Optimisation et convexité : introduction, motivations, exemples

Claude LEMARÉCHAL   Jérôme MALICK

École de recherche "Optimisation & Convexité" – Partie 1

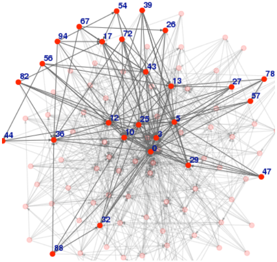
ENS Lyon – Janvier 2011

Plan de la présentation

Quel est le point commun ?

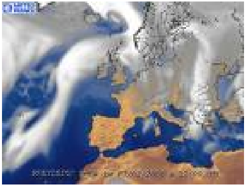

production électrique


robotique, mécanique



graphes, réseaux

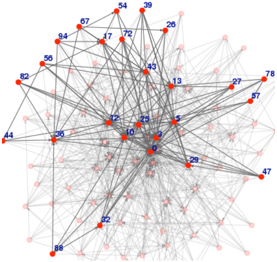
risque en finance


météo

Quel est le point commun ?

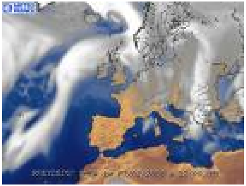
production électrique

robotique, mécanique



graphes, réseaux

risque en finance



météo

Réponse :

optimisation
(convexe)

Optimisation... convexe

L'optimisation en 2 mots

- "la science du mieux-faire" ou "la science de la décision"
- Discipline mature des maths applis (théorie, algos, logiciels)
- Explosion récente des applications dans les sciences et technologies

Optimisation... convexe

L'optimisation en 2 mots

- "la science du mieux-faire" ou "la science de la décision"
- Discipline mature des maths applis (théorie, algos, logiciels)
- Explosion récente des applications dans les sciences et technologies

Et pourquoi convexe ?

- Propriétés géométriques : globalité, garanties,...
- Précieux outils: dualité, analyse de sensibilité...
- Résoudre des problèmes non-convexes à l'aide de l'optimisation convexe

Optimisation... convexe

L'optimisation en 2 mots

- "la science du mieux-faire" ou "la science de la décision"
- Discipline mature des maths applis (théorie, algos, logiciels)
- Explosion récente des applications dans les sciences et technologies

Et pourquoi convexe ?

- Propriétés géométriques : globalité, garanties,...
- Précieux outils: dualité, analyse de sensibilité...
- Résoudre des problèmes non-convexes à l'aide de l'optimisation convexe

Un cours en optimisation convexe ?

- Optim. linéaire vs. non-linéaire, déterministe vs. stochastique...
- Théorie, algorithmes, applications,...
- Il y a beaucoup (trop) à dire, on va se concentrer sur un thème...

Optimisation convexe non-différentiable

Objectif : Optimisation convexe non-différentiable

- Encore peu connue mais en pleine expansion (théorie, applications)
- Application en optimisation combinatoire (antonymiques ?)
- Applications en ingénierie, en apprentissage,...

Optimisation convexe non-différentiable

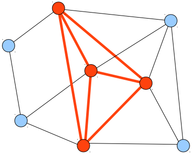
Objectif : Optimisation convexe non-différentiable

- Encore peu connue mais en pleine expansion (théorie, applications)
- Application en optimisation combinatoire (antonymiques ?)
- Applications en ingénierie, en apprentissage,...

Contenu du cours

- Introduction aux bases de l'optimisation (algorithmes)
- Illustrer avec des exemples d'applications (apprentissage, météo...)
- Présentation de l'analyse convexe et de l'algorithmique non-diff.
- Insister sur leur utilisation dans deux domaines :

production électrique



graphes et applications

Plan de cette semaine : "Optimisation & Convexité"

- Lundi (10h30 - 12h 13h30 - 15h30)
  - Introduction générale, exemples (partie 1)
- Mardi (10h30 - 12h 13h30 - 17h30)
  - "Rappels" d'optimisation différentiable (partie 2)
  - TP – optimisation numérique (scilab)
- Mercredi (10h30 - 12h 13h30 - 17h30)
  - Introduction à optimisation non-différentiable (partie 3.1)
  - TD – dualité convexe
- Jeudi (10h30 - 12h 13h30 - 17h30)
  - Théorie de l'optimisation convexe (partie 3.2)
  - TD – analyse convexe
- Vendredi (9h - 12h 13h30 - 15h30)
  - Algorithmique de l'optimisation non-différentiable (partie 3.3)
  - Examen – optimisation dans les réseaux telecom

Plan d'aujourd'hui (Partie 1) : Introductions, exemples

1. **Vue d'ensemble**
2. **Introduction à l'optimisation**
3. **Introduction à l'analyse convexe**
4. **Exemples de problèmes d'optimisation convexe**
   - Exemple dans l'industrie : gestion de la production électrique
   - Exemples dans un domaine scientifique : apprentissage statistique

Plan de la présentation

C'est quoi l'optimisation ?

- **Problème d'optimisation :** formulation mathématique

$$\boxed{\left\{ \begin{array}{ll} \min & f(x) \\ & x \in C \end{array} \right.}$$

  - variable $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$
  - fonction objectif $f\colon \mathbb{R}^n \to \mathbb{R}$
  - ensemble des contraintes $C \subset \mathbb{R}^n$

C'est quoi l'optimisation ?

- **Problème d'optimisation :** formulation mathématique

$$\boxed{\left\{\begin{array}{cc} \min & f(x) \\ & x \in C \end{array}\right.}$$

  - variable $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$
  - fonction objectif $f \colon \mathbb{R}^n \to \mathbb{R}$
  - ensemble des contraintes $C \subset \mathbb{R}^n$

- **Solution :** trouver $\bar{f} \in \mathbb{R}$ et $\bar{x} \in C$ tel que

$$f(\bar{x}) = \bar{f} \leqslant f(x) \quad \text{pour tout } x \in C$$

C'est quoi l'optimisation ?

- **Problème d'optimisation :** formulation mathématique

$$\boxed{\begin{cases} \min & f(x) \\ & x \in C \end{cases}}$$

  - variable $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$
  - fonction objectif $f \colon \mathbb{R}^n \to \mathbb{R}$
  - ensemble des contraintes $C \subset \mathbb{R}^n$

- **Solution :** trouver $\bar{f} \in \mathbb{R}$ et $\bar{x} \in C$ tel que

$$f(\bar{x}) = \bar{f} \leqslant f(x) \quad \text{pour tout } x \in C$$

- L'**"optimisation"**, c'est au moins trois choses :
  1. L'art de formuler les problèmes (de décision)
  2. Une théorie mathématique
  3. Des techniques algorithmiques

Exemple : moindres carrés parcimonieux

- Pb : trouver quelques facteurs socio-économiques importants qui expliquent les prix de l'immobiler (prix du m^2)
- Données :
  - n régions, p facteurs : données normalisées $A \in \mathbb{R}^{n \times p}$
  - prix moyen par région $b \in \mathbb{R}^n$

Exemple : moindres carrés parcimonieux

- Pb : trouver quelques facteurs socio-économiques importants qui expliquent les prix de l'immobiler (prix du m^2)
- Données :
  - n régions, p facteurs : données normalisées $A \in \mathbb{R}^{n \times p}$
  - prix moyen par région $b \in \mathbb{R}^n$
- Modélisation :
  - trouver $x \in \mathbb{R}^p$ qui donne $Ax \approx b$
  - avec peu de $x_i \neq 0$

Exemple : moindres carrés parcimonieux

- Pb : trouver quelques facteurs socio-économiques importants qui expliquent les prix de l'immobiler (prix du m^2)
- Données :
  - n régions, p facteurs : données normalisées $A \in \mathbb{R}^{n \times p}$
  - prix moyen par région $b \in \mathbb{R}^n$
- Modélisation :
  - trouver $x \in \mathbb{R}^p$ qui donne $Ax \approx b$
  - avec peu de $x_i \neq 0$
- Problème d'optimisation (convexe, non-différentiable)

$$\min_{x \in \mathbb{R}^p} \quad \|A\, x - b\|^2 + \alpha \, \|x\|_1$$

Exemple : moindres carrés parcimonieux

- Pb : trouver quelques facteurs socio-économiques importants qui expliquent les prix de l'immobiler (prix du m^2)
- Données :
  - n régions, p facteurs : données normalisées $A \in \mathbb{R}^{n \times p}$
  - prix moyen par région $b \in \mathbb{R}^n$
- Modélisation :
  - trouver $x \in \mathbb{R}^p$ qui donne $Ax \approx b$
  - avec peu de $x_i \neq 0$
- Problème d'optimisation (convexe, non-différentiable)

$$\min_{x \in \mathbb{R}^p} \quad \|A\,x - b\|^2 + \alpha\,\|x\|_1$$

- Problème au coeur du "compressed sensing" (traitement du signal)

Exemple : moindres carrés parcimonieux

- Pb : trouver quelques facteurs socio-économiques importants qui expliquent les prix de l'immobiler (prix du m^2)
- Données :
  - n régions, p facteurs : données normalisées $A \in \mathbb{R}^{n \times p}$
  - prix moyen par région $b \in \mathbb{R}^n$
- Modélisation :
  - trouver $x \in \mathbb{R}^p$ qui donne $Ax \approx b$
  - avec peu de $x_i \neq 0$
- Problème d'optimisation (convexe, non-différentiable)

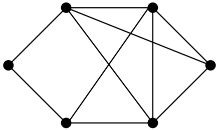
$$\min_{x \in \mathbb{R}^p} \quad \|A\,x - b\|^2 + \alpha\,\|x\|_1$$

- Problème au coeur du "compressed sensing" (traitement du signal)
- Plus généralement : Problèmes inverses (très fréquents)

$$\boxed{\text{minimiser} \quad \text{erreur(données, modèle)} \ + \text{terme régularisateur}}$$

Exemple dans les graphes : coupe maximale

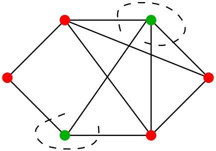
Graphe : ensemble de n noeuds
dont certains sont reliés par des arêtes



Exemple dans les graphes : coupe maximale

Graphe : ensemble de n noeuds
dont certains sont reliés par des arêtes

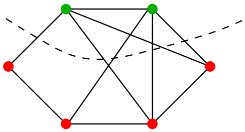
Coupe : partition de l'ensemble des
noeuds en deux (5 arêtes coupées)



Exemple dans les graphes : coupe maximale

Graphe : ensemble de n noeuds
dont certains sont reliés par des arêtes

Coupe : partition de l'ensemble des
noeuds en deux (6 arêtes coupées)

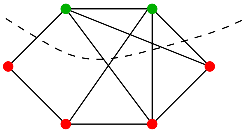


- Problème de la coupe maximale : trouver une coupe qui maximise le nombre d'arêtes coupées – ex d'appli. : physique statistique (!)

Exemple dans les graphes : coupe maximale

Graphe : ensemble de n noeuds
dont certains sont reliés par des arêtes

Coupe : partition de l'ensemble des noeuds en deux (6 arêtes coupées)



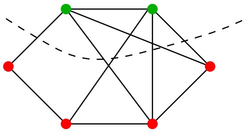
- Problème de la coupe maximale : trouver une coupe qui maximise le nombre d'arêtes coupées – ex d'appli. : physique statistique (!)
- Modélisation :
  - variable : $x \in \mathbb{R}^n$ (x_i pour le noeud i)
  - contrainte : $x_i = 1$ ou -1 (choix de l'ensemble)
  - objectif : nombre d'arêtes coupées (arête (ij) $\iff$ $a_{ij} \in \{0,1\}$)
    (ij) coupée $\iff$ $\big(a_{ij} = 1$ et $1 - x_i x_j = 0\big)$ $\iff$ $a_{ij}(1 - x_i x_j)/2 = 1$
- Formulation :

$$\begin{cases} \max \quad \sum_{ij} a_{ij}(1 - x_i x_j)/2 \\ \quad x \in \{-1, 1\}^n \end{cases}$$

Exemple dans les graphes : coupe maximale

Graphe : ensemble de n noeuds
dont certains sont reliés par des arêtes

Coupe : partition de l'ensemble des noeuds en deux (6 arêtes coupées)



- Problème de la coupe maximale : trouver une coupe qui maximise le nombre d'arêtes coupées – ex d'appli. : physique statistique (!)
- Modélisation :
  - variable : $x \in \mathbb{R}^n$ (x_i pour le noeud i)
  - contrainte : $x_i = 1$ ou -1 (choix de l'ensemble)
  - objectif : nombre d'arêtes coupées (arête $(ij) \iff a_{ij} \in \{0,1\}$)
    (ij) coupée $\iff \big(a_{ij} = 1$ et $1 - x_i x_j = 0\big) \iff a_{ij}(1 - x_i x_j)/2 = 1$
- Formulation : $Q = (-a_{ij}/2)_{ij}$ matrice d'adjacence (facteur $-1/2$)

$$\begin{cases} \min & x^\top Q\, x + \text{cste} \\ & x \in \{-1,1\}^n \end{cases}$$

- Problème "fini" mais dur ! (NP dur... OK pour $n \leqslant 500$)

Résoudre un problème d'optimisation

- En général

  $\boxed{\text{résoudre un problème d'optimisation est très difficile}}$

- Ce qu'on voudrait...
  - Situation idéale : calculer $\bar{x}$ et $\bar{f}$ explicitement
  - Bonne situation : avoir un algorithme qui génère une suite

$$x_k \longrightarrow \bar{x} \qquad f(x_k) \longrightarrow \bar{f}$$

Résoudre un problème d'optimisation

- En général

  $\boxed{\text{résoudre un problème d'optimisation est très difficile}}$

- Ce qu'on voudrait...
  - Situation idéale : calculer $\bar{x}$ et $\bar{f}$ explicitement
  - Bonne situation : avoir un algorithme qui génère une suite

$$x_k \to \bar{x} \qquad f(x_k) \to \bar{f}$$

- **Exemple** : optimisation linéaire

$$\left\{ \begin{array}{ll} \min & c^\top x \\ & Ax = b \\ & x \geqslant 0 \end{array} \right.$$

- théorie et algorithmes (2e guerre mondiale)
- logiciels efficaces et disponibles (20 ans)
- outils pour modéliser sous forme linéaire

Résoudre un problème d'optimisation

- En général

  $\boxed{\text{résoudre un problème d'optimisation est très difficile}}$

- Ce qu'on voudrait...
  - Situation idéale : calculer $\bar{x}$ et $\bar{f}$ explicitement
  - Bonne situation : avoir un algorithme qui génère une suite

$$x_k \to \bar{x} \qquad f(x_k) \to \bar{f}$$

- **Exemple** : optimisation linéaire

$$\begin{cases} \min \quad c^\top x \\ \quad Ax = b \\ \quad x \geqslant 0 \end{cases}$$

  - théorie et algorithmes (2e guerre mondiale)
  - logiciels efficaces et disponibles (20 ans)
  - outils pour modéliser sous forme linéaire

- Ce qu'on a souvent...
  - Pas de globalité : on a une sous-suite $x_{k'} \to \bar{x}$ minimum local

$$f(x_{k'}) \to f(\bar{x}) = \bar{f} \leqslant f(x) \quad \text{pour tout } x \in C \cap \mathrm{B}(\bar{x}, r)$$

  - Sous-optimalité : on a un minorant de la valeur optimale

$$m_k \to \bar{m} < \bar{f} \leqslant f(x) \quad \text{pour tout } x \in C$$

Résoudre un problème d'optimisation convexe

$$\left\{ \begin{array}{ll} \min & f(x) \\ & x \in C \end{array} \right. \qquad \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \text{ convexe} \\ C = \left\{ x \in \mathbb{R}^n : \ {a_i}^\top x = b_i, \ g_j(x) \leqslant 0 \right\} \text{ convexe} \end{array}$$

- Problèmes précédents $=$ manque de convexité !

Résoudre un problème d'optimisation convexe

$$\begin{cases} \min & f(x) \\ & x \in C \end{cases} \qquad \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \text{ convexe} \\ C = \left\{x \in \mathbb{R}^n : \ {a_i}^\top x = b_i,\ g_j(x) \leqslant 0\right\} \text{ convexe} \end{array}$$

- Problèmes précédents = manque de convexité !
- Pour les problèmes convexes, on est dans la "bonne" situation
- On dispose d'algorithmes : $f(x_k) \to \bar{f}$, $x_k \to \bar{x}$
- Contrôle du comportement $\longrightarrow$ théorie de la complexité
- En général,

$$\boxed{\text{résoudre un problème d'optimisation convexe est plus facile}}$$

Résoudre un problème d'optimisation convexe

$$\left\{\begin{array}{ll} \min & f(x) \\ & x \in C \end{array}\right. \qquad \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \text{ convexe} \\ C = \left\{x \in \mathbb{R}^n :\ {a_i}^\top x = b_i,\ g_j(x) \leqslant 0\right\} \text{ convexe} \end{array}$$

- Problèmes précédents = manque de convexité !
- Pour les problèmes convexes, on est dans la "bonne" situation
- On dispose d'algorithmes : $f(x_k) \to \bar{f}$, $x_k \to \bar{x}$
- Contrôle du comportement $\longrightarrow$ théorie de la complexité
- En général,

résoudre un problème d'optimisation convexe est plus facile

- Beaucoup de problèmes se modélisent comme des problèmes d'optimisation convexe (ex: prix de l'immobilier)
- Beaucoup de problèmes se résolvent à l'aide d'optimisation convexe (ex: résoudre un max-cut par branch-and-bound)

Se situer dans le temps...

Repères historiques pour l'optimisation convexe:

1900 Début de l'étude mathématique de la convexité (ex: H. Minkowski)

1947 Algorithme du simplexe pour l'optimisation linéaire (G. Dantzig)
Premières applications militaires puis en "recherche opérationnelle"

1970 Analyse convexe (W. Fenchel, J.-J. Moreau, T. Rockafellar)

1994 Algorithme de points intérieurs (Y. Nesterov & A. Nemirovski)

1990 $\rightarrow$ 2010

- nombreuses applications en sciences de l'ingénieur
(traitement du signal, réseaux, statistiques, robotique...)
- nouvelles familles de problèmes
(optimisation semidéfinie, optimisation robuste, stochastique...)

Se situer dans l'optimisation...

Nomenclature en optimisation:

- linéaire vs. non-linéaire (dichotomie classique)
- continue vs. discrète
- déterministe vs. stochastique (dans l'incertain)
- statique vs. dynamique (ex: commande optimale)
- unique vs. multi : critère ou décideur (ex: théorie des jeux)
- convexe vs. non-convexe
- autres: grande taille, incrémentale... (ex: apprentissage)

Se situer dans l'optimisation...

Nomenclature en optimisation:

- linéaire vs. non-linéaire (dichotomie classique)
- continue vs. discrète
- déterministe vs. stochastique (dans l'incertain)
- statique vs. dynamique (ex: commande optimale)
- unique vs. multi : critère ou décideur (ex: théorie des jeux)
- convexe vs. non-convexe
- autres: grande taille, incrémentale... (ex: apprentissage)

$\longrightarrow$ objectif : optimisation convexe non-différentiable
applications en combinatoire

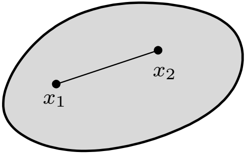
$\longrightarrow$ début : introduction aux base de l'optimisation numérique

Plan de la présentation

Ensemble convexe

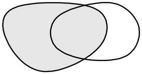
$C \subset \mathbb{R}^n$ est convexe si

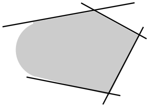
$$x_1, x_2 \in C, \quad 0 \leqslant \alpha \leqslant 1$$
$$\Longrightarrow \ \alpha\, x_1 + (1-\alpha) x_2 \in C$$



Exemples :

- Un espace affine $\{x \in \mathbb{R}^n : \ Ax = b\}$ est convexe
- Un demi-espace $\{x \in \mathbb{R}^n : \ a^\top x \leqslant \beta\}$ est convexe
- L'intersection (quelconque) d'ensembles convexes est convexe



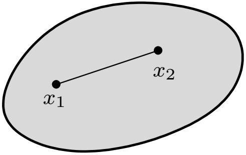


Ensemble convexe

$C \subset \mathbb{R}^n$ est convexe si

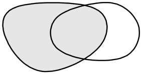
$$x_1, x_2 \in C, \quad 0 \leqslant \alpha \leqslant 1$$

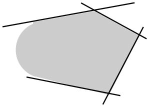
$$\Longrightarrow \; \alpha\, x_1 + (1-\alpha) x_2 \in C$$

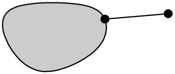


Exemples :

- Un espace affine $\{x \in \mathbb{R}^n : \; Ax = b\}$ est convexe
- Un demi-espace $\{x \in \mathbb{R}^n : \; a^\top x \leqslant \beta\}$ est convexe
- L'intersection (quelconque) d'ensembles convexes est convexe







Propriété de base :

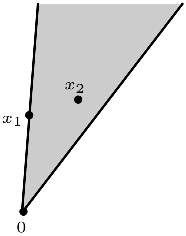
- Soit C convexe fermé; alors tout point $x \in \mathbb{R}^n$ a une unique projection sur C ! (caractéristique des ensembles convexes fermés)

Cône convexe

$C \subset \mathbb{R}^n$ est un cône convexe si

C est convexe

$x \in C,\ 0 < \alpha \implies \alpha\, x \in C$

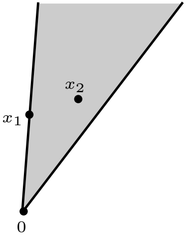


Cône convexe

$C \subset \mathbb{R}^n$ est un cône convexe si

C est convexe

$x \in C,\ 0 < \alpha \implies \alpha\, x \in C$



Exemples :

- L'ensemble des matrices (semidéfinies) positives

$$\mathcal{S}_n^+ = \{X \in \mathcal{S}_n :\ u^\top X u \geqslant 0 \text{ pour tout } u \in \mathbb{R}^n\}$$

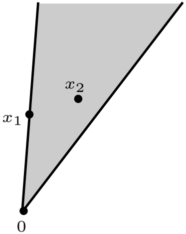
est un cône convexe (fermé)

Cône convexe

$C \subset \mathbb{R}^n$ est un cône convexe si

C est convexe

$x \in C,\ 0 < \alpha \implies \alpha\, x \in C$



Exemples :

- L'ensemble des matrices (semidéfinies) positives

$$\mathcal{S}_n^+ = \{X \in \mathcal{S}_n :\ u^\top X u \geqslant 0 \text{ pour tout } u \in \mathbb{R}^n\}$$

  est un cône convexe (fermé)

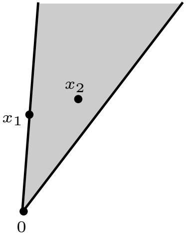
  Rem: ensemble convexe en statistiques : matrices de corrélation

Cône convexe

$C \subset \mathbb{R}^n$ est un cône convexe si

C est convexe

$x \in C,\ 0 < \alpha \implies \alpha\, x \in C$



Exemples :

- L'ensemble des matrices (semidéfinies) positives

$$\mathcal{S}_n^+ = \{X \in \mathcal{S}_n :\ u^\top X u \geqslant 0 \text{ pour tout } u \in \mathbb{R}^n\}$$

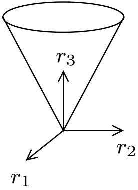
est un cône convexe (fermé)

Rem: ensemble convexe en statistiques : matrices de corrélation

- Cône "du second ordre" dans $\mathbb{R}^3$

$$K_\mu = \left\{ r \in \mathbb{R}^3 :\ \sqrt{r_1^2 + r_2^2} \leqslant \mu\, r_3 \right\}$$

Rem: cône convexe en mécanique



Exemple : optimisation conique

Optimisation linéaire

$$\begin{cases} \min \quad c^\top x \\ \quad Ax = b \\ \quad x \geqslant 0 \end{cases}$$

Optimisation conique (linéaire)

$$\begin{cases} \min \quad c^\top x \\ \quad Ax = b \\ \quad x \in K \end{cases}$$

Optimisation conique

- On garde presque les mêmes propriétés théoriques
- Travail pour adapter/développer des algorithmes
- Formalisme pour de nouvelles applications...

Exemple : optimisation conique

Optimisation linéaire

$$\begin{cases} \min \quad c^\top x \\ \quad Ax = b \\ \quad x \geqslant 0 \end{cases}$$

Optimisation conique (linéaire)

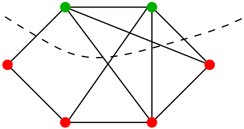
$$\begin{cases} \min \quad c^\top x \\ \quad Ax = b \\ \quad x \in K \end{cases}$$

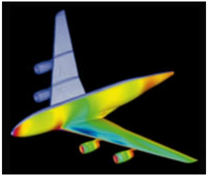
Optimisation conique

- On garde presque les mêmes propriétés théoriques
- Travail pour adapter/développer des algorithmes
- Formalisme pour de nouvelles applications...
- En particulier : optimisation semidéfinie ! (SDP)

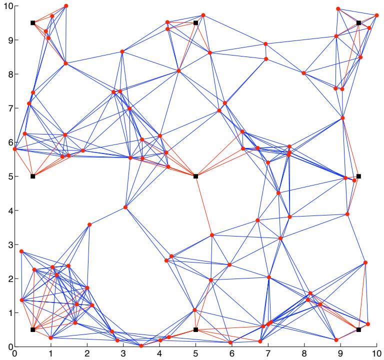
Optimisation semidéfinie

$$\begin{cases} \min \quad \langle C, X \rangle \\ \quad \mathcal{A}X = b \\ \quad X \in \mathcal{S}_n^+ \end{cases}$$

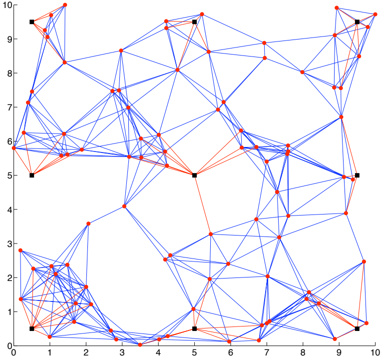




Exemple : localisation de réseaux de capteurs



Exemple : localisation de réseaux de capteurs

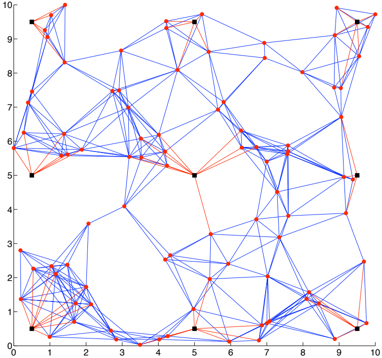


- D matrice "distance euclidienne"
  $\exists\, p_1, \ldots, p_n \in \mathbb{R}^r$

$$D_{ij} = \|p_i - p_j\|^2$$

- Problème : retrouver les p_i connaissant certains D_{ij}...

Exemple : localisation de réseaux de capteurs



- D matrice "distance euclidienne"
  $\exists\, p_1, \ldots, p_n \in \mathbb{R}^r$

$$D_{ij} = \|p_i - p_j\|^2$$

- Problème : retrouver les p_i connaissant certains D_{ij}...

D'où vient la modélisation avec l'optimisation semidéfinie ?

- $P = [p_1, \ldots, p_n]$ et $Y = P^\top P \in \mathcal{S}_n^+$

$$D_{ij} \;=\; p_i^\top p_i + p_j^\top p_j - 2p_j^\top p_i \;=\; Y_{ii} + Y_{jj} - 2Y_{ij} \;=:\; K(Y)$$

- D distance euclidienne $\iff D = K(Y)$ et $Y \in \mathcal{S}_n^+$
- ...

Fonction convexe

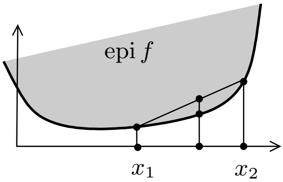
La fonction $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est convexe si

$$x_1, x_2 \in \mathbb{R}^n \quad 0 \leqslant \alpha \leqslant 1 \quad f(\alpha\, x_1 + (1-\alpha)x_2) \leqslant \alpha\, f(x_1) + (1-\alpha) f(x_2)$$

ce qui équivaut à :

$$\operatorname{epi} f = \left\{(x,t) \in \mathbb{R}^{n+1} :\ f(x) \leqslant t\right\}$$

est un ensemble convexe de $\mathbb{R}^{n+1}$



Fonction convexe

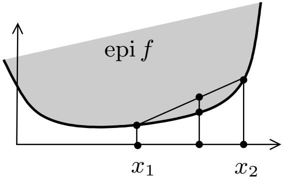
La fonction $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est convexe si

$$x_1, x_2 \in \mathbb{R}^n \quad 0 \leqslant \alpha \leqslant 1 \quad f(\alpha\, x_1 + (1-\alpha) x_2) \leqslant \alpha\, f(x_1) + (1-\alpha) f(x_2)$$

ce qui équivaut à :

$$\operatorname{epi} f = \left\{ (x,t) \in \mathbb{R}^{n+1} : \ f(x) \leqslant t \right\}$$

est un ensemble convexe de $\mathbb{R}^{n+1}$



Exemples :

- Les fonctions affines sont convexes
- Les normes sont convexes (ex: $\|\cdot\|_1$)

Fonction convexe

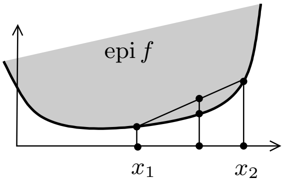
La fonction $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est convexe si

$$x_1, x_2 \in \mathbb{R}^n \quad 0 \leqslant \alpha \leqslant 1 \quad f(\alpha\, x_1 + (1-\alpha)x_2) \leqslant \alpha\, f(x_1) + (1-\alpha) f(x_2)$$

ce qui équivaut à :

$$\operatorname{epi} f = \left\{(x,t) \in \mathbb{R}^{n+1} : \; f(x) \leqslant t\right\}$$

est un ensemble convexe de $\mathbb{R}^{n+1}$

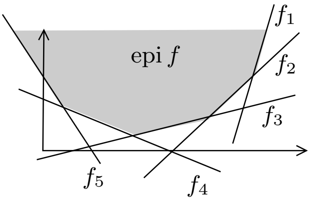


Exemples :

- Les fonctions affines sont convexes
- Les normes sont convexes (ex: $\|\cdot\|_1$)
- Un sup de fonctions convexes est convexe

$$f(x) = \sup_{i \in I} f_i(x) \quad \text{convexe}$$

- Apparition de la non-différentiabilité...



Fonction convexe différentiable

Rappel: La fonction $f\colon \mathbb{R}^n \to \mathbb{R}$ est différentiable en x si

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + o(y - x)$$

Fonction convexe différentiable

Rappel: La fonction $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable en x si

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + o(y - x)$$

Caractérisation des fonctions convexes différentiables :

Pour f différentiable sur U (ouvert convexe), on a l'équivalence entre

- f est convexe sur U
- "le graphe de f est au-dessus de ses tangentes"

$$f(y) \geqslant f(x) + \nabla f(x)^\top (y - x)$$

- $\nabla^2 f(x)$ est semidéfini positif pour tout $x \in U$

Fonction convexe différentiable

Rappel: La fonction $f:\mathbb{R}^n \to \mathbb{R}$ est différentiable en x si

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + o(y-x)$$

Caractérisation des fonctions convexes différentiables :

Pour f différentiable sur U (ouvert convexe), on a l'équivalence entre

- f est convexe sur U
- "le graphe de f est au-dessus de ses tangentes"

$$f(y) \geqslant f(x) + \nabla f(x)^\top (y-x)$$

- $\nabla^2 f(x)$ est semidéfini positif pour tout $x \in U$

Ex : $f(x) = x^\top A\, x + b^\top x$ convexe $\iff A \in \mathcal{S}_n^+$

Fonction convexe différentiable

Rappel: La fonction $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable en x si

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + o(y - x)$$

Caractérisation des fonctions fortement convexes différentiables :

Pour f différentiable sur U (ouvert convexe), on a l'équivalence entre

- f est fortement convexe sur U ($f(x) = g(x) + c\|x\|^2$ avec g convexe)
- "le graphe de f est au-dessus de ses tangentes" (+ du quadratique)

$$f(y) \geqslant f(x) + \nabla f(x)^\top (y - x) + c\|y - x\|^2$$

- $\nabla^2 f(x)$ est uniformément ~~semi~~défini positif pour tout $x \in U$

Ex : $f(x) = x^\top A\, x + b^\top x$ fortement convexe $\iff A \in \mathcal{S}_n^{++}$

Premières conséquences de la convexité en optimisation

$$\left\{\begin{array}{ll} \min & f(x) \\ & x \in C \end{array}\right.$$

$C \subset \mathbb{R}^n$ ensemble convexe

$f : \mathbb{R}^n \to \mathbb{R}$ fonction convexe

- **Globalité :** les minimums locaux sont globaux
  (et l'ensemble des minimums est un convexe)

Premières conséquences de la convexité en optimisation

$$\begin{cases} \min & f(x) \\ & x \in C \end{cases}$$

$C \subset \mathbb{R}^n$ ensemble convexe

$f : \mathbb{R}^n \to \mathbb{R}$ fonction convexe

- **Globalité :** les minimums locaux sont globaux
  (et l'ensemble des minimums est un convexe)
- **Unicité :** si f est strictement convexe ($\Longleftarrow f$ fortement convexe)
  alors il existe au plus un minimum

Premières conséquences de la convexité en optimisation

$$\begin{cases} \min & f(x) \\ & x \in C \end{cases}$$

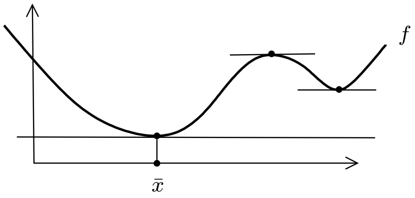
$C \subset \mathbb{R}^n$ ensemble convexe

$f : \mathbb{R}^n \to \mathbb{R}$ fonction convexe

- **Globalité :** les minimums locaux sont globaux
  (et l'ensemble des minimums est un convexe)
- **Unicité :** si f est strictement convexe ($\Longleftarrow$ f fortement convexe)
  alors il existe au plus un minimum
- **Conditions d'optimalité :** nécessaires

  Ex: sans contrainte et f différentiable

$$\bar{x} \text{ sol. de } \min_{x \in \mathbb{R}^n} f(x) \quad \Longrightarrow \quad \nabla f(\bar{x}) = 0$$



Premières conséquences de la convexité en optimisation

$$\begin{cases} \min & f(x) \\ & x \in C \end{cases}$$

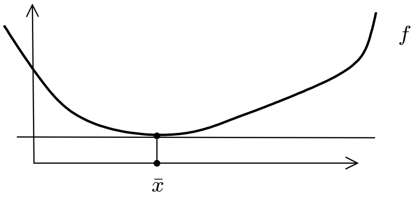
$C \subset \mathbb{R}^n$ ensemble convexe

$f : \mathbb{R}^n \to \mathbb{R}$ fonction convexe

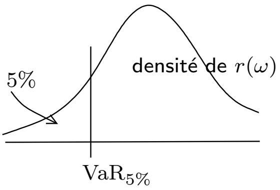
- **Globalité :** les minimums locaux sont globaux
  (et l'ensemble des minimums est un convexe)
- **Unicité :** si f est strictement convexe ($\Longleftarrow f$ fortement convexe)
  alors il existe au plus un minimum
- **Conditions d'optimalité :** nécessaires et suffisantes

  Ex: sans contrainte et f différentiable

$$\bar{x} \text{ sol. de } \min_{x \in \mathbb{R}^n} f(x) \quad \Longleftrightarrow \quad \nabla f(\bar{x}) = 0$$



Exemple : (non)convexité en finance



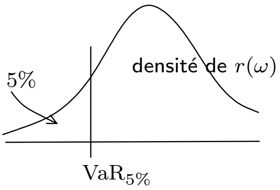
Value-at-risk : mesure de risque pour un placement financier ($\omega \in \mathbb{R}^n$)

- popularisée par JP Morgan (1993)
- institutionalisée par "Bâle II" (2004)

Soit $r(\omega)$ rendement (variable aléatoire)

$$\mathrm{VaR}_{5\%}(\omega) = \max\left\{\alpha : P\big(r(\omega) < \alpha\big) < 5\%\right\}$$

Exemple : (non)convexité en finance



Value-at-risk : mesure de risque pour un placement financier ($\omega \in \mathbb{R}^n$)

- popularisée par JP Morgan (1993)
- institutionalisée par "Bâle II" (2004)

Soit $r(\omega)$ rendement (variable aléatoire)

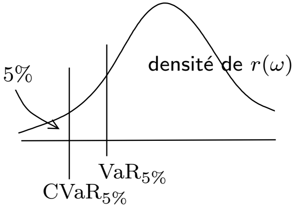
$$\mathrm{VaR}_{5\%}(\omega) = \max\left\{\alpha : P\big(r(\omega) < \alpha\big) < 5\%\right\}$$

Mais "VaR a sous-estimé l'importance des pertes du marché du crédit" (2008)
comportement non-intuitif : il existe $\mathrm{VaR}(\omega_1) = \mathrm{VaR}(\omega_2)$

$$\mathrm{VaR}((\omega_1 + \omega_2)/2) > \mathrm{VaR}(\omega_1) = (\mathrm{VaR}(\omega_1) + \mathrm{VaR}(\omega_2))/2$$

$\longrightarrow$ manque de convexité !...

Exemple : (non)convexité en finance



Value-at-risk : mesure de risque pour un placement financier ($\omega \in \mathbb{R}^n$)

- popularisée par JP Morgan (1993)
- institutionalisée par "Bâle II" (2004)

Soit $r(\omega)$ rendement (variable aléatoire)

$$\mathrm{VaR}_{5\%}(\omega) = \max\left\{\alpha : P\big(r(\omega) < \alpha\big) < 5\%\right\}$$

Mais "VaR a sous-estimé l'importance des pertes du marché du crédit" (2008) comportement non-intuitif : il existe $\mathrm{VaR}(\omega_1) = \mathrm{VaR}(\omega_2)$

$$\mathrm{VaR}((\omega_1 + \omega_2)/2) > \mathrm{VaR}(\omega_1) = (\mathrm{VaR}(\omega_1) + \mathrm{VaR}(\omega_2))/2$$

$\longrightarrow$ manque de convexité !... Solution : Conditional VaR

$$\mathrm{CVaR}_\beta(\omega) = \frac{1}{\beta}\int_0^\beta \mathrm{VaR}_\alpha(\omega)\mathrm{d}\alpha \qquad \text{qui est convexe}$$

$\longrightarrow$ notion de mesure "cohérente" du risque (incluant la convexité)

Conclusion (provisoire) sur la convexité

- La convexité : notion incroyablement simple...
- ...mais qui donne naissance à une géométrie et une analyse riches
- Aussi : porte d'entrée de l'analyse non-différentiable
- Convexité et optimisation
  - apporte des propriétés globales
  - permet de construire des algorithmes
  - théorie de la dualité
- Propriété recherchée...
- "Convexification" de problèmes : relaxation

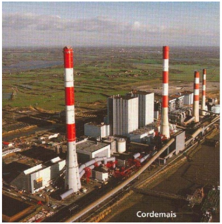
Plan de la présentation

Plan de la présentation

Production électrique en France

- La production d'électricité en France est assurée $n \simeq 200$ centrales

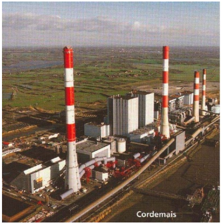
nucléaire 80% pétrole + charbon 3% hydraulique 17%



Production électrique en France

- La production d'électricité en France est assurée $n \simeq 200$ centrales

nucléaire 80% pétrole + charbon 3% hydraulique 17%

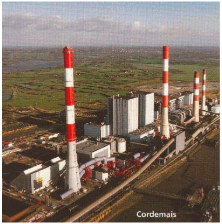



- Question de l'organisation de la production : quelle unité produit quoi et quand ? pour satisfaire la demande à chaque instant ?
- Modélisation comme un problème d'optimisation... dur !
  (unit-commitment)

Production électrique en France

- La production d'électricité en France est assurée $n \simeq 200$ centrales

nucléaire 80% pétrole + charbon 3% hydraulique 17%



- Question de l'organisation de la production : quelle unité produit quoi et quand ? pour satisfaire la demande à chaque instant ?
- Modélisation comme un problème d'optimisation... dur ! (unit-commitment)
- Depuis 2003, EDF optimise sa production d'électricité par un algorithme d'optimisation convexe

Modèle (simplifié) de la planification de production

- Données : centrales : n $[\simeq 200]$ centrales
  intervalle : T $[= 96]$ (2 jours × 48 demi-heures)

- Variables : programme de production pour chaque centrale i

$$p_i = \left(p_i^1, \ldots, p_i^T\right) \in P_i \quad \text{contraintes technologiques}$$

- Objectif : minimiser les coûts de production

- Chaque centrale i a ses coûts $c_i(p_i)$ et ses contraintes $p_i \in P_i$

- Contrainte : satisfaire les demandes (connues) d^t aux temps t

- Problème d'optimisation dur : grande taille, hétérogène, délais serrés

$$\left\{ \begin{array}{lll} \min \quad \sum_i c_i(p_i) & & \text{(somme des coûts)} \\ \quad p_i \in P_i & i = 1, \ldots, n & \text{(contraintes techniques)} \\ \quad \sum_i p_i^t = d^t & t = 1, \ldots, T & \text{(répondre à la demande)} \end{array} \right.$$

"Décomposition par les prix"

- On "dualise" les contraintes couplantes... plus mercredi !
- Variables primales : plannings de production $p \in P = P_1 \times \cdots \times P_n$
  Variables duales : les "prix" $\lambda = (\lambda^1, \ldots, \lambda^T) \in \mathbb{R}^T$
- Problème frère ("dual") : minimiser "la perte" θ
  $\longrightarrow$ problème convexe non-différentiable non-explicite !

"Décomposition par les prix"

- On "dualise" les contraintes couplantes... plus mercredi !
- Variables primales : plannings de production $p \in P = P_1 \times \cdots \times P_n$
  Variables duales : les "prix" $\lambda = (\lambda^1, \ldots, \lambda^T) \in \mathbb{R}^T$
- Problème frère ("dual") : minimiser "la perte" θ
  $\longrightarrow$ problème convexe non-différentiable non-explicite !
- Calcul de la fonction duale à λ fixé $=$ n problèmes indépendants !
- Algorithme efficace d'optimisation convexe (type "faisceaux", cf jeudi)

  $\longrightarrow$ résolution en 1/2h

  recherche en cours...

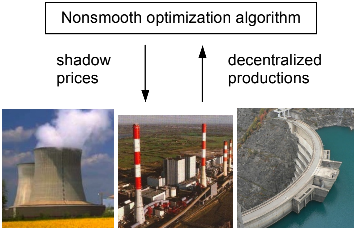
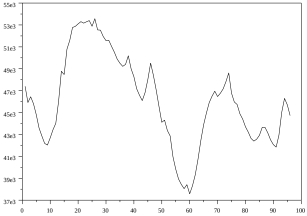


Illustration numérique

Exemple de
demande sur 2 jours

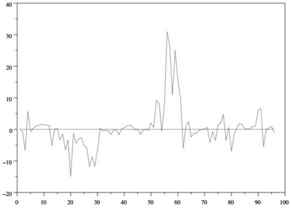
(Ex: de 35000 MW à 70000 MW)



Écart apriori à la demande $\sum_i p_i - d$

(Ex: de -15 MW à 30MW)

à corriger...



Plan de la présentation

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités
- Trois applications de l'optimisation convexe en apprentissage
  1. Filtrage collaboratif
  2. Classification supervisée
  3. Classification supervisée multiclasse

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités
- Trois applications de l'optimisation convexe en apprentissage
  1. Filtrage collaboratif
  2. Classification supervisée
  3. Classification supervisée multiclasse
- Bien d'autres ! (interaction optimisation $\leftrightarrow$ apprentissage)
  - Déjà vu : moindres carrés (parcimonieux)
  - Classification non-supervisée $\rightarrow$ optimisation combinatoire
  - ...

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités
- Trois applications de l'optimisation convexe en apprentissage
  1. Filtrage collaboratif
  2. Classification supervisée
  3. Classification supervisée multiclasse
- Bien d'autres ! (interaction optimisation $\leftrightarrow$ apprentissage)
  - Déjà vu : moindres carrés (parcimonieux)
  - Classification non-supervisée $\rightarrow$ optimisation combinatoire
  - ...

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités
- Trois applications de l'optimisation convexe en apprentissage
  1. Filtrage collaboratif
  2. Classification supervisée
  3. Classification supervisée multiclasse
- Bien d'autres ! (interaction optimisation $\leftrightarrow$ apprentissage)
  - Déjà vu : moindres carrés (parcimonieux)
  - Classification non-supervisée $\rightarrow$ optimisation combinatoire
  - ...

Exemples dans un domaine scientifique : apprentissage

- Apprentissage (...) : construire une représentation à partir d'observation
- Applications : bio-stats, vison par ordinateur (semaine prochaine !),...
- Caractéristiques : problèmes de très grande dimension, bruités
- Trois applications de l'optimisation convexe en apprentissage
  1. Filtrage collaboratif
  2. Classification supervisée
  3. Classification supervisée multiclasse
- Bien d'autres ! (interaction optimisation $\leftrightarrow$ apprentissage)
  - Déjà vu : moindres carrés (parcimonieux)
  - Classification non-supervisée $\rightarrow$ optimisation combinatoire
  - ...

Exemple 3 : classification supervisée multiclasse

- But : classer des objets (auxquels sont associés des descripteurs)
  - données : couples descripteurs/classes $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assigner une classe à un nouvel objet décrit par x ?
- Ex : vision par ordinateur

Exemple 3 : classification supervisée multiclasse

- But : classer des objets (auxquels sont associés des descripteurs)
  - données : couples descripteurs/classes $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assigner une classe à un nouvel objet décrit par x ?
- Ex : vision par ordinateur
- Optimisation : apprendre un classifieur à partir des données
  - calculer une matrice de poids $W \in \mathbb{R}^{p\times K}$
  - minimiser une fonction d'erreur (+ régularisation)

$$\min_{W\in\mathbb{R}^{p\times K}} \; \frac{1}{n}\sum_{i=1}^{n} L(y_i, W^\top x_i) + \alpha\,\mathrm{Reg(W)}$$

  - pour classer suivant $\max_{k=1,\ldots,K} {w_k}^\top x$

Exemple 3 : classification supervisée multiclasse

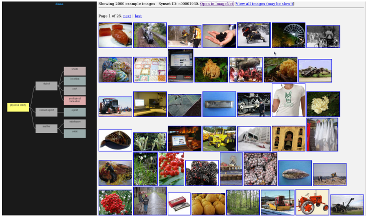
- But : classer des objets (auxquels sont associés des descripteurs)
  - données : couples descripteurs/classes $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assigner une classe à un nouvel objet décrit par x ?
- Ex : vision par ordinateur
- Optimisation : apprendre un classifieur à partir des données
  - calculer une matrice de poids $W \in \mathbb{R}^{p\times K}$
  - minimiser une fonction d'erreur (+ régularisation)

$$\min_{W \in \mathbb{R}^{p\times K}} \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i) + \alpha \,\mathrm{Reg(W)}$$

  - pour classer suivant $\max_{k=1,\ldots,K} {w_k}^\top x$
- Situation : OK pour $K = 2$ (et $K \approx 10$... en revenant au cas 2!)
- Défi : $\boxed{K \text{ grand}}$ (avec une véritable approche multiclasse)
- Ex: Pascal Challenge '10 "imagenet" : $K \approx 1000$ mais peu fiable

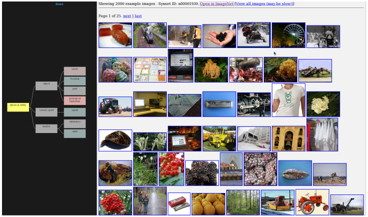
Exemple 3 : classification multiclasse avec sous-structure

- Structure sous-jacente inconnue
- $\mathrm{rang}(W)$ faible
- Factorisation de l'information
  (et calculs moins chers)
- $W = UV^\top,\ U \in \mathbb{R}^{p\times r}, V \in \mathbb{R}^{K\times r}$



Exemple 3 : classification multiclasse avec sous-structure

- Structure sous-jacente inconnue
- $\mathrm{rang}(W)$ faible
- Factorisation de l'information (et calculs moins chers)
- $W = UV^\top,\ U \in \mathbb{R}^{p\times r}, V \in \mathbb{R}^{K\times r}$

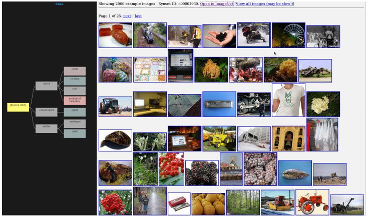


- Optimisation : problème grande taille non-convexe

$$\min_{W\in\mathbb{R}^{p\times K}} \ \alpha\ \mathrm{rang}(W)\ + \frac{1}{n}\sum_{i=1}^{n} L(y_i, W^\top x_i)$$

Exemple 3 : classification multiclasse avec sous-structure

- Structure sous-jacente inconnue
- $\mathrm{rang}(W)$ faible
- Factorisation de l'information
  (et calculs moins chers)
- $W = UV^\top, \ U \in \mathbb{R}^{p \times r}, V \in \mathbb{R}^{K \times r}$



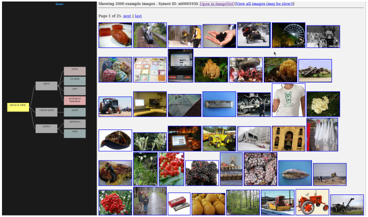
- Optimisation : relaxation convexe non-différentiable

$$\min_{W \in \mathbb{R}^{p \times K}} \ \alpha \, \|\sigma(W)\|_1 \ + \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i)$$

où $\sigma(W)$ est le vecteur des valeurs singulières

Exemple 3 : classification multiclasse avec sous-structure

- Structure sous-jacente inconnue
- $\text{rang}(W)$ faible
- Factorisation de l'information
  (et calculs moins chers)
- $W = UV^\top, \; U \in \mathbb{R}^{p\times r}, V \in \mathbb{R}^{K\times r}$



- Optimisation : relaxation convexe non-différentiable

$$\min_{W\in\mathbb{R}^{p\times K}} \; \alpha \, \|\sigma(W)\|_1 \; + \frac{1}{n}\sum_{i=1}^{n} L(y_i, W^\top x_i)$$

  où $\sigma(W)$ est le vecteur des valeurs singulières

- Recherche en cours : approche constructive + contrôle du rang

Plan d'aujourd'hui (Partie 1) : Introductions, exemples

1. **Vue d'ensemble**
2. **Introduction à l'optimisation**
3. **Introduction à l'analyse convexe**
4. **Exemples de problèmes d'optimisation convexe**
    - Exemple dans l'industrie : gestion de la production électrique
    - Exemples dans un domaine scientifique : apprentissage statistique