in X_2$	$x_3 \in X_3$
coût de production	$c_1^\top x_1$	$c_2^\top x_2$	$c_3^\top x_3$

$$\binom{\text{modèle}}{\text{simplifié}} \left\{\begin{array}{lll} \min & \sum_{i=1}^{N} c_i^\top x_i & \text{(coûts de production)} \\ & \sum_{i=1}^{N} x_i = d \quad \in \mathbb{R}^T & \text{(contraintes de demande)} \\ & (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N & \text{(contraintes techniques)} \end{array}\right.$$

Optimisation : tout dépend des propriétés

Problème d'optimisation, sous forme abstraite

$$\begin{cases} \min \quad f(x) & \text{(objectif)} \\ \quad g(x) \leqslant 0 & \text{(contraintes)} \\ \quad x \in X & \text{(domaine)} \end{cases}$$

Analyse et résolution $\leftarrow$ propriétés mathématiques de f, g et X

Différentes classes de problèmes, différents sous-domaines de l'optimisation

Ex : en optimisation combinatoire : f, g simples et X compliqué (parties discrètes)

$$\begin{cases} \min \quad c^\top x \ (+x^\top Q\, x) \\ \quad Ax \leqslant b \\ \quad x \in \mathbb{Z}^n \end{cases}$$

Ex : en apprentissage : f terme d'attache aux données, g impose une structure

$$\begin{cases} \min \quad \sum_{i=1}^{N} (a_i^\top x - y_i)^2 \\ \quad \|x\|_{\ell_1} \leqslant \mu \\ \quad x \in \mathbb{R}^n \end{cases}$$

Optimisation : tout dépend des propriétés

Problème d'optimisation, sous forme abstraite

$$\begin{cases} \min & f(x) \quad \text{(objectif)} \\ & g(x) \leqslant 0 \quad \text{(contraintes)} \\ & x \in X \quad \text{(domaine)} \end{cases}$$

Analyse et résolution $\leftarrow$ propriétés mathématiques de f, g et X

Différentes classes de problèmes, différents sous-domaines de l'optimisation

Ex : en optimisation combinatoire : f, g simples et X compliqué (parties discrètes)

$$\begin{cases} \min & c^\top x \ (+x^\top Q\, x) \\ & Ax \leqslant b \\ & x \in \mathbb{Z}^n \end{cases} \qquad \begin{cases} \max & \theta(u) \\ & u \in \mathbb{R}^m \end{cases}$$

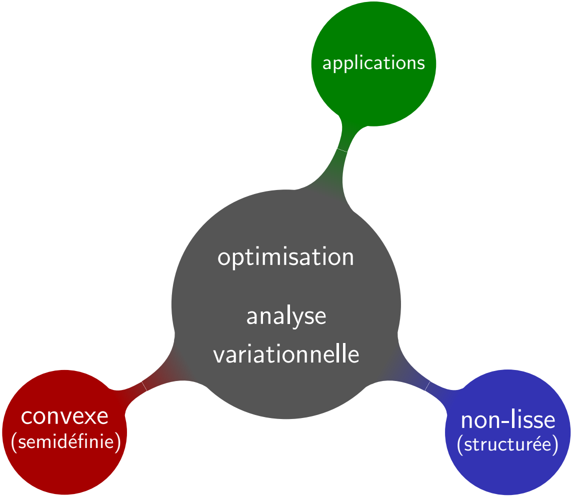
Ex : en apprentissage : f terme d'attache aux données, g impose une structure

$$\begin{cases} \min & \sum_{i=1}^{N} (a_i^\top x - y_i)^2 \\ & \|x\|_{\ell_1} \leqslant \mu \\ & x \in \mathbb{R}^n \end{cases}$$

Remarque : rôle-clé de l'optimisation convexe non-lisse dans les deux cas

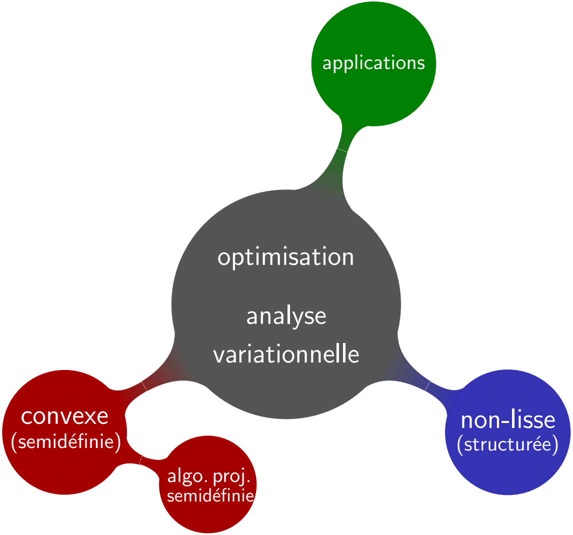
Mon travail en optimisation

Contributions (28 articles journaux)



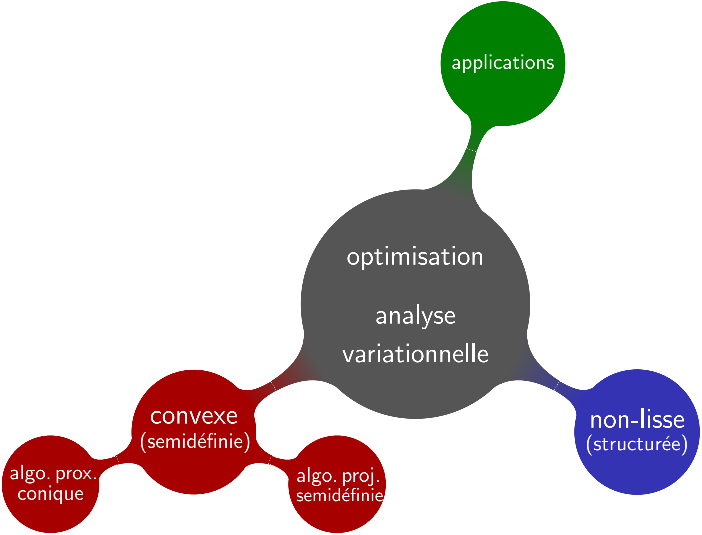
Mon travail en optimisation

Contributions (28 articles journaux)



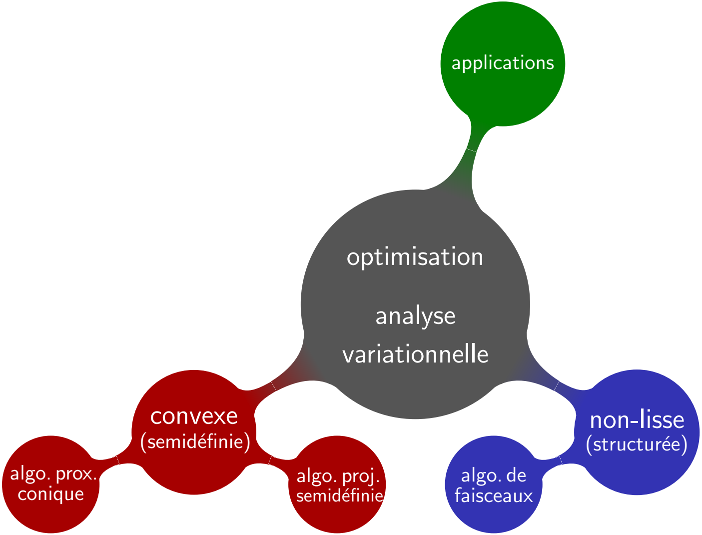
Mon travail en optimisation

Contributions (28 articles journaux)



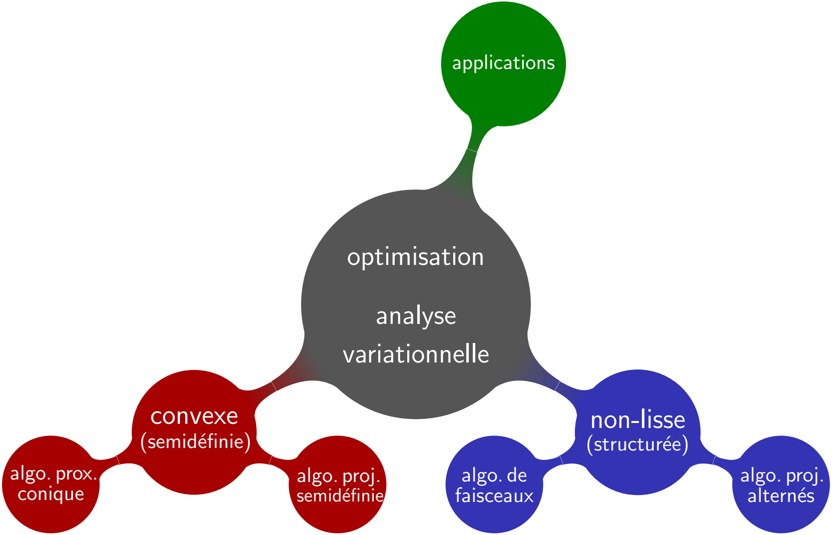
Mon travail en optimisation

Contributions (28 articles journaux)



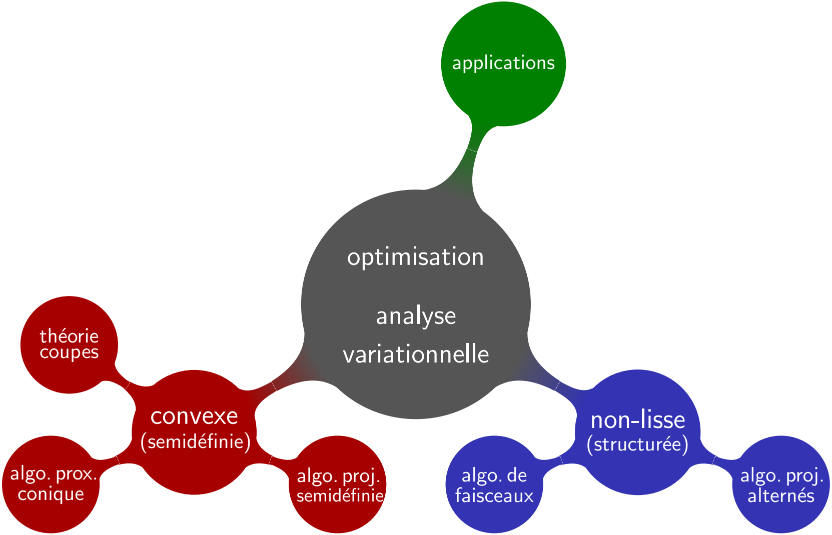
Mon travail en optimisation

Contributions (28 articles journaux)



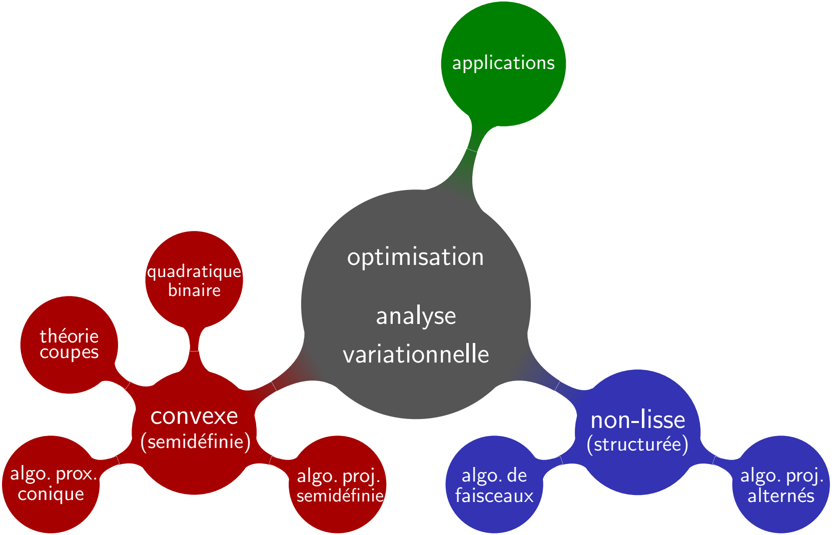
Mon travail en optimisation

Contributions (28 articles journaux)



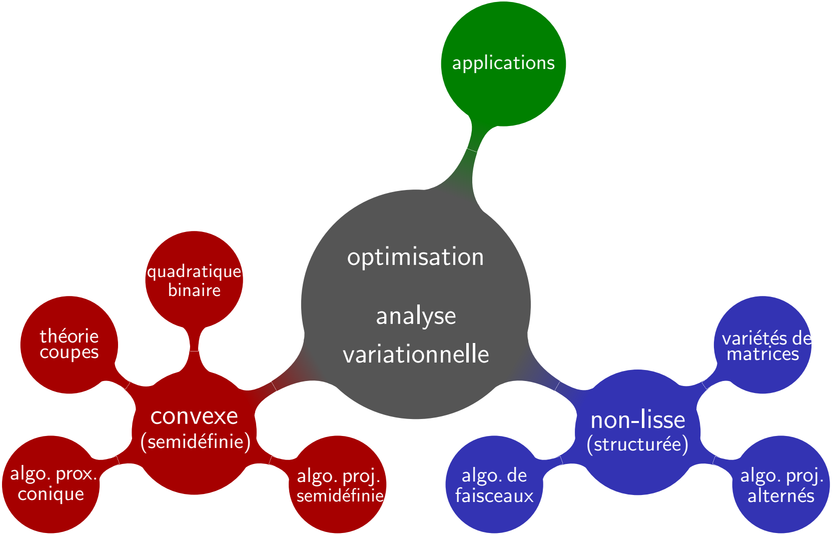
Mon travail en optimisation

Contributions (28 articles journaux)



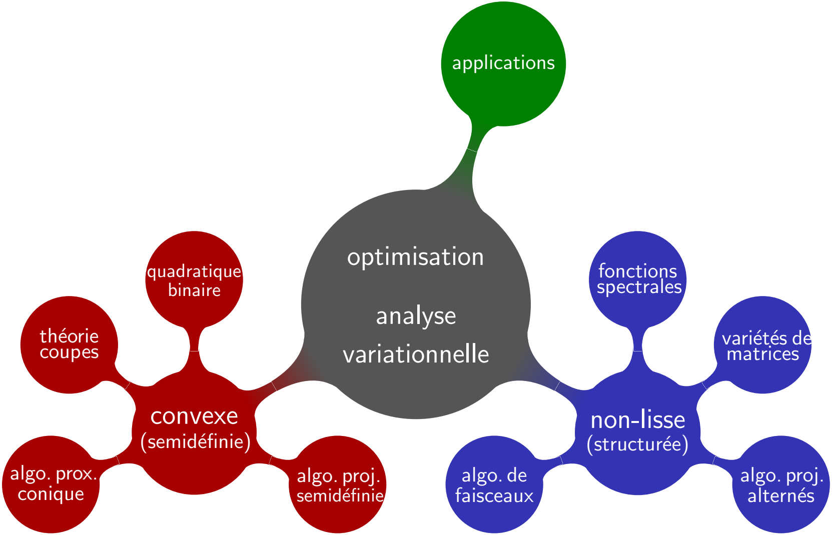
Mon travail en optimisation

Contributions (28 articles journaux)



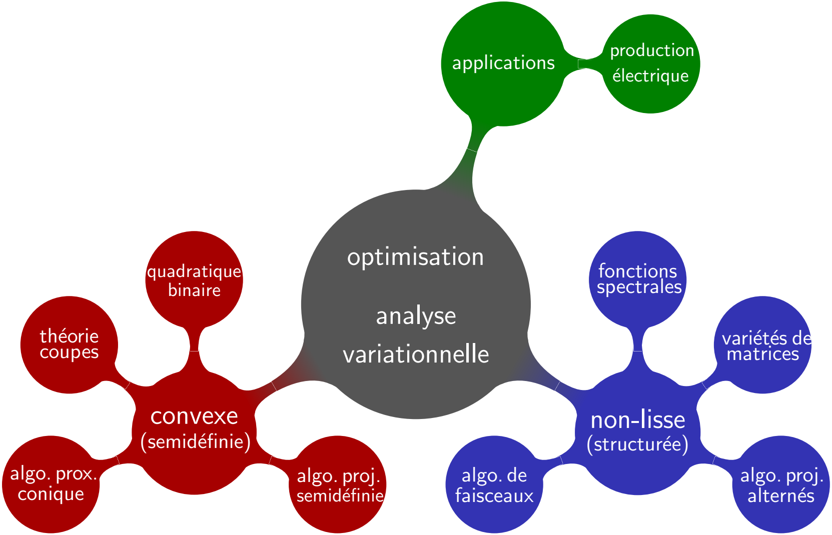
Mon travail en optimisation

Contributions (28 articles journaux)



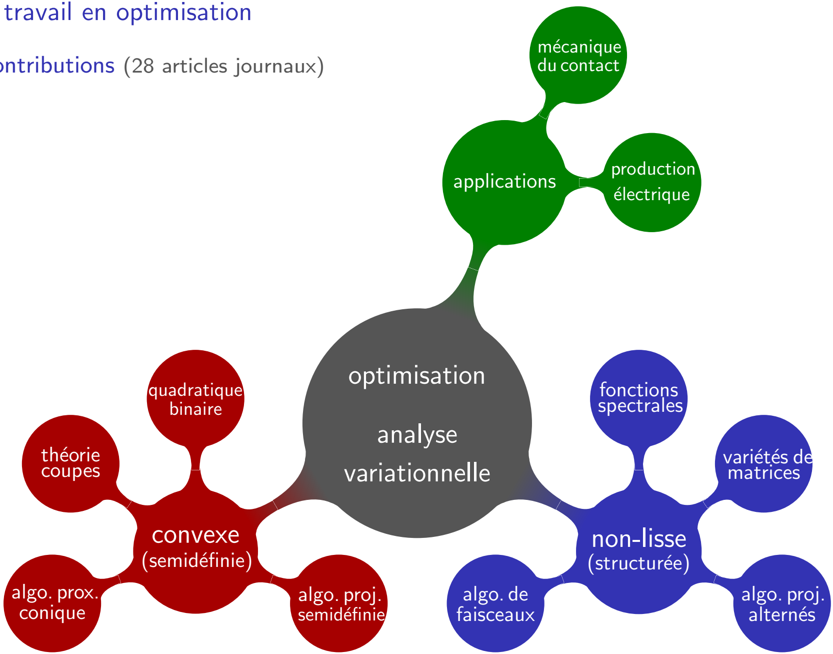
Mon travail en optimisation

Contributions (28 articles journaux)



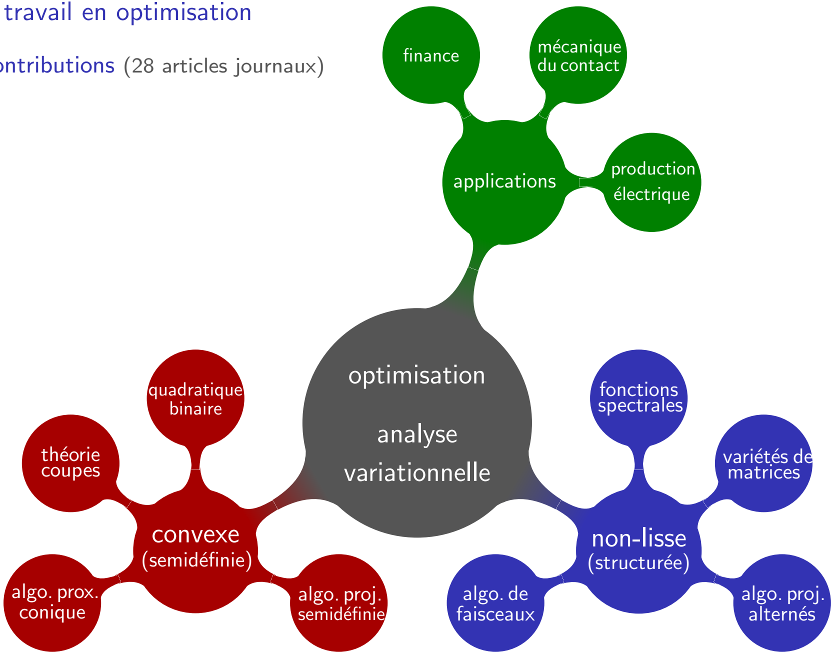
Mon travail en optimisation

Contributions (28 articles journaux)



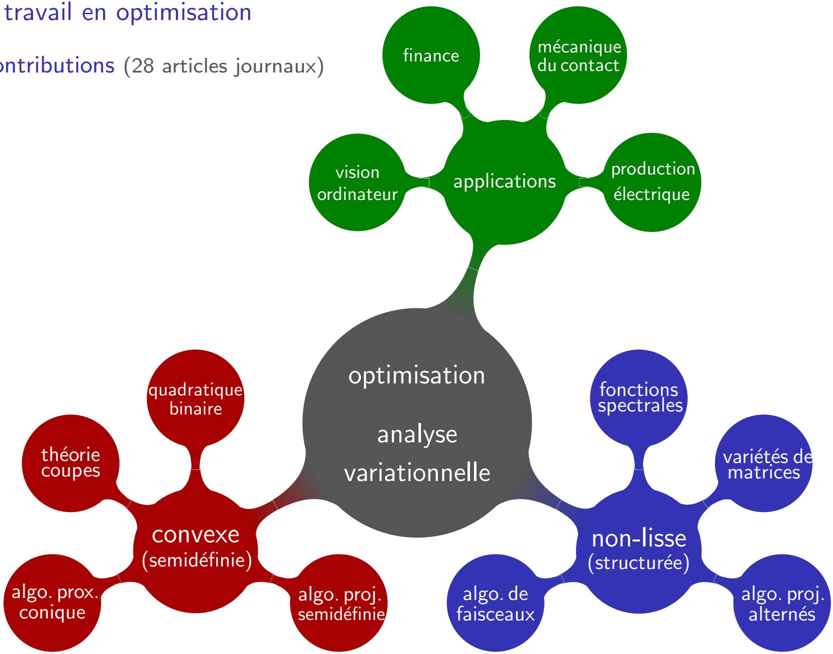
Mon travail en optimisation

Contributions (28 articles journaux)



Mon travail en optimisation

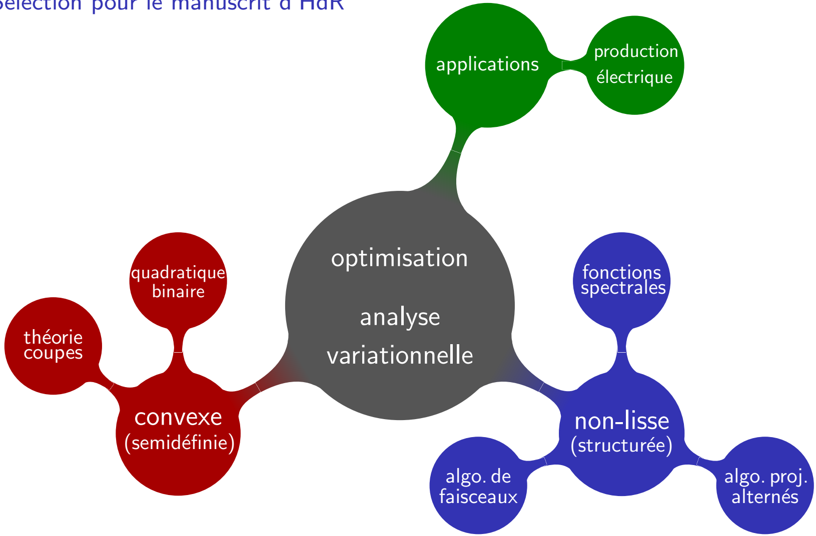
Contributions (28 articles journaux)



Mon travail en optimisation

Contributions (28 articles journaux)

Sélection pour le manuscrit d'HdR



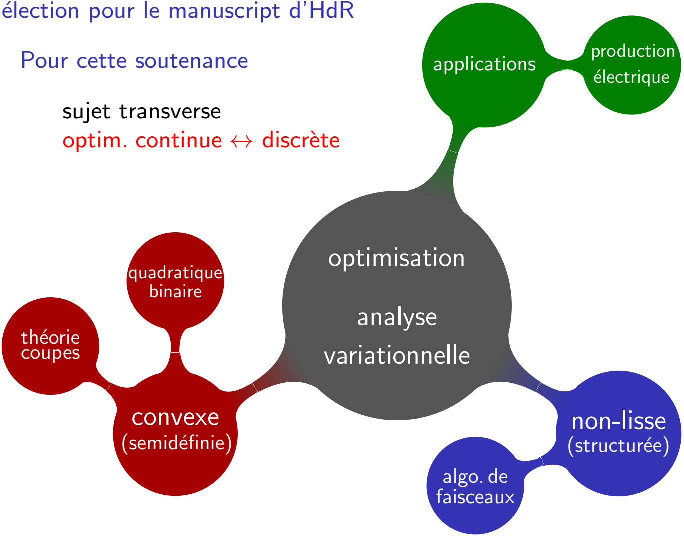
Mon travail en optimisation

Contributions (28 articles journaux)

Sélection pour le manuscrit d'HdR

Pour cette soutenance

sujet transverse
optim. continue $\leftrightarrow$ discrète



Plan de la présentation

1. Introduction
   - Un travail en optimisation mathématique...
   - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
   - Optimisation discrète : analyse des fonctions de coupes
   - Optimisation semidéfinie : problèmes binaires quadratiques
   - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
   - Optimisation avec données incertaines
   - Optimisation avec données distribuées
   - Optimisation avec données structurées

Deux techniques de l'optimisation discrète

Optimisation discrète

(maximiser f simple sur un ensemble compliqué $D = \{g(x) \leqslant 0, x \in X\})$

Deux techniques fondamentales

1. borner ou relaxer/relâcher $=$ agrandir D en un ensemble P plus simple
2. couper $=$ séparer D et un point de P pour préciser la relaxation

Deux techniques de l'optimisation discrète

Optimisation discrète

(maximiser f simple sur un ensemble compliqué $D = \{g(x) \leqslant 0, x \in X\}$)

Deux techniques fondamentales

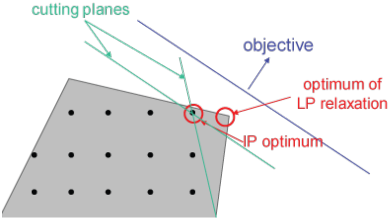
1. borner ou relaxer/relâcher $=$ agrandir D en un ensemble P plus simple
2. couper $=$ séparer D et un point de P pour préciser la relaxation

Illustration : optimisation entière (IP)

$D = \{Ax \leqslant b, x \in \mathbb{Z}^n\}$

$P = \{Ax \leqslant b, x \in \mathbb{R}^n\}$ relax. linéaire

$D = \{A'x \leqslant b', x \in \mathbb{Z}^n\}$



Deux techniques de l'optimisation discrète

Optimisation discrète

(maximiser f simple sur un ensemble compliqué $D = \{g(x) \leqslant 0, x \in X\}$)

Deux techniques fondamentales

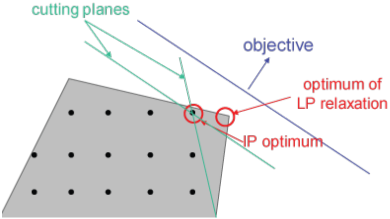
1. borner ou relaxer/relâcher = agrandir D en un ensemble P plus simple
2. couper = séparer D et un point de P pour préciser la relaxation

Illustration : optimisation entière (IP)

$D = \{Ax \leqslant b, x \in \mathbb{Z}^n\}$

$P = \{Ax \leqslant b, x \in \mathbb{R}^n\}$ relax. linéaire

$D = \{A'x \leqslant b', x \in \mathbb{Z}^n\}$



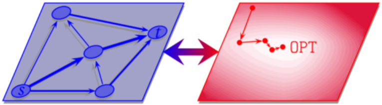
Deux techniques... intrinsèquement liées à l'optimisation convexe

borne/relaxation $\longleftrightarrow$  dualité

plan séparateur $\longrightarrow$ géométrie et analyse convexe

$\longrightarrow$ algorithmes de faisceaux

Interaction entre optimisation discrète et continue



- Liens de fond entre optimisation discrète et continue
- Rapprochement des deux communautés, en partie poussé par des problèmes industriels complexes et des applications en science des données
- Une partie de mon activité entre ces deux domaines, sur trois sujets
- Pour chaque sujet : présentation en trois temps
    1. Contexte
    2. Contributions
    3. Zoom sur un point plus technique

Plan de la présentation

1. Introduction
   - Un travail en optimisation mathématique...
   - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
   - Optimisation discrète : analyse des fonctions de coupes
   - Optimisation semidéfinie : problèmes binaires quadratiques
   - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
   - Optimisation avec données incertaines
   - Optimisation avec données distribuées
   - Optimisation avec données structurées

Plan de la présentation

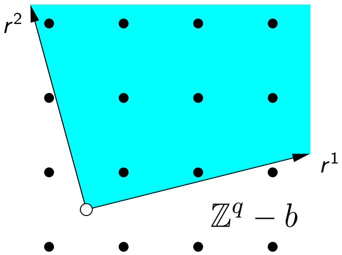
Contexte : coupes en optimisation discrète

Optimisation en nombres entiers

- les coupes permettent d'affiner les domaines réalisables
- ajout de coupes dans les logiciels $\rightarrow$ temps de calculs / 1 000
- coupes génériques, variées et peu coûteuses

Exemple : vision géométrique

séparer 0 de l'ensemble des éléments de $\mathbb{Z}^q - b$ dans la partie bleue



Contexte : coupes en optimisation discrète

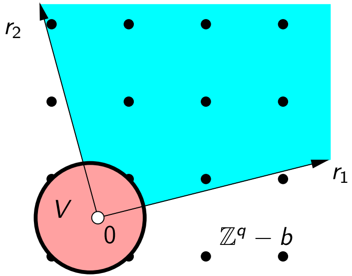
Optimisation en nombres entiers

- les coupes permettent d'affiner les domaines réalisables
- ajout de coupes dans les logiciels $\rightarrow$ temps de calculs / 1 000
- coupes génériques, variées et peu coûteuses

Exemple : vision géométrique

séparer 0 de l'ensemble des éléments de $\mathbb{Z}^q - b$ dans la partie bleue

On peut construire des coupes à partir d'ensembles V ne contenant pas $\mathbb{Z}^q - b$



Contexte : coupes en optimisation discrète

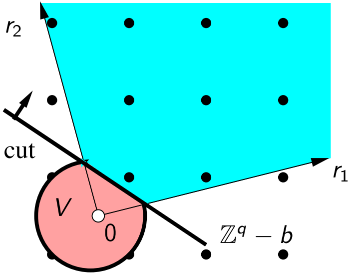
Optimisation en nombres entiers

- les coupes permettent d'affiner les domaines réalisables
- ajout de coupes dans les logiciels $\rightarrow$ temps de calculs / 1 000
- coupes génériques, variées et peu coûteuses

Exemple : vision géométrique

séparer 0 de l'ensemble des éléments de $\mathbb{Z}^q - b$ dans la partie bleue

On peut construire des coupes à partir d'ensembles V ne contenant pas $\mathbb{Z}^q - b$



Contexte : coupes en optimisation discrète

Optimisation en nombres entiers

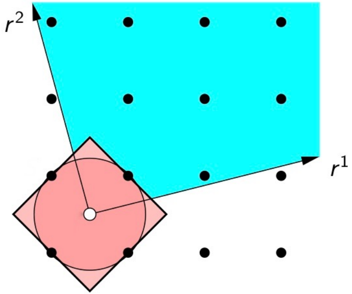
- les coupes permettent d'affiner les domaines réalisables
- ajout de coupes dans les logiciels $\rightarrow$ temps de calculs / 1 000
- coupes génériques, variées et peu coûteuses

Exemple : vision géométrique

séparer 0 de l'ensemble des éléments de $\mathbb{Z}^q - b$ dans la partie bleue

On peut construire des coupes à partir d'ensembles V ne contenant pas $\mathbb{Z}^q - b$

Des ensembles plus grands donnent de meilleures coupes



Contexte : coupes en optimisation discrète

Optimisation en nombres entiers

- les coupes permettent d'affiner les domaines réalisables
- ajout de coupes dans les logiciels $\rightarrow$ temps de calculs / 1 000
- coupes génériques, variées et peu coûteuses

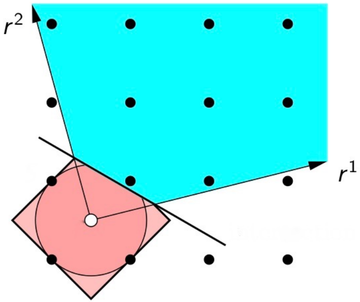
Exemple : vision géométrique

séparer 0 de l'ensemble des éléments de $\mathbb{Z}^q - b$ dans la partie bleue

On peut construire des coupes à partir d'ensembles V ne contenant pas $\mathbb{Z}^q - b$

Des ensembles plus grands donnent de meilleures coupes

Cela se formalise mathématiquement...



Contributions sur ce thème

Travail avec Gérard Cornuéjols (CMU, USA), Michele Conforti (Padova, Italie), Aris Daniilidis (Santiago, Chile), Claude Lemaréchal (ex-Inria)

Contributions (mathématiques)

1. introduction des fonctions génératrices de coupes (généralisant Gomory)
2. étude théorique des coupes $\longleftarrow$ rôle-clé de l'analyse convexe !
3. équivalence entre coupes "minimales" et les ensembles "maximaux"
   dans des situations générales

Contributions sur ce thème

Travail avec Gérard Cornuéjols (CMU, USA), Michele Conforti (Padova, Italie), Aris Daniilidis (Santiago, Chile), Claude Lemaréchal (ex-Inria)

Contributions (mathématiques)

1. introduction des fonctions génératrices de coupes (généralisant Gomory)
2. étude théorique des coupes $\longleftarrow$ rôle-clé de l'analyse convexe !
3. équivalence entre coupes "minimales" et les ensembles "maximaux" dans des situations générales

Zoom : affiner des résultats d'analyse convexe

- Analyse convexe : $V \longleftrightarrow V^\circ = \left\{d : d^\top r \leqslant 1 \text{ pour tout } r \in V\right\}$ pour V voisinage convexe compact de 0
- Généralisation à des voisinages non-bornés

$$G^\circ = V \quad \Longleftrightarrow \quad V^\bullet \subset \overline{\mathrm{conv}}\,(G) \subset V^\circ$$

- Propriétés subtiles de $V^\bullet$ + définition des coupes $\rho(r) = \sup_{v \in V^\bullet} v^\top r$

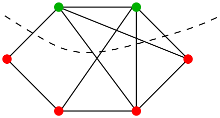
Plan de la présentation

Contexte : optimisation quadratique sous contraintes quadratiques

Optimisation quadratique ($\subset$ optimisation polynomiale, Lasserre 01)

$$\left\{\begin{array}{ll} \max & x^\top Q_0\, x + {q_0}^\top x \\ & x^\top Q_i\, x + {q_i}^\top x = (\text{ou} \leqslant)\ a_i, \quad i \in \{1,\dots,m\} \end{array}\right.$$

Exemple le plus simple : max-cut



$$\left\{\begin{array}{ll} \max & x^\top Q\, x \\ & x \in \{0,1\}^n \quad (\Leftrightarrow x^2 - x = 0) \end{array}\right.$$

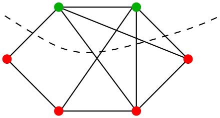
faire une partition de l'ensemble des sommets du graphe telle que le nombre d'arêtes à cheval soit maximal

Contexte : optimisation quadratique sous contraintes quadratiques

Optimisation quadratique ($\subset$ optimisation polynomiale, Lasserre 01)

$$\begin{cases} \max \quad x^\top Q_0\, x + {q_0}^\top x \\ \quad x^\top Q_i\, x + {q_i}^\top x = (\text{ou} \leqslant)\ a_i, \quad i \in \{1,\dots,m\} \end{cases}$$

Exemple le plus simple : max-cut

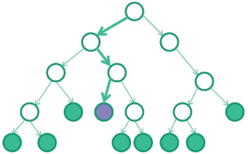


$$\begin{cases} \max \quad x^\top Q\, x \\ \quad x \in \{0,1\}^n \quad (\Leftrightarrow x^2 - x = 0) \end{cases}$$

faire une partition de l'ensemble des sommets du graphe telle que le nombre d'arêtes à cheval soit maximal

En général :

- Problèmes (NP-)difficiles en théorie
  ... et en pratique !
- Résolution effective par énumération intelligente : branch & bound



- "bons" majorants avec bon équilibre : précision $\longleftrightarrow$ temps de calculs

Contexte : relaxation semidéfinie (SDP)

La relaxation SDP $\longleftarrow$ dualisation des contraintes quadratiques
(optimisation sur le cône convexe des matrices semidéfinies positives)

Exemple : relaxation SDP de max-cut

$$(\text{max-cut}) \left\{ \begin{array}{l} \max \quad x^\top Q\, x \\ \quad x \in \{0,1\}^n \end{array} \right. \qquad (\text{SDP}) \left\{ \begin{array}{l} \max \quad \operatorname{trace}(QX) \\ \quad \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \end{array} \right.$$

Contexte : relaxation semidéfinie (SDP)

La relaxation SDP $\longleftarrow$ dualisation des contraintes quadratiques
(optimisation sur le cône convexe des matrices semidéfinies positives)

Exemple : relaxation SDP de max-cut

$$(\text{max-cut}) \begin{cases} \max & x^\top Q\, x \\ & x \in \{0,1\}^n \end{cases} \qquad (\text{SDP}) \begin{cases} \max & \operatorname{trace}(QX) \\ & \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \end{cases}$$

En général, relaxations SDP : précises + bonnes propriétés théoriques

Approximation : $\frac{2}{\pi}\,\mathrm{val}\,(\text{SDP}) \leqslant \mathrm{val}\,(\text{max-cut}) \leqslant \mathrm{val}\,(\text{SDP})$ si $Q \succcurlyeq 0$ [Nesterov '98]

Renforcer avec des relaxations d'ordres supérieurs [Lasserre 01] ou des coupes
(ex : inégalités triangulaires)

$$\begin{cases} \max & \operatorname{trace}(QX) \\ & \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \\ & A_I(X) + \mathbf{1} \geqslant 0 \end{cases}$$

Contexte : relaxation semidéfinie (SDP)

La relaxation SDP $\longleftarrow$ dualisation des contraintes quadratiques
(optimisation sur le cône convexe des matrices semidéfinies positives)

Exemple : relaxation SDP de max-cut

$$(\text{max-cut}) \begin{cases} \max & x^\top Q\, x \\ & x \in \{0,1\}^n \end{cases} \qquad (\text{SDP}) \begin{cases} \max & \operatorname{trace}(QX) \\ & \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \end{cases}$$

En général, relaxations SDP : précises + bonnes propriétés théoriques

Approximation : $\frac{2}{\pi}\,\text{val}\,(\text{SDP}) \leqslant \text{val}\,(\text{max-cut}) \leqslant \text{val}\,(\text{SDP})$ si $Q \succcurlyeq 0$ [Nesterov '98]

Renforcer avec des relaxations d'ordres supérieurs [Lasserre 01] ou des coupes
(ex : inégalités triangulaires)

$$\begin{cases} \max & \operatorname{trace}(QX) \\ & \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \\ & A_I(X) + \mathbf{1} \geqslant 0 \end{cases}$$

Problème : coût numérique exorbitant... SDP inutilisable en pratique?

$$\boxed{\text{précision}} \longleftrightarrow \boxed{\text{temps de calcul}}$$

relax. SDP — relax. linéaires ou quadratiques

Contexte : relaxation semidéfinie (SDP)

La relaxation SDP $\longleftarrow$ dualisation des contraintes quadratiques
(optimisation sur le cône convexe des matrices semidéfinies positives)

Exemple : relaxation SDP de max-cut

$$(\text{max-cut}) \left\{ \begin{array}{l} \max \quad x^\top Q\, x \\ \qquad x \in \{0,1\}^n \end{array} \right. \qquad\qquad (\text{SDP}) \left\{ \begin{array}{l} \max \quad \operatorname{trace}(QX) \\ \qquad \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \end{array} \right.$$

En général, relaxations SDP : précises + bonnes propriétés théoriques

Approximation : $\frac{2}{\pi}\,\text{val}\,(\text{SDP}) \leqslant \text{val}\,(\text{max-cut}) \leqslant \text{val}\,(\text{SDP})$ si $Q \succcurlyeq 0$ [Nesterov '98]

Renforcer avec des relaxations d'ordres supérieurs [Lasserre 01] ou des coupes
(ex : inégalités triangulaires)

$$\left\{ \begin{array}{l} \max \quad \operatorname{trace}(QX) \\ \qquad \operatorname{diag}(X) = \mathbf{1}, \quad X \succcurlyeq 0 \\ \qquad A_l(X) + \mathbf{1} \geqslant 0 \end{array} \right.$$

Problème : coût numérique exorbitant... SDP inutilisable en pratique?

précision $\longleftrightarrow$ temps de calcul

relax. SDP

relax. linéaires
ou quadratiques

nouvelles relax. SDP (α)

Contributions sur ce thème

Travail avec Frédéric Roupin, Paris XIII (6 articles)

Contributions (théorique, algorithmique, numérique)

1. Nouvelle famille de relaxations de type SDP adjustables par un curseur

   Exemple pour max-cut :

$$F_I^{\alpha}(y,z) = \mathbf{1}^T y + \mathbf{1}^T z + \sum_{\text{valeur propre} \geqslant 0} \lambda_i \Big(Q - \mathrm{Diag}(y) + A_I^*(z)\Big)^2 / 2\alpha \quad \geqslant \text{(max-cut)}$$

2. Procédure de relaxation avec gestion automatique des paramètres
   - Entrelacement de la diminution de α_k et de la gestion coupes I_k
   - Preuve de convergence asymptotique vers la relaxation SDP
   - Solution réalisable et β_k par une procédure du type [Goemans-Williamson '95]

$$\beta_k \leqslant \quad \text{valeur optimale du (sous-)problème} \quad \leqslant F_k = F_{I_k}^{\alpha_k}(y_k, z_k)$$

3. Algorithme de résolution exacte pour ces problèmes quadratiques
   - Expérimentation numérique, comparaisons
   - Logiciel libre... et efficace ☺ $\quad \rightarrow$ oui, SDP est utile en pratique aussi !

Contributions sur ce thème

Travail avec Frédéric Roupin, Paris XIII (6 articles)

Contributions (théorique, algorithmique, numérique)

1. Nouvelle famille de relaxations de type SDP adjustables par un curseur

   Exemple pour max-cut :

$$F_I^{\alpha}(y,z) = \mathbf{1}^T y + \mathbf{1}^T z + \sum_{\text{valeur propre} \geqslant 0} \lambda_i \Big(Q - \mathrm{Diag}(y) + A_I^*(z) \Big)^2 / 2\alpha \quad \geqslant (\text{max-cut})$$

2. Procédure de relaxation avec gestion automatique des paramètres
   - Entrelacement de la diminution de α_k et de la gestion coupes I_k
   - Preuve de convergence asymptotique vers la relaxation SDP
   - Solution réalisable et β_k par une procédure du type [Goemans-Williamson '95]

$$\beta_k \leqslant \text{ valeur optimale du (sous-)problème } \leqslant F_k = F_{I_k}^{\alpha_k}(y_k, z_k)$$

3. Algorithme de résolution exacte pour ces problèmes quadratiques
   - Expérimentation numérique, comparaisons
   - Logiciel libre... et efficace ☺ $\rightarrow$ oui, SDP est utile en pratique aussi !

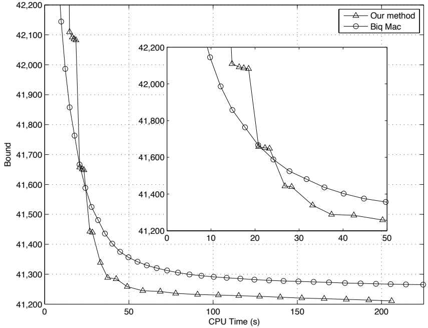
Zoom : illustration numérique sur max-cut

Numériquement, c'est le pire cas pour nous

– relaxation SDP standard simple
– logiciel concurrent

calcul de majorant

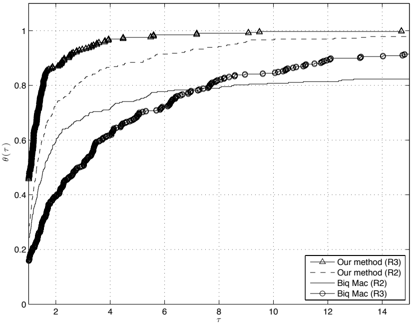
the lower the better!



sur une instance particulière (Beasley bqp250.6)

résolution exacte

the higher the better!



Profil de performance sur toutes les instances publiques

Plan de la présentation

Contexte : gestion de la production électrique

Problème hétérogène et de grande taille ($\sim 10^6$ variables, $\sim 10^6$ contraintes)

$$\left\{\begin{array}{lll} \min & \sum_i {c_i}^\top x_i & \text{(coûts de production)} \\ & \sum_i x_i = d & \text{(contraintes de demande couplantes)} \\ & (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N & \text{(contraintes techniques)} \end{array}\right.$$

Contexte : gestion de la production électrique

Problème hétérogène et de grande taille ($\sim 10^6$ variables, $\sim 10^6$ contraintes)

$$\left\{\begin{array}{lll} \min & \sum_i c_i^\top x_i & \text{(coûts de production)} \\ & \sum_i x_i = d \quad \twoheadleftarrow u \in \mathbb{R}^T & \text{(contraintes de demande couplantes)} \\ & (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N & \text{(contraintes techniques)} \end{array}\right.$$

La solution d'EDF :  approche par dualité [Lemaréchal *et al* '05]

Contexte : gestion de la production électrique

Problème hétérogène et de grande taille ($\sim 10^6$ variables, $\sim 10^6$ contraintes)

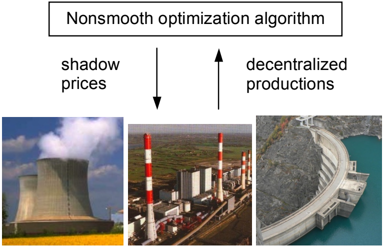
$$\left\{\begin{array}{ll} \min \quad \sum_i c_i^\top x_i & \text{(coûts de production)} \\ \quad \sum_i x_i = d \quad \leftarrow u \in \mathbb{R}^T & \text{(contraintes de demande couplantes)} \\ \quad (x_1, \ldots, x_N) \in X_1 \times \cdots \times X_N & \text{(contraintes techniques)} \end{array}\right.$$

La solution d'EDF :  approche par dualité [Lemaréchal *et al* '05]

– coordination (calcul d'un "prix" u) par un algorithme de faisceaux proximal inexact

– décomposition des calculs sur chaque centrale

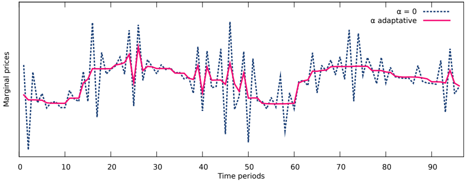
$$\left\{\begin{array}{l} \min \quad (c_i + A^\top u)^\top x_i \\ \quad x_i \in X_i \end{array}\right.$$

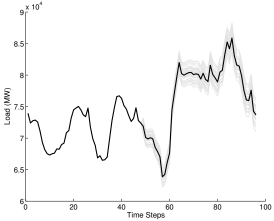


Recherche reliée à cette application industrielle

Contributions (modélisation, théorie, algorithme)

1. Sur le modèle EDF (Lemaréchal Malick '11)
   → étude théorique de l'interprétation des variables duales
2. Débruiter les solutions duales (Zaourar Malick '13)
   → implémentée dans le logiciel EDF
3. Une accélération de l'algorithme de faisceaux
   → utiliser des coupes imprécises disponibles (Malick Oliveira Zaourar '15)
4. Le modèle prenait mal en compte les énergies renouvelables
   → ajout de l'aléatoire dans le modèle (météo) (van Ackooij Malick '16)
   - problème d'optimisation stochastique à deux niveaux énorme
   - double décomposition par les scénarios et par les contraintes

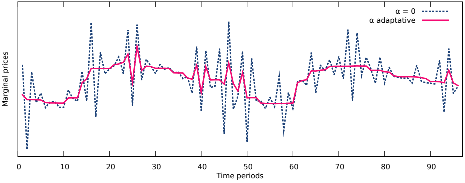


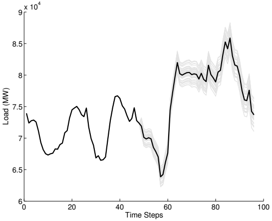


Recherche reliée à cette application industrielle

Contributions (modélisation, théorie, algorithme)

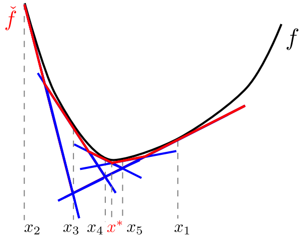
1. Sur le modèle EDF (Lemaréchal Malick '11)
   → étude théorique de l'interprétation des variables duales
2. Débruiter les solutions duales (Zaourar Malick '13)
   → implémentée dans le logiciel EDF
3. Une accélération de l'algorithme de faisceaux
   → utiliser des coupes imprécises disponibles (Malick Oliveira Zaourar '15)
4. Le modèle prenait mal en compte les énergies renouvelables
   → ajout de l'aléatoire dans le modèle (météo) (van Ackooij Malick '16)
   - problème d'optimisation stochastique à deux niveaux énorme
   - double décomposition par les scénarios et par les contraintes





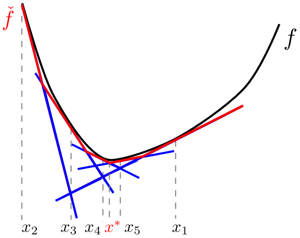
Zoom : coupes inexactes imprécises

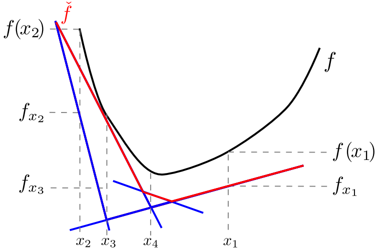
- Les méthodes de faisceaux accumulent des coupes
- Coupes exactes...
  ou inexactes controlées [Kiwiel '06]
- Et les coupes non-controlées ?



Zoom : coupes inexactes imprécises

- Les méthodes de faisceaux accumulent des coupes
- Coupes exactes...
  ou inexactes controlées [Kiwiel '06]
- Et les coupes non-controlées ?





Contribution avec Sofia Zaourar (Xerox) et Welington de Oliveira (Rio, Brésil)

- Algorithme de faisceaux adapté
- Propriétes théoriques préservées
  convergence asymptotique de l'algo, sans hypothèse de compacité
- Illustration du gain en pratique

Une difficulté technique : gérer les coupes inexactes dans l'algorithme de faisceaux par niveau [Lemaréchal Nesterov Nemirovski '95] $\longrightarrow$ règle implicite pour gérer la taille du pas en fonction du bruit

Plan de la présentation

1. Introduction
   - Un travail en optimisation mathématique...
   - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
   - Optimisation discrète : analyse des fonctions de coupes
   - Optimisation semidéfinie : problèmes binaires quadratiques
   - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
   - Optimisation avec données incertaines
   - Optimisation avec données distribuées
   - Optimisation avec données structurées

Programme de recherche en science des données

Données collectées en permanence :

- données des personnes et des clients (smart phone, réseaux sociaux, caméras, internet...)
- données scientifiques (biologie, génomique, astronomie,...)

Emergence d'un nouveau domaine scientifique : science des données
vise à extraire de l'information de ces masses de données complexes

→ défis sur toute la chaîne de traitement

→ l'optimisation est au coeur des outils numériques

3 projets en optimisation pour la science des données
$\subset$ direction de recherche de DAO au LJK

Pour chaque projet : présentation en deux temps

1. Contexte + idées
2. Zoom sur un point plus précis

Plan de la présentation

Optimisation avec données incertaines

Défi : prendre de “bonnes” décisions avec des données incertaines

Exemple : énergies renouvelables sujettes à de l'aléa météo
$\longrightarrow$ incertitude sur les réseaux électriques

En général : deux manières de modéliser l'incertain

optimisation robuste

$$\begin{cases} \min_x \quad \max_{\xi \in D} \quad f(x,\xi) \\ \quad g(x,\xi) \leqslant 0 \quad \text{pour tout } \xi \in D \end{cases}$$

optimisation stochastique

$$\begin{cases} \min_x \quad \mathbb{E}[f(x,\xi)] \\ \quad \mathbb{P}\big(g(x,\xi) \leqslant 0\big) \geqslant p \end{cases}$$

Optimisation avec données incertaines

Défi : prendre de "bonnes" décisions avec des données incertaines

Exemple : énergies renouvelables sujettes à de l'aléa météo
$\longrightarrow$ incertitude sur les réseaux électriques

En général : deux manières de modéliser l'incertain

optimisation robuste

$$\begin{cases} \min_x \quad \max_{\xi \in D} \quad f(x,\xi) \\ \quad g(x,\xi) \leqslant 0 \quad \text{pour tout } \xi \in D \end{cases}$$

optimisation stochastique

$$\begin{cases} \min_x \quad \mathbb{E}[f(x,\xi)] \\ \quad \mathbb{P}\big(g(x,\xi) \leqslant 0\big) \geqslant p \end{cases}$$

Constat : écart énorme entre ce qu'on peut modéliser et ce qu'on peut résoudre
En particulier : contraintes en probabilité peu utilisées en pratique

Directions de recherche

- théorie : propriétés des fonctions en probabilité (différentiabilité/convexité)
- algorithme : méthodes faisceaux adaptées
  - pour l'optimisation robuste : naturel ? (D généraux vs reformulation)
  - pour optimisation avec contraintes en probabilité

    + approximations numériques – complications dues à la non-convexité

Optimisation avec données incertaines

Défi : prendre de “bonnes” décisions avec des données incertaines

Exemple : énergies renouvelables sujettes à de l'aléa météo
$\longrightarrow$ incertitude sur les réseaux électriques

En général : deux manières de modéliser l'incertain

optimisation robuste

$$\begin{cases} \min_x \quad \max_{\xi \in D} \quad f(x,\xi) \\ \quad g(x,\xi) \leqslant 0 \quad \text{pour tout } \xi \in D \end{cases}$$

optimisation stochastique

$$\begin{cases} \min_x \quad \mathbb{E}[f(x,\xi)] \\ \quad \mathbb{P}\big(g(x,\xi) \leqslant 0\big) \geqslant p \end{cases}$$

Constat : écart énorme entre ce qu'on peut modéliser et ce qu'on peut résoudre
En particulier : contraintes en probabilité peu utilisées en pratique

Directions de recherche

- théorie : propriétés des fonctions en probabilité (différentiabilité/convexité)
- algorithme : méthodes faisceaux adaptées
  - pour l'optimisation robuste : naturel ? (D généraux vs reformulation)
  - pour optimisation avec contraintes en probabilité

    $+$ approximations numériques $-$ complications dues à la non-convexité

Zoom : analyse des fonctions contraintes en probabilité

Etude mathématique de $\varphi\colon x \mapsto \mathbb{P}\big(g(x,\xi) \leqslant 0\big)$

- continuité, convexité, différentiabilité de g ne passent pas à φ
- analyse variationnelle (différentiabilité généralisée) [Henrion '14]
- différentiabilité seconde sous des hypothèses légères pour de nombreuses lois classiques (normal, log-normal, student,...) [van Ackooij Malick '16]

Zoom : analyse des fonctions contraintes en probabilité

Etude mathématique de $\varphi\colon x \mapsto \mathbb{P}\big(g(x,\xi) \leqslant 0\big)$

- continuité, convexité, différentiabilité de g ne passent pas à φ
- analyse variationnelle (différentiabilité généralisée) [Henrion '14]
- différentiabilité seconde sous des hypothèses légères pour de nombreuses lois classiques (normal, log-normal, student,...) [van Ackooij Malick '16]

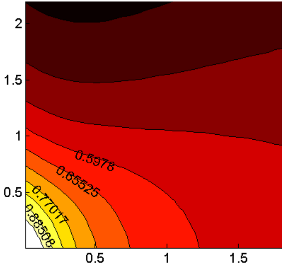
Projet avec Wim van Ackooij (EDF)

Question importante pour l'optimisation : convexité de la contrainte $\varphi(x) \geqslant p$?
(Résultats connus [Prepoka '95] ne couvrent pas $A(\xi)x \leqslant b$)

Avec des hypothèses géométriques sur g, on a convexité finale : il existe $\bar{p} < 1$ tel que $\bar{p} \leqslant p \leqslant 1$

$$\{x : \mathbb{P}\big(g(x,\xi) \leqslant 0\big) \geqslant p\} \quad \text{convexe}$$

C'est ce qu'on observe :
sur les courbes de niveaux de φ



Plan de la présentation

Optimisation avec données distribuées

Défi : faire passer à l'échelle les algorithmes pour résoudre les problèmes d'optimisation de très grande taille en science des données

Distribution des données/calculs : spécificités de l'optimisation

- problèmes d'optimisation : fortement structurés
- algorithmes d'optimisation : capacité de s'auto-corriger + capacité de gérer des calculs inexacts

Technologies récentes aident à franchir le pas

logiciel calcul distribué   virtualisation   des plateformes de calculs

Direction de recherche avec Nabil Layaida (Inria) et toute une équipe

- Conception, analyse, et déploiement réel d'algorithmes d'optimisation
- Réduire l'écart entre les algorithmes et les nouveaux systèmes distribués

Optimisation avec données distribuées

Défi : faire passer à l'échelle les algorithmes pour résoudre les problèmes d'optimisation de très grande taille en science des données

Distribution des données/calculs : spécificités de l'optimisation

- problèmes d'optimisation : fortement structurés
- algorithmes d'optimisation : capacité de s'auto-corriger + capacité de gérer des calculs inexacts

Technologies récentes aident à franchir le pas

logiciel calcul distribué | virtualisation | des plateformes de calculs

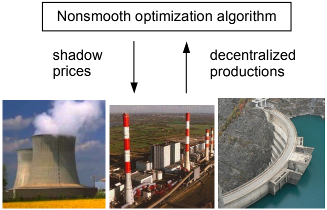
Direction de recherche avec Nabil Layaida (Inria) et toute une équipe

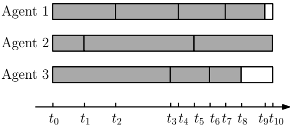
- Conception, analyse, et déploiement réel d'algorithmes d'optimisation
- Réduire l'écart entre les algorithmes et les nouveaux systèmes distribués

Zoom : algorithmes de décomposition asynchrones

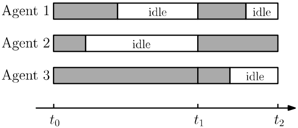
Constat : les algorithmes d'optimisation sont séquentiels ou parallélisables mais de manière synchrone

Ex : algorithme décomposition
(RO ou optim. stochastique)





asynchrone



synchrone

Projet avec Franck Iutzeler (LJK, DAO)

Versions asynchrones des méthodes de décomposition en optimisation convexe

- méthodes de faisceaux (information asynchrone $\simeq$ coupes incontrolées)
- algorithme d'optimisation stochastique : "progressive hedging" $\simeq$ ADMM

Plan de la présentation

Problèmes d'optimisation fortement structurés en analyse de données

Exemple : problèmes inverses régularisés

$$\min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^{N} f_i(x) + \lambda\, R(x)$$

avec R non-lisse $\rightarrow$ structure de faible complexité (parcimonie, par bloc, du spectre...)

La décomposition des calcul augmente encore la taille... de manière structurée !

Problèmes d'optimisation fortement structurés en analyse de données

Exemple : problèmes inverses régularisés

$$\min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^{N} f_i(x) + \lambda\, R(x)$$

avec R non-lisse $\rightarrow$ structure de faible complexité (parcimonie, par bloc, du spectre...)

La décomposition des calcul augmente encore la taille... de manière structurée !

Direction de recherche

Analyse mathématique : déjà une riche activité (ex : [Bolte *et al* '16])

Aller au-delà des problèmes convexes bien posés

- non-convexité en apprentissage et statistique (ex : modèles non-linéaires)
- mauvais conditionnements (ex : imagerie médicale)

Première piste : stabilité des problèmes inverses dans les cas limites

$\rightarrow$ Analyse de stabilité théorique (identification)

$\rightarrow$ Accélérations des algorithmes ? réglages de paramètres ?

Problèmes d'optimisation fortement structurés en analyse de données

Exemple : problèmes inverses régularisés

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{N} f_i(x) + \lambda\, R(x)$$

avec R non-lisse $\rightarrow$ structure de faible complexité (parcimonie, par bloc, du spectre...)

La décomposition des calcul augmente encore la taille... de manière structurée !

Direction de recherche

Analyse mathématique : déjà une riche activité (ex : [Bolte *et al* '16])

Aller au-delà des problèmes convexes bien posés

- non-convexité en apprentissage et statistique (ex : modèles non-linéaires)
- mauvais conditionnements (ex : imagerie médicale)

Première piste : stabilité des problèmes inverses dans les cas limites

$\rightarrow$ Analyse de stabilité théorique (identification)

$\rightarrow$ Accélérations des algorithmes ? réglages de paramètres ?

Zoom : analyse de stabilité généralisée des problèmes inverses

Exemple "compressed sensing"

Retrouver $\bar{x}$ parcimonieux à partir de l'observation $y = \Phi\bar{x} + w$

On résout $(\text{lasso})_y \quad \min_x \|y - \Phi x\|^2 + \lambda\|x\|_{\ell_1}$

[Candès *et al* '05] : on retrouve le support de $\bar{x}$ si il est suffisamment petit
(avec probabilité 1, si Φ gaussienne et bruit petit)

Zoom : analyse de stabilité généralisée des problèmes inverses

Exemple "compressed sensing"

Retrouver $\bar{x}$ parcimonieux à partir de l'observation $y = \Phi\bar{x} + w$

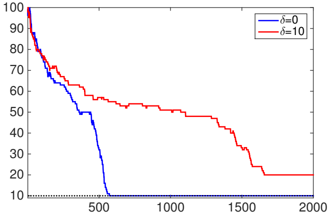
On résout $(\text{lasso})_y \quad \min_x \|y - \Phi x\|^2 + \lambda\|x\|_{\ell_1}$

[Candès *et al* '05] : on retrouve le support de $\bar{x}$ si il est suffisamment petit
(avec probabilité 1, si Φ gaussienne et bruit petit)

Illustration hors du cadre du résultat

$\bar{x}_1$

on résout $(\text{lasso})_{y_1}$
par gradient-proximal, et on affiche $\|x_k\|_0$



Zoom : analyse de stabilité généralisée des problèmes inverses

Exemple "compressed sensing"

Retrouver $\bar{x}$ parcimonieux à partir de l'observation $y = \Phi\bar{x} + w$

On résout $(\text{lasso})_y \quad \min_x \|y - \Phi x\|^2 + \lambda\|x\|_{\ell_1}$

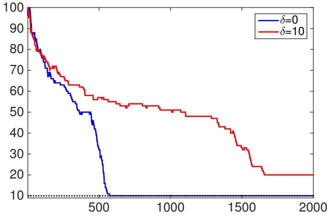
[Candès *et al* '05] : on retrouve le support de $\bar{x}$ si il est suffisamment petit
(avec probabilité 1, si Φ gaussienne et bruit petit)

Illustration hors du cadre du résultat

$\bar{x}_1$ et $\bar{x}_2$ avec $\|\bar{x}_1\|_0 = \|\bar{x}_2\|_0$

on résout $(\text{lasso})_{y_1}$ et $(\text{lasso})_{y_2}$
par gradient-proximal, et on affiche $\|x_k\|_0$

on observe une certaine identification...



Zoom : analyse de stabilité généralisée des problèmes inverses

Exemple "compressed sensing"

Retrouver $\bar{x}$ parcimonieux à partir de l'observation $y = \Phi\bar{x} + w$

On résout (lasso)$_y$ $\min_x \|y - \Phi x\|^2 + \lambda\|x\|_{\ell_1}$

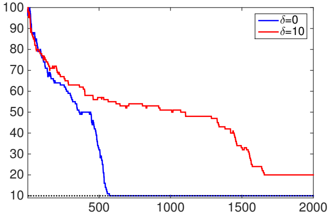
[Candès *et al* '05] : on retrouve le support de $\bar{x}$ si il est suffisamment petit
(avec probabilité 1, si Φ gaussienne et bruit petit)

Illustration hors du cadre du résultat

$\bar{x}_1$ et $\bar{x}_2$ avec $\|\bar{x}_1\|_0 = \|\bar{x}_2\|_0$

on résout (lasso)$_{y_1}$ et (lasso)$_{y_2}$
par gradient-proximal, et on affiche $\|x_k\|_0$

on observe une certaine identification...



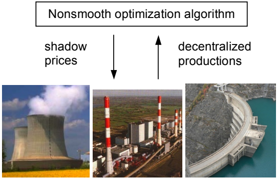
Projet avec Gabriel Peyré, ENS Ulm

- résultat d'identification général (couvrant les cas limites/dégénérés)
- exploiter la structure des régularisations pour obtenir des garanties de stabilité dans une variété élargie

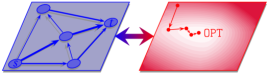
En guise de conclusion

Messages de cette présentation

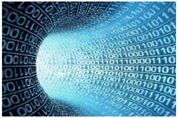
optimisation mathématique :
des maths utiles !



intersection de l'optimisation
combinatoire et convexe

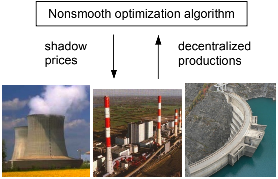


applications en science des données

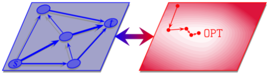


En guise de conclusion

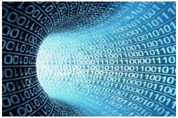
Messages de cette présentation



optimisation mathématique :
des maths utiles !



intersection de l'optimisation
combinatoire et convexe



applications en science des données

Sélection de contributions et de projets

une philosophie commune : réduire l'écart en théorie et pratique

- motiver les développements théoriques par des applications
- unifier, clarifer les résultats pour les rendre accessibles aux non-spécialistes

Merci à tous mes co-auteurs !

Et merci à vous pour votre présence aujourd'hui !

Plan de la présentation

1. Introduction
    - Un travail en optimisation mathématique...
    - ...à l'intersection entre optimisation continue et discrète
2. Sélection de sujets et contributions
    - Optimisation discrète : analyse des fonctions de coupes
    - Optimisation semidéfinie : problèmes binaires quadratiques
    - Optimisation non-lisse : méthodes de faisceaux inexactes
3. Programme de recherche
    - Optimisation avec données incertaines
    - Optimisation avec données distribuées
    - Optimisation avec données structurées