

Optimisation Conique - La programmation semi-définie

Amélie Lambert

Cnam

MPRO

- 1 Ensemble convexes
 - Les ensembles convexes
 - Les cônes convexes

- 2 Les matrices semi-définies positives

- 3 Les programmes semi-définis

- 4 Algorithmes de résolution des programmes semi-définis
 - Algorithmes des points intérieurs

Les ensembles convexes

Les ensembles convexes

- Un ensemble C est **convexe** si et seulement si pour tous points x et y dans C , le segment connectant x et y est inclu dans C

Exemples : les cubes, les boules ou les ellipsoïdes

- Pour utiliser les ensembles convexes dans des algorithmes il faut pouvoir les représenter en machine :
 - ▶ description implicite par l'intersection de demi-espaces,
 - ▶ description explicite de la combinaison convexe de ses points extrêmes
- Ici, nous allons parler de programmation conique, nous parlerons de cônes convexes.

Exemples de cônes convexes pour l'optimisation : \mathbb{R}_+^n , cône des matrices semi-définie positives ou le cônes des matrices copositives

Contexte

- On travaille dans un espace euclidien E , de n dimensions muni d'un produit scalaire : ici \mathbb{R}^n
- Pour les vecteurs colonnes $x = (x_1, \dots, x_n)^T$ et $y = (y_1, \dots, y_n)^T$ appartenant à \mathbb{R}^n , le produit scalaire est

$$x^T y = \sum_{i=1}^n x_i y_i$$

Topologie dans les espaces métriques : définitions (1/2)

- On définit la **boule de centre** $x \in \mathbb{R}^n$ et de **rayon** r comme

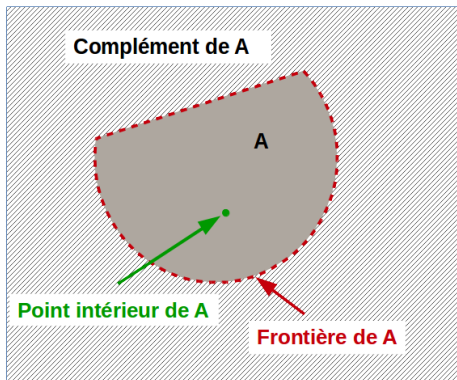
$$B(x, r) = \{y \in \mathbb{R}^n : d(x, y) \leq r\}$$

- Si $A \subset \mathbb{R}^n$, un point $x \in A$ est un **point intérieur** si il existe un rayon $\epsilon > 0$ tel que $B(x, \epsilon) \subseteq A$
- On note $\text{int } A$ l'ensemble des points intérieurs de A , et on dit que A est **ouvert** si tous les points de A sont intérieurs.
- L'ensemble A est **fermé** si son complément $\mathbb{R}^n \setminus A$ est ouvert.

Topologie dans les espaces métriques : définitions (2/2)

- La **fermeture** de A , notée \bar{A} , est l'intersection de tous les fermés contenant A , c'est donc le plus petit fermé contenant A .
- Un point x appartient à la **frontière** de A , notée ∂A , si pour tout $\epsilon > 0$, la boule $B(x, \epsilon)$ a des points dans A et dans son complément $\mathbb{R}^n \setminus A$.
- La **frontière** ∂A est un ensemble fermé et nous avons $\bar{A} = A \cup \partial A$ et $\partial A = \bar{A} \setminus \text{int } A$.
- L'ensemble A est **compact** si et seulement si il est fermé et borné. (i.e. il peut être contenu dans une boule suffisamment grande mais de rayon fini).

Topologie dans les espaces métriques : Illustration



Ensembles convexes : définitions (1/4)

- Le **segment entre les points x et y** est l'ensemble des barycentres à coefficients positifs ou nuls de x et y , il est défini par les points z :

$$[x, y] = \{z = (1 - \alpha)x + \alpha y : \alpha \text{ élément de } [0, 1]\}$$

- Un ensemble C est **convexe** si et seulement si pour tous points x et y dans C , le segment connectant x et y est inclus dans C
- Toute **intersection** d'ensembles convexes est un **ensemble convexe**.

Ensembles convexes : définitions (2/4)

- Soient $n \in \mathbb{N}^*$ et $x_1, \dots, x_n \in C \subseteq \mathbb{R}^n$, une **combinaison convexe** des x_i est un point p de la forme :

$$p = \sum_{i=1}^n \alpha_i x_i \text{ où } \alpha \in \mathbb{R}_+^n \text{ et } \sum_{i=1}^n \alpha_i = 1$$

- Un **sous-ensemble convexe est stable aux combinaisons convexes** :

$$\forall \tilde{n} \in \mathbb{N}, \forall x_1, \dots, x_{\tilde{n}} \in C \subseteq \mathbb{R}^n, \forall \alpha \in \mathbb{R}_+^{\tilde{n}}, \text{ tq } \sum_{i=1}^{\tilde{n}} \alpha_i = 1,$$

$$\text{on a } p = \sum_{i=1}^{\tilde{n}} \alpha_i x_i \in C$$

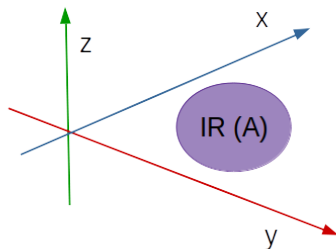
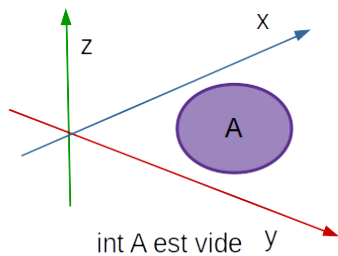
Ensembles convexes : définitions (3/4)

Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe. Un point $x \in C$ est dans l'**intérieur relatif** de C si et seulement si

$$\forall y \in C, \exists z \in C, \alpha \in]0, 1[: x = \alpha y + (1 - \alpha)z$$

En d'autres termes, c'est l'intérieur de C relativement au sous-espace affine engendré par C .

Exemple : Soit $A = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0, x^2 + y^2 \leq 1\}$
 $\forall x \in A, \nexists \epsilon > 0$ tq $B(x, \epsilon) \subseteq A$, (car $z = 0$) $\Rightarrow \text{int } A = \emptyset$



Ensembles convexes : définitions (4/4)

- L'**enveloppe convexe** de $A \in \mathbb{R}^n$ (notée $\text{conv } A$) est le plus petit ensemble convexe contenant A .
- C'est l'ensemble des combinaisons convexes des points $x_1, \dots, x_{\tilde{n}}$, $\forall \tilde{n} \in \mathbb{N}$, c'est-à-dire de leurs barycentres à coefficients positifs :

$$\text{conv } A = \left\{ p = \sum_{i=1}^{\tilde{n}} \alpha_i x_i : \tilde{n} \in \mathbb{N}, \forall x_1, \dots, x_{\tilde{n}} \in C, \forall \alpha \in \mathbb{R}_+^{\tilde{n}}, \text{ tq } \sum_{i=1}^{\tilde{n}} \alpha_i = 1 \right\}$$

- L'**enveloppe convexe** d'un ensemble fini de points est appelé **polytope** (ou **polygone** en 2 dimensions).

Description implicite : intersection de demi-espaces (1/2)

- Un **hyperplan H** correspondant au vecteur normal $c \in \mathbb{R}^n \setminus \{0\}$ passant par $x \in \mathbb{R}^n$ est :

$$H = \{x \in \mathbb{R}^n : c^T x = \beta\}$$

c'est un **sous-espace affine** de dimension $n - 1$.

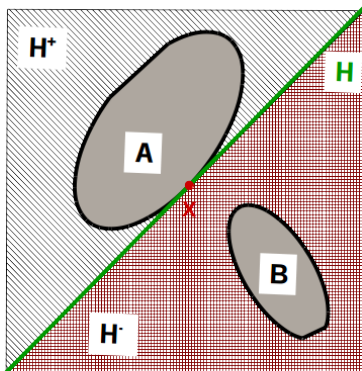
- L'hyperplan H divise \mathbb{R}^n en 2 **demi-espaces fermés** :

$$H^+ = \{y \in \mathbb{R}^n : c^T y \geq c^T x = \beta\} \text{ et } H^- = \{y \in \mathbb{R}^n : c^T y \leq c^T x = \beta\}$$

- H sépare $A \subseteq \mathbb{R}^n$ et $B \subseteq \mathbb{R}^n$ si $A \subseteq H^+$ et $B \subseteq H^-$ (ou inversement).
i.e. si $\exists c \in \mathbb{R}^n \setminus \{0\}, \beta \in \mathbb{R} : \forall x \in A, y \in B : c^T x \leq \beta \leq c^T y$
- Si les deux inégalités sont strictes, la séparation est dite **stricte**.

Description implicite : intersection de demi-espaces (2/2)

- H est le **support de A en x** , si $x \in A$, $x \in H$ et A est contenu dans un des 2 sous-espaces H^+ ou H^- .
- Alors on dit que H^+ (ou H^-) est le **demi-espace support**.



Exemple : L'hyperplan H supporte A en x et sépare A et B

Projection orthogonale

Soit $C \subset \mathbb{R}^n$ un espace convexe et fermé. Il est possible de projeter tous les points x de \mathbb{R}^n sur C : on prend le point de C qui est le plus proche de x .

- Soit $x \in \mathbb{R}^n \setminus C$ (x est extérieur à C), $\exists! \pi_C(x) \in C$, qui est le plus proche de x . De plus, $\pi_C(x) \in \partial C$.
- La **projection** $\pi_C : \mathbb{R}^n \rightarrow C$ est caractérisée par la propriété

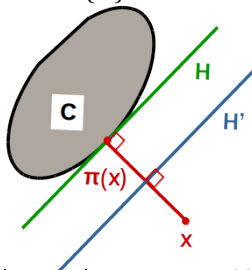
$$\forall y \in C : d(\pi_C(x), x) \leq d(y, x)$$

- Pour tout $y \in \partial C$, il y a un point x extérieur à C , tel que $y = \pi_C(x)$

Hyperplans supports et séparateurs

Soient $C \subset \mathbb{R}^n$ un espace convexe et fermé, $x \in \mathbb{R}^n \setminus C$, $\pi_C(x) \in C$ la projection de x sur C , on a :

- 1 l'hyperplan H qui passe par x avec comme normale $x - \pi_C(x)$ supporte C en $\pi_C(x)$ et donc sépare $\{x\}$ et C .
- 2 l'hyperplan H' qui passe par $(x + \pi_C(x))/2$ avec comme normale $x - \pi_C(x)$ sépare strictement $\{x\}$ et C .



Donc on peut construire un hyperplan support H sur chaque point $x \in \partial C$.
 \Rightarrow description implicite de l'ensemble convexe.

Description explicite : ensemble des points extrêmes (1/2)

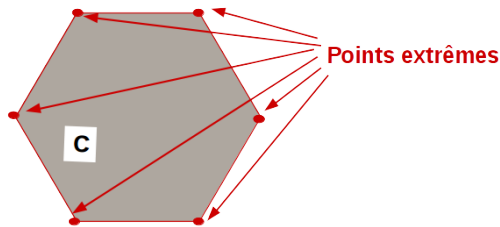
Définition : un point $x \in C$ convexe est un **point extrême** si $C \setminus \{x\}$ est encore convexe (i.e. tout segment s de C , x n'est pas un point de l'intérieur relatif de s .)

i.e. : si on ne peut pas écrire x sous la forme $x = (1 - \alpha)y + \alpha z$, avec $y, z \in C$ et $0 < \alpha < 1$.

On note **ext** C l'ensemble de tous les points extrêmes de C .

Propriété : Soit $C \subset \mathbb{R}^n$ un ensemble compact et convexe, alors

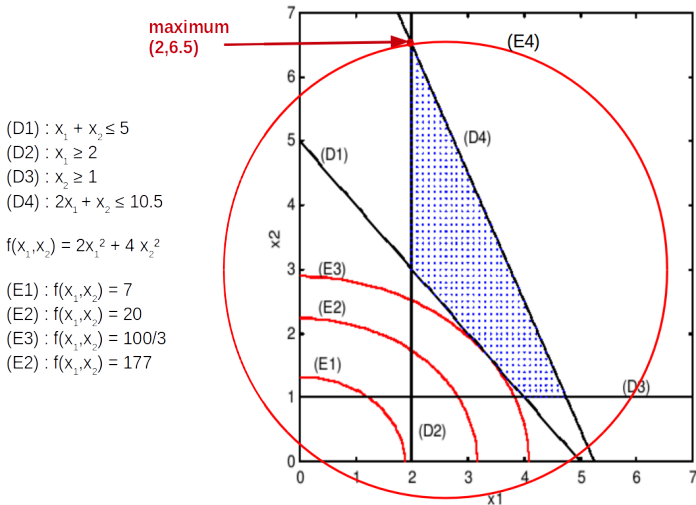
$$C = \text{conv}(\text{ext } C)$$



Les points extrêmes d'un ensemble convexe.

Description explicite : ensemble des points extrêmes (2/2)

Application : Soit $C \subset \mathbb{R}^n$ compact et convexe, et $f : \mathbb{R}^n \mapsto \mathbb{R}$ continue et convexe, alors $\max_{x \in C} f(x)$ a un maximum qui est un point extrême de C .



Les cônes convexes

Les cônes convexes (1/4)

- Un ensemble non vide $K \subset \mathbb{R}^n$ est un **cône convexe** si et seulement si il est stable par rapport aux combinaisons linéaires positives :

$$\forall \alpha, \beta \in \mathbb{R}_+, \forall x, y \in K : \alpha x + \beta y \in K$$

- Un cône convexe est **pointé** s'il ne contient pas de droite :

$$x \in K \cup -x \in K \Rightarrow x = 0$$

- Le **dual** K^* d'un cône convexe K est défini de la façon suivante :

$$K^* = \{y \in \mathbb{R}^n : x^T y \geq 0, \forall x \in K\}$$

K^* est un cône fermé et convexe.

Les cônes convexes (2/4)

Un **cône convexe, fermé, pointé avec un intérieur non vide** $K \subset \mathbb{R}^n$ définit un ordre partiel sur \mathbb{R}^n , pour $x, y \in \mathbb{R}^n$:

$$x \succeq y \iff x - y \in K$$

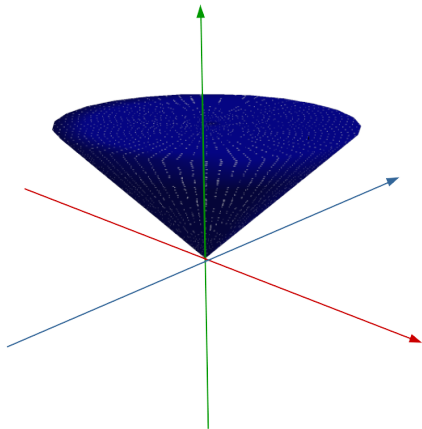
- Cet ordre partiel satisfait les conditions suivantes :

- 1 Reflexivité : $\forall x \in \mathbb{R}^n : x \succeq x$
- 2 Anti-symmetrie : $\forall x, y \in \mathbb{R}^n : x \succeq y, y \succeq x \implies x = y$
- 3 Transitivité : $\forall x, y, z \in \mathbb{R}^n : x \succeq y, y \succeq z \implies x \succeq z$
- 4 Homogénéité : $\forall x, y \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}_+ : x \succeq y, \implies \alpha x \succeq \alpha y$
- 5 Additivité : $\forall x, y, x', y' \in \mathbb{R}^n : x \succeq y, x' \succeq y' \implies x + x' \succeq y + y'$

- On définit l'inégalité stricte par :

$$\forall x, y \in \mathbb{R}^n : x \succ y \iff x - y \in \text{int } K$$

Les cônes convexes (3/4)

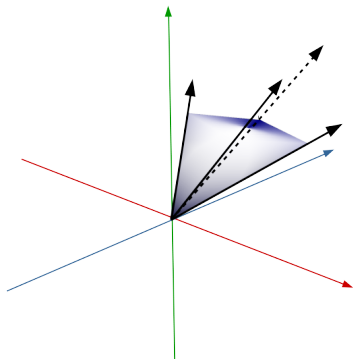


Exemple de cône qui n'est pas une combinaison conique d'un nombre fini de vecteurs.

Les cônes convexes (4/4)

Le cône convexe généré par un ensemble de vecteurs $A \subset \mathbb{R}^n$ est le plus petit cône convexe contenant A :

$$\text{cône } A = \left\{ p = \sum_{i=1}^n \alpha_i x_i : n \in \mathbb{N}, x \in A, \alpha \in \mathbb{R}_+^n \right\}$$

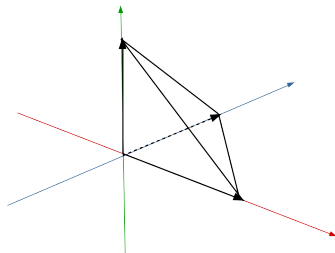


Exemple de cône généré par la combinaison conique des 4 vecteurs noirs

Exemple : \mathbb{R}_+^n

\mathbb{R}_+^n est un cône pointé et convexe, il est défini par :

$$\mathbb{R}_+^n = \{x = (x_1, \dots, x_n)^T \in \mathbb{R}^n : x_1, \dots, x_n \geq 0\}$$



Pour le PL suivant :

$$(PL) \begin{cases} \max c^T x \\ \text{s.t.} \\ Ax \succeq b \end{cases}$$

l'ordre partiel $x \succeq y$ signifie $x_i \geq y_i$ pour tout $i \in \{1, \dots, n\}$

Le cône \mathbb{R}_+^3 : un polyèdre.

- 1 Ensemble convexes
 - Les ensembles convexes
 - Les cônes convexes
- 2 Les matrices semi-définies positives
- 3 Les programmes semi-définis
- 4 Algorithmes de résolution des programmes semi-définis
 - Algorithmes des points intérieurs

Théorème de décomposition spectrale

Toute matrice $X \in \mathcal{S}^n$ (ensemble des matrices symétriques de dimension n) peut être décomposée comme :

$$X = \sum_{i=1}^n \lambda_i u_i u_i^T$$

où les $\lambda_i \in \mathbb{R}$ sont les **valeurs propres** de X , et les $u_i \in \mathbb{R}^n$ les **vecteurs propres** associés, formant une base orthonormale de \mathbb{R}^n .

Matriciellement, $X = PDP^T$, où $D = \text{diag}(\lambda)$ est la matrice diagonale dont les éléments non-nuls sont les λ_i , et P la matrice orthogonale dont les colonnes sont les u_i .

Sous-matrices principales et symétriques, mineurs

- **Définition** : Soit $X \in \mathcal{S}^n$, on appelle **sous-matrice symétrique** de X toute sous matrice de X obtenue en supprimant des lignes et des colonnes de X en même nombre et de mêmes indices.

$$\text{Ex : } X = \begin{pmatrix} 3 & 5 & 6 \\ 5 & 1 & 9 \\ 6 & 9 & 2 \end{pmatrix}, \text{ ordre 2 : } \begin{pmatrix} 3 & 5 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} 3 & 6 \\ 6 & 2 \end{pmatrix} \begin{pmatrix} 1 & 9 \\ 9 & 2 \end{pmatrix},$$

ordre 1 : (3) (1) (2)

- **Définition** : Soit $X \in \mathcal{S}^n$, on appelle **sous-matrice principale** de X toute sous matrice symétrique de X obtenue en supprimant les dernières lignes et les dernières colonnes de X .

$$\text{Ex : } X = \begin{pmatrix} 3 & 5 & 6 \\ 5 & 1 & 9 \\ 6 & 9 & 2 \end{pmatrix}, \text{ ordre 2 : } \begin{pmatrix} 3 & 5 \\ 5 & 1 \end{pmatrix}, \text{ ordre 1 : (3)}$$

- **Définition** : On appelle **mineur symétrique (ou principal)**, le déterminant d'une sous-matrice symétrique (ou principale).

Matrices semi-définies positives

Soit une matrice $X \in \mathcal{S}^n$, les assertions suivantes sont équivalentes :

- ① X est **semi-définie positive (SDP)**, noté $X \succeq 0$, si

$$\forall x \in \mathbb{R}^n, x^T X x = \sum_{i=1}^n \sum_{j=1}^n X_{ij} x_i x_j = \langle X, x x^T \rangle \geq 0$$

- ② La plus petite valeur propre de X est positive ou nulle.
- ③ Tous les mineurs symétriques de X sont positifs ou nuls.
- ④ Décomposition de Cholesky : Pour $k \geq 1$, il existe une matrice triangulaire inférieure $L \in \mathbb{R}^{n \times k}$ tel que $X = L L^T$

Matrices semi-définies positives : exemple

$$X = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

2 Les valeurs propres de X sont $(0, 4, 3)$.

3 Mineurs symétriques :

▶ Mineur symétrique d'ordre 3 :

$$\begin{vmatrix} 3 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{vmatrix} = 3 \begin{vmatrix} 3 & 1 \\ 1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 0 \geq 0$$

▶ Mineurs symétriques d'ordre 2 :

$$\begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} = 8 \geq 0, \quad \begin{vmatrix} 3 & -1 \\ -1 & 1 \end{vmatrix} = 2 \geq 0, \quad \begin{vmatrix} 3 & 1 \\ 1 & 1 \end{vmatrix} = 2 \geq 0$$

▶ Mineurs symétriques d'ordre 1 : $|3| = 3 \geq 0$, $|3| = 3 \geq 0$, $|1| = 1 \geq 0$.

\implies La matrice X est semi-définie positive.

Matrices semi-définies positives : exemple

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 8 \end{pmatrix}$$

④ Décompositions de Choleski non unique :

$$\blacktriangleright \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 8 \end{pmatrix} = LL^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & \frac{4}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{2} & \frac{4}{\sqrt{2}} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\blacktriangleright \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 8 \end{pmatrix} = L'L'^T = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

\implies La matrice X est semi-définie positive.

Matrices définies positives

Soit une matrice $X \in \mathcal{S}^n$, les assertions suivantes sont équivalentes :

- ① X est **définie positive (DP)**, noté $X \succ 0$, si

$$\forall x \in \mathbb{R}^n, x^T X x = \sum_{i=1}^n \sum_{j=1}^n X_{ij} x_i x_j = \langle X, x x^T \rangle > 0$$

- ② La plus petite valeur propre de X est strictement positive.
- ③ Tous les mineurs principaux de X sont strictement positifs.
- ④ Décomposition de Cholesky : pour $k \geq 1$, il existe une unique matrice triangulaire inférieure $L \in \mathbb{R}^{n \times k}$ inversible tel que $X = LL^T$.

Matrices définies positives : exemple

$$X = \begin{pmatrix} 4 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

② Les valeurs propres de X sont (0.13, 3.2, 4.7).

③ Mineurs symétriques :

▶ Mineur symétrique d'ordre 3 :

$$\begin{vmatrix} 4 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{vmatrix} = 4 \begin{vmatrix} 3 & 1 \\ 1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 2 > 0$$

▶ Mineur principal d'ordre 2 :

$$\begin{vmatrix} 4 & 1 \\ 1 & 3 \end{vmatrix} = 11 > 0$$

▶ Mineurs principaux d'ordre 1 : $|4| = 4 > 0$

\implies La matrice X est définie positive.

Matrices définies positives : exemple

$$X = \begin{pmatrix} 4 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

④ Décomposition de Choleski unique :

$$\begin{pmatrix} 4 & 1 & -1 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix} = LL^T = \begin{pmatrix} 2 & 0 & 0 \\ \frac{1}{2} & \frac{\sqrt{11}}{2} & 0 \\ -\frac{1}{2} & \frac{5}{2\sqrt{11}} & \frac{\sqrt{2}}{\sqrt{11}} \end{pmatrix} \begin{pmatrix} 2 & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{11}}{2} & \frac{5}{2\sqrt{11}} \\ 0 & 0 & \frac{\sqrt{2}}{\sqrt{11}} \end{pmatrix}$$

\Rightarrow La matrice X est définie positive.

Produit tensoriel de deux vecteurs

- **Définition** : Soient les vecteurs $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, on définit le produit tensoriel de deux vecteurs :

$$u \otimes v = uv^T = \begin{pmatrix} u_1 v_1 & \dots & u_1 v_n \\ \vdots & \ddots & \vdots \\ u_n v_1 & \dots & u_n v_n \end{pmatrix}$$

- **Propriété** : uv^T n'est pas symétrique, et $u \otimes v$ n'est pas commutatif.

Exemple : Soient $u = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ et $v = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$, on a :

$$u \otimes v = \begin{pmatrix} -2 & 0 & 4 \\ -1 & 0 & 2 \\ -3 & 0 & 6 \end{pmatrix} \text{ et } v \otimes u = \begin{pmatrix} -2 & -1 & -3 \\ 0 & 0 & 0 \\ 4 & 2 & 6 \end{pmatrix}$$

Remarques

- 1 Si X est SDP et $X_{ii} = 0$, alors $X_{ij} = X_{ji} = 0, \forall j = 1, \dots, n$
- 2 Si X est une matrice diagonale, elle est SDP, si tous ces éléments diagonaux sont positifs ou nuls (ses valeurs propres)
- 3 Si de plus ils sont tous strictement positifs, alors elle est DP.

4 Soit $x \in \mathbb{R}^n$, $x \otimes x = xx^T = \begin{pmatrix} x_1^2 & \dots & x_1 x_n \\ \vdots & \ddots & \vdots \\ x_n x_1 & \dots & x_n^2 \end{pmatrix}$ est SDP.

Le produit scalaire et la trace

- La **trace d'une matrice** $X \in \mathbb{R}^{n \times n}$ est $\text{Tr}(X) = \sum_{i=1}^n X_{ii} = \sum_{i=1}^n \lambda_i$
- **Propriétés**
 - ▶ $\text{Tr}(\lambda X) = \lambda \text{Tr}(X)$
 - ▶ $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$.
 - ▶ $\text{Tr}(A) = \text{Tr}(A^T)$
 - ▶ $\text{Tr}(AB) = \text{Tr}(BA)$
- On définit le **produit scalaire** : $\langle A, B \rangle = \text{Tr}(A^T B) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij}$

Matrices par blocs

- **Définition** : Soient les matrices $X_i \in \mathcal{S}^{n_i}, \forall i = 1, \dots, r$,
 $X = X_1 \oplus \dots \oplus X_r$ est la **matrice par blocs** suivante :

$$X = X_1 \oplus \dots \oplus X_r = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & X_r \end{pmatrix}$$

- **Propriété** : $X \succeq 0$ si et seulement si $X_i \succeq 0, \forall i = 1, \dots, r$

Le cône des matrices semi-définie positives

- On note $\mathcal{S}_+^n \subset \mathcal{S}^n$ le cône convexe des matrices SDP $\mathcal{S}_+^n \subset \mathcal{S}^n$:

$$\mathcal{S}_+^n = \{X \in \mathcal{S}^n : X \text{ est sdp, i.e } \forall x \in \mathbb{R}^n : x^T X x \geq 0\}$$

- C'est bien un cône convexe puisque :

$\forall \alpha, \beta > 0$ et $(X, Y) \in (\mathcal{S}_+^n)^2$, on a

$$x^T (\alpha X + \beta Y) x = \alpha x^T X x + \beta x^T Y x \geq 0, \text{ et donc } \alpha X + \beta Y \in \mathcal{S}_+^n.$$

- Soit $X \in \mathcal{S}_+^n$, $X = \sum_{i=1}^n \lambda_i u_i u_i^T$ avec $\lambda_i \geq 0$, le cône des matrices SDP est généré par les matrices de rang 1 :

$$\mathcal{S}_+^n = \text{cône} \{xx^T : x \in \mathbb{R}^n\}$$

- A l'intérieur du cône \mathcal{S}_+^n se trouvent les matrices définies positives.

Le dual du cône des matrices SDP

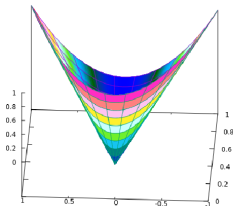
- Le dual de S_+^n est :

$$(S_+^n)^* = \left\{ B \in S^n : \langle A, B \rangle \geq 0, \forall A \in S_+^n \right\}$$

- **Théorème** : Le cône S_+^n coïncide avec son dual : $(S_+^n)^* = S_+^n$,
i.e. $A \succeq 0 \iff \forall B \in S_+^n : \langle A, B \rangle \geq 0$

Exercice : preuve

Exemple : Le cône des matrices \mathcal{S}_+^2



$$\begin{pmatrix} x & y \\ y & z \end{pmatrix} \succeq 0 \\ \iff x \geq 0, z \geq 0, xz - y^2 \geq 0$$

Le cône des matrices $X \in \mathcal{S}_+^2$

- 1 Ensemble convexes
 - Les ensembles convexes
 - Les cônes convexes

- 2 Les matrices semi-définies positives

- 3 Les programmes semi-définis

- 4 Algorithmes de résolution des programmes semi-définis
 - Algorithmes des points intérieurs

La programmation semi-définie

- La **programmation semi-définie (SDP)** est une extension de la programmation linéaire (LP).
- La SDP a été étudiée pour la première fois dans les années 60.
- Un problème SDP est un programme dont l'objectif est linéaire en les termes de la matrice variable X et dont les contraintes sont également linéaires en les termes de cette matrice.

Inégalités matricielles linéaires (LMI)

Définition : Soient $A_i \in \mathcal{S}^n$ et $x \in \mathbb{R}^n$, une **inegalité matricielle linéaire**

est de la forme : $g(x) = \sum_{i=1}^m x_i A_i \succeq A_0$

Exemple : Soient 2 réels x_1 et $x_2 \in \mathbb{R}$, et l'inégalité matricielle suivante :

$$x_1 \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \succeq \begin{pmatrix} -1 & 0 & -1 \\ 0 & -2 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

est équivalent à : $\begin{pmatrix} x_2 + 1 & x_1 & 1 \\ x_1 & 2 & 0 \\ 1 & 0 & x_2 \end{pmatrix} \succeq 0$

ou bien encore à l'infinité d'inégalités : $\forall v \in \mathbb{R}^3$,

$$v^T \begin{pmatrix} x_2 + 1 & x_1 & 1 \\ x_1 & 2 & 0 \\ 1 & 0 & x_2 \end{pmatrix} v \geq 0$$

Inégalités matricielles linéaires (LMI)

$$g(x) = \sum_{i=1}^m x_i A_i - A_0 \succeq 0$$

$g(x)$ caractérise un ensemble convexe, qui est l'ensemble \mathcal{C} :

$$\mathcal{C} = \{x \in \mathbb{R}^n : g(x) \succeq 0\}$$

\mathcal{C} est dans l'intersection d'un espace affine et du cône des matrices semi-définies positives.

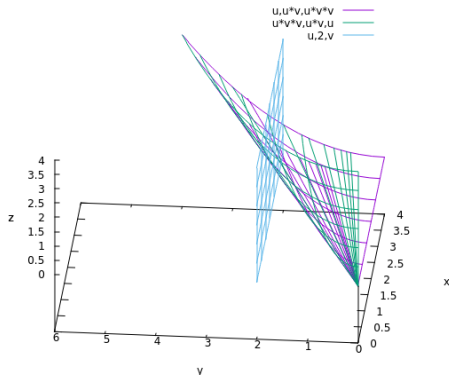
\mathcal{C} n'est en général pas un polytope pour $n \geq 2$

Inégalités matricielles linéaires (LMI) : exemple

Dans \mathcal{S}_+^2 , on cherche les matrices $\begin{pmatrix} x & y \\ y & z \end{pmatrix} \succeq 0$ telles que $y = 2$, i.e. :

$$\begin{pmatrix} x & 2 \\ 2 & z \end{pmatrix} \succeq 0 \iff x \geq 0, z \geq 0, xz \geq 4$$

Ce sont celles qui sont à l'intersection du cône \mathcal{S}_+^2 et de l'hyperplan $y = 2$.



Le problème primal semi-défini

Soient $Q, A_1, \dots, A_m \in \mathcal{S}^n$, et $b \in \mathbb{R}^m$, le primal (SDP) s'écrit :

$$(SDP) \begin{cases} \max f(\mathbf{X}) = \langle Q, \mathbf{X} \rangle \\ \text{s.t.} \\ g_r(\mathbf{X}) = \langle A_r, \mathbf{X} \rangle = b_r \quad \forall r = 1, \dots, m \\ \mathbf{X} \succeq 0 \end{cases}$$

Les contraintes $g_r(x)$ indiquent la structure de la matrice \mathbf{X} .

rappel : $\forall A, B \in \mathcal{S}^n, \langle A, B \rangle = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij}$

Exemple : un problème primal SDP (1/2)

$$(E_{X_1}) \begin{cases} \min \langle \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, X \rangle \\ \text{s.t.} \\ \langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \rangle = 2 \\ X \succeq 0 \end{cases} \quad (E_{X_1}) \begin{cases} \min X_{11} + X_{22} \\ \text{s.t.} \\ X_{12} + X_{21} = 2 \\ X \succeq 0 \end{cases}$$

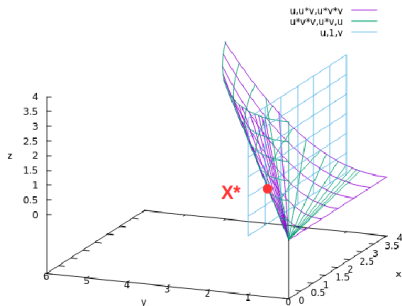
Ce qui est équivalent au problème :

$$(E_{X_1}) \begin{cases} \min X_{11} + X_{22} \\ \text{s.t.} \\ \begin{pmatrix} X_{11} & 1 \\ 1 & X_{22} \end{pmatrix} \succeq 0 \end{cases}$$

Exemple : un problème primal SDP (2/2)

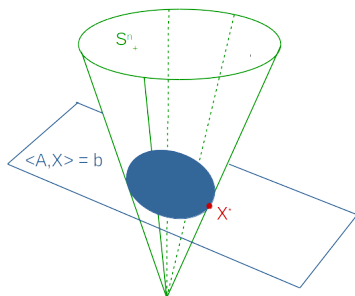
$$(E_{X_1}) \begin{cases} \min X_{11} + X_{22} \\ \text{s.t.} \\ \begin{pmatrix} X_{11} & 1 \\ 1 & X_{22} \end{pmatrix} \succeq 0 \end{cases}$$

La solution optimale est la matrice $X^* = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ de valeur 2.



Le problème primal SDP : géométrie

La matrice optimale X^* se trouve à l'intersection de l'hyperplan $\langle A, X \rangle = b$ et du cône \mathcal{S}_+^n des matrices SDP



Exemple de région admissible pour un programme SDP

Problème primal non nécessairement borné

- Le problème SDP peut ne pas être borné.
- **Exemple** : Soient un scalaire $\alpha \in \mathbb{R}$, et la famille de problèmes (E_{X^α})

$$(E_{X^\alpha}) \begin{cases} \max X_{11} \\ \text{s.t.} \\ \begin{pmatrix} X_{11} & \alpha \\ \alpha & 0 \end{pmatrix} \succeq 0 \end{cases}$$

- ▶ $\det(X) = -\alpha^2$ et donc la région admissible de (E_{X^α}) est vide si $\alpha \neq 0$.
- ▶ Lorsque $\alpha = 0$, (E_{X^α}) admet une solution admissible, mais pas de solution strictement admissible.

La valeur optimale est $+\infty$, (E_{X^α}) est non borné.

- En cas de minimisation de la fonction objectif, la valeur optimale est 0 avec la solution $X = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Optimum non nécessairement atteint

- La valeur optimale d'un problème SDP peut ne pas être atteinte
- Exemple

$$(SDP) \begin{cases} \min X_{11} \\ \text{s.t.} \\ \begin{pmatrix} X_{11} & 1 \\ 1 & X_{22} \end{pmatrix} \succeq 0 \end{cases}$$

La valeur optimale est 0 dont la solution est $X_{11} = \frac{1}{X_{22}}$, avec $X_{22} \rightarrow +\infty$, et elle n'est donc jamais atteinte.

Programmation SDP et Programmation Linéaire

Considérons un programme SDP où :

- $Q, A_1, \dots, A_m \in \mathcal{S}^n$ sont des matrices diagonales, dont les éléments diagonaux sont les vecteurs q et a_r ($r = 1, \dots, m$), $b \in \mathbb{R}^m$,
- la matrice variable X est diagonale, et ses termes diagonaux sont les éléments du vecteur x

Le problème SDP s'écrit

$$(SDP_d) \begin{cases} \max \langle q, x \rangle \\ \text{s.t.} \\ \langle a_r, x \rangle = b_r \\ x \succeq 0 \end{cases} \iff (PL) \begin{cases} \max q^T x \\ \text{s.t.} \\ a_r^T x = b_r \\ x \geq 0 \end{cases}$$

La programmation SDP inclut donc la programmation linéaire.

Le problème dual SDP

Soit $\beta \in \mathbb{R}^m$ les variables duales de (SDP) , le dual $(DSDP)$ est :

$$(SDP) \begin{cases} \max f(\mathbf{X}) = \langle \mathbf{Q}, \mathbf{X} \rangle \\ \text{s.t.} \\ g_r(\mathbf{X}) = \langle \mathbf{A}_r, \mathbf{X} \rangle = b_r \leftarrow \beta_r \\ \mathbf{X} \succeq 0 \end{cases} \quad (DSDP) \begin{cases} \min h(\beta) = \langle \mathbf{b}, \beta \rangle \\ \text{s.t.} \\ \sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{S} = \mathbf{Q} \\ \beta \in \mathbb{R}^m \\ \mathbf{S} \succeq 0 \end{cases}$$

Remarques :

- $\sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}$ doit appartenir à \mathcal{S}_+^n
- la matrice \mathbf{S} joue ici le rôle d'une variable d'écart.

Exemple : le problème dual SDP

$$(E_{X_1}) \begin{cases} \min X_{11} + X_{22} \\ \text{s.t.} \\ \begin{pmatrix} X_{11} & 1 \\ 1 & X_{22} \end{pmatrix} \succeq 0 \end{cases}$$

$$(E_{X_1}) \begin{cases} \max -\langle \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, X \rangle \\ \text{s.t.} \\ \langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \rangle = 2 \leftarrow \beta \\ X \succeq 0 \end{cases}$$

$$(DE_{X_1}) \begin{cases} \min 2\beta \\ \text{s.t.} \\ \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \succeq 0 \\ \beta \in \mathbb{R}^2 \end{cases}$$

Dual d'un SDP, relation avec le Lagrangien

$$(SDP) : \max_{\mathbf{X}} \{ \langle \mathbf{Q}, \mathbf{X} \rangle : \langle \mathbf{A}_r, \mathbf{X} \rangle = b_r, r = 1, \dots, m, \mathbf{X} \succeq 0 \}$$

On introduit pour chaque contrainte β_r , son Lagrangien est :

$$\mathcal{L}(\mathbf{X}, \beta) = \langle \mathbf{Q}, \mathbf{X} \rangle - \sum_{r=1}^m \beta_r (\langle \mathbf{A}_r, \mathbf{X} \rangle - b_r) = \langle \mathbf{b}, \beta \rangle - \langle \sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}, \mathbf{X} \rangle$$

dont la fonction duale est : $L(\beta) = \max_{\mathbf{X} \succeq 0} \langle \mathbf{b}, \beta \rangle - \langle \sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}, \mathbf{X} \rangle$

et son dual Lagrangien est : $(DL) = \min_{\beta \in \mathbb{R}^m} \max_{\mathbf{X} \succeq 0} \langle \mathbf{b}, \beta \rangle - \langle \sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}, \mathbf{X} \rangle$

$$\text{Or, } \max_{\mathbf{X} \succeq 0} \mathcal{L}(\mathbf{X}, \beta) = \begin{cases} \langle \mathbf{b}, \beta \rangle & \text{ssi } (\sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}) \succeq 0 \\ \infty & \text{sinon} \end{cases}$$

Donc : $(DL) = \min_{\beta \in \mathbb{R}^m} \{ \langle \mathbf{b}, \beta \rangle : (\sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{Q}) \succeq 0 \}$ qui est exactement $(DSDP)$

Admissibilité et admissibilité stricte

- Admissibilité :

- ▶ **primale** : Une matrice $X \succeq 0$ est réalisable pour (*SDP*) ssi :

$$\langle A_r, X \rangle = b_r \quad r = 1, \dots, m.$$

- ▶ **duale** : Le vecteur $\beta \in \mathbb{R}^m$ est réalisable pour (*DSDP*) ssi

$$\sum_{r=1}^m \beta_r A_r - Q \succeq 0.$$

- Admissibilité stricte :

- ▶ **primale** : Une matrice $X \succ 0$ est strictement réalisable pour (*SDP*) ssi

$$\langle A_r, X \rangle = b_r \quad r = 1, \dots, m.$$

- ▶ **duale** : Le vecteur $\beta \in \mathbb{R}^m$ est strictement réalisable pour (*DSDP*) ssi

$$\sum_{r=1}^m \beta_r A_r - Q \succ 0.$$

Saut de dualité - Dualité faible et forte

Etant donné X et β solutions réalisables de (SDP) et $(DSDP)$, la différence entre les valeurs des objectifs de (SDP) et $(DSDP)$ est appelée le **saut de dualité**. Il est donné par :

$$\langle b, \beta \rangle - \langle Q, X \rangle$$

Dualité faible : Etant donné X et β solutions réalisables de (SDP) et $(DSDP)$ alors

$$\langle Q, X \rangle \leq \langle b, \beta \rangle$$

Dualité forte : Si pour X et β solutions admissibles de (SDP) et $(DSDP)$, le saut de dualité vaut 0, alors il y a dualité forte et X et β sont les valeurs optimales de (SDP) et $(DSDP)$. Et on a :

$$\langle b, \beta \rangle - \langle Q, X \rangle = 0 \iff \langle X, \sum_{r=1}^m \beta_r A_r - Q \rangle = 0$$

Cependant, il n'y a pas nécessairement dualité forte en SDP.

Dualité faible et forte : exemple [Vandenberghe and Boyd]

$$(SDP_{ex}) \begin{cases} \max -X_{12} \\ \text{s.t.} \\ \begin{pmatrix} 0 & X_{12} & 0 \\ X_{12} & X_{22} & 0 \\ 0 & 0 & 1 + X_{12} \end{pmatrix} \succeq 0 \end{cases}$$

$$(DSDP_{ex}) \begin{cases} \min \beta_1 \\ \text{s.t.} \\ \begin{pmatrix} \beta_2 & \frac{1-\beta_1}{2} & \frac{\beta_3}{2} \\ \frac{1-\beta_1}{2} & 0 & \frac{\beta_4}{2} \\ \frac{\beta_3}{2} & \frac{\beta_4}{2} & \beta_1 \end{pmatrix} \succeq 0 \end{cases}$$

Exercice : Retrouvez ce dual et montrer quel est le saut de dualité

Certificat de saut de dualité nul (conditions de Slater)

Théorème : Soient p^* = la valeur optimale de (SDP) et d^* la valeur optimale de $(DSDP)$

- i si (SDP) admet une solution strictement réalisable et p^* fini, alors $p^* = d^*$ et cette valeur est atteinte pour $(DSDP)$
- ii si $(DSDP)$ admet une solution strictement réalisable et d^* fini, alors $p^* = d^*$ et cette valeur est atteinte pour (SDP)
- iii si (SDP) et $(DSDP)$ ont des solutions strictement réalisables alors $p^* = d^*$ et cette valeur est atteinte pour les deux problèmes.

Conditions d'optimalité

Pour les problèmes à dualité forte, nous avons les conditions nécessaires et suffisantes d'optimalité suivantes :

$$(OPT) \left\{ \begin{array}{ll} g_r(\mathbf{X}) = \langle \mathbf{A}_r, \mathbf{X} \rangle - b_r = 0 & \forall r, \mathbf{X} \succeq 0 \text{ (admissibilité du primal)} \\ \sum_{r=1}^m \beta_r \mathbf{A}_r - \mathbf{S} = \mathbf{Q} & \beta \in \mathbb{R}^m \text{ (admissibilité du dual)} \\ \mathbf{S}\mathbf{X} = 0 & \text{complémentarité de } \mathbf{X} \text{ et } \mathbf{S} \\ \mathbf{S} \succeq 0, \mathbf{X} \succeq 0 & \end{array} \right.$$

- 1 Ensemble convexes
 - Les ensembles convexes
 - Les cônes convexes

- 2 Les matrices semi-définies positives

- 3 Les programmes semi-définis

- 4 Algorithmes de résolution des programmes semi-définis
 - Algorithmes des points intérieurs

Quels algorithmes pour résoudre (*SDP*)

- **Méthode des points intérieurs** (implementations CSDP, SeDuMi, SDPA).
En général ne sont pas capables de résoudre des instances de dimension supérieures à 1000 ou de plus de 10000 contraintes.
- **Méthode des faisceaux spectraux** basée sur l'optimisation des valeurs propres (implementations Spectral Bundle)
L'idée est d'exploiter le fait que $X \succeq 0 \iff \lambda_{\min}(X) \geq 0$
- **Méthodes des faisceaux** (callable Conic Bundle library)
Permet de résoudre des instances significativement plus grandes, doit être bien paramétrée car sa convergence peut-être très longue.
- D'autres classes de méthodes comme les **algorithmes du lagrangien augmenté**.

Algorithmes des points intérieurs

Les méthodes de points intérieurs - Principe

- Méthodes développées dans les années 90, il existe plusieurs variantes.
- **La méthode primale-duale** : On suppose que le primal (SDP) et le dual ($DSDP$) satisfont les conditions de Slater (saut de dualité nul).
- **Principe** : suivre un chemin central dans l'intérieur de la région admissible qui est obtenu en remplaçant les conditions d'optimalités par des conditions d'optimalités approchées

Méthodes de points intérieurs - Principe

- **Idée** : se débarrasser de la contrainte difficile $X \succeq 0$
- **Méthode** : on pénalise l'objectif par une fonction barrière qui tend vers 0 à l'intérieur du cône \mathcal{S}_+^n et vers l'infini ailleurs.

\implies si on démarre d'une solution $X \succ 0$, on peut retirer les contraintes $X \succeq 0$ puisqu'elles seront forcées par la fonction barrière (on reste dans l'intérieur du cône \mathcal{S}_+^n).

Points intérieurs - Problème auxiliaire

Soit la fonction barrière : $B(\mathbf{X}) = \begin{cases} \log(\det(\mathbf{X})) & \mathbf{X} \succ 0 \\ -\infty & \text{sinon} \end{cases}$

Remarque : On a bien pour $\mathbf{X} \succ 0$, $\det(\mathbf{X}) = \prod_{i=1}^n \lambda_i > 0$.

Pour $\mu \geq 0$, on définit le problème auxiliaire suivant :

$$(SDP_{\mu}) \begin{cases} \max f_{\mu}(\mathbf{X}) = \langle \mathbf{Q}, \mathbf{X} \rangle + \mu \log(\det(\mathbf{X})) \\ \text{s.t.} \\ g_r(\mathbf{X}) = \langle \mathbf{A}_r, \mathbf{X} \rangle - b_r = 0 & \forall r = 1, \dots, m \\ \mathbf{X} \succ 0 \end{cases}$$

Lemme : $\mathbf{X} \mapsto \log(\det(\mathbf{X}))$ est strictement concave dans l'intérieur de \mathcal{S}_+^n , i.e. quand $\mathbf{X} \succ 0$.

\implies Pour $\mu \geq 0$ fixé, la solution optimale \mathbf{X}_{μ}^* de (SDP_{μ}) est unique.

Résoudre (SDP) avec son problème auxiliaire

Pour $\mu \rightarrow 0$, il faut que $\langle Q, X_{\mu}^* \rangle$ converge vers la valeur optimale de (SDP).

Nous allons :

- i Ecrire les CN de l'existence de X_{μ}^* , sous la forme d'un système que X_{μ}^* doit satisfaire. Ces conditions vont aussi impliquer la convergence de $\langle Q, X_{\mu}^* \rangle$ vers la valeur optimale de (SDP).
- ii Observer que les CN sont aussi suffisantes : le système issu de (i) a une solution unique $X = X_{\mu}^*$

$\forall g_r(\mathbf{X}) = 0$, nous associons un multiplicateur de lagrange $\beta_r \in \mathbb{R}$.

On considère ensuite le lagrangien :

$$\mathcal{L}_\mu(\mathbf{X}, \beta) = \langle Q, \mathbf{X} \rangle + \mu \log(\det(\mathbf{X})) - \sum_{r=1}^m \langle \beta_r, g_r(\mathbf{X}) \rangle$$

Théorème de Lagrange : \mathbf{X}_μ^* est un maximum (local) si :

- ① $g_r(\mathbf{X}_\mu^*) = 0$, pour tout r
- ② $\exists! \beta^* \in \mathbb{R}^m$ tel que $\nabla \mathcal{L}_\mu(\mathbf{X}_\mu^*, \beta^*) = \nabla f_\mu(\mathbf{X}_\mu^*) + \sum_{r=1}^m \beta_r^* \nabla g_r(\mathbf{X}_\mu^*) = 0$

Cela s'applique si :

- ① si f_μ et g_r sont définies sur un ouvert de \mathbb{R}^n
- ② si f_μ et g_r sont différentiables
- ③ les $\nabla g_r(\mathbf{X})$ sont indépendants (satisfait car les g_r sont des fonctions linéaires).

Etape (i) : CN de l'existence de X_μ^*

$$(SDP_\mu) \begin{cases} \max f_\mu(\mathbf{X}) = \langle \mathbf{Q}, \mathbf{X} \rangle + \mu \log(\det(\mathbf{X})) \\ \text{s.t.} \\ g_r(\mathbf{X}) = \langle \mathbf{A}_r, \mathbf{X} \rangle - b_r = 0 \\ \mathbf{X} \succ 0 \end{cases} \quad \forall r = 1, \dots, m$$

Propriété : $\nabla \log(\det(\mathbf{X})) = (\mathbf{X}^T)^{-1}$

$$\text{Calculons } \nabla \mathcal{L}_\mu(\mathbf{X}, \beta) = \nabla f_\mu(\mathbf{X}) + \sum_{r=1}^m \beta_r \nabla g_r(\mathbf{X})$$

- On a $\nabla \log(\det(\mathbf{X})) = (\mathbf{X}^T)^{-1}$ et $\nabla \langle \mathbf{Q}, \mathbf{X} \rangle = \nabla \text{Tr}(\mathbf{Q}^T \mathbf{X}) = \mathbf{Q}^T$
- Donc $\nabla f_\mu(\mathbf{X}) = \mathbf{Q}^T + \mu (\mathbf{X}^T)^{-1}$
- On a $\nabla g_r(\mathbf{X}) = \nabla(\langle \mathbf{A}_r, \mathbf{X} \rangle - b_r) = \nabla(\text{Tr}(\mathbf{A}_r^T \mathbf{X}) - b_r) = \mathbf{A}_r^T$

$$\text{On a donc } \nabla \mathcal{L}_\mu(\mathbf{X}, \beta) = \mathbf{Q}^T + \mu (\mathbf{X}^T)^{-1} - \sum_{r=1}^m \beta_r \mathbf{A}_r^T = 0$$

CN d'optimalité pour (SDP_μ)

En introduisant la variable d'écart $S = \sum_{r=1}^m \beta_r A_r^T - Q^T = \mu (X^T)^{-1}$, on obtient les CN suivantes :

$$(OPT_\mu) \begin{cases} g_r(X) = \langle A_r, X \rangle = b_r & \forall r, X \succ 0 \text{ (admissibilité du primal)} \\ \sum_{r=1}^m \beta_r A_r - S = Q & \beta \in \mathbb{R}^m, S \succ 0 \text{ (admissibilité du dual)} \\ SX = \mu I_n & \text{complémentarité de } X \text{ et } S \end{cases}$$

En fait, par rapport à (SDP) , on a remplacé la dernière condition par $SX = 0$ par $SX = \mu I_n$, avec $\mu > 0$ et $\mu \rightarrow 0$.

Une interprétation primale duale

- Puisque X_μ^* satisfait (OPT_μ) si il existe, on peut en déduire que $f_\mu(X_\mu^*) = \langle Q, X_\mu^* \rangle + \mu \log(\det(X_\mu^*))$ converge vers la valeur du problème initial (SDP) : $f(X) = \langle Q, X \rangle$ lorsque μ tend vers 0.
- En plus, on a une paire primale-duale (X, β, S) dont le saut de dualité dépend de μ :
Si $(\hat{X}, \hat{\beta}, \hat{S})$ satisfont (OPT_μ) , alors \hat{X} est une solution strictement réalisable de (SDP) , $(\hat{\beta}, \hat{S})$ est une solution strictement réalisable de $(DSDP)$ et le saut de dualité est :

$$\langle b, \hat{\beta} \rangle - \langle Q, \hat{X} \rangle = n\mu$$

Exercice : preuve

Une interprétation primale duale

- Du coup si nous pouvons calculer X_μ^* pour μ très petit, nous aurons une solution "presque" optimale de (SDP) .
- De plus comme par la faible dualité, $\langle Q, X \rangle \leq \langle b, \beta \rangle$, la valeur de la solution $\langle Q, X_\mu^* \rangle$ sera au plus à $n\mu$ de la solution optimale de (SDP) .

Etape (ii) : CS pour l'optimalité de X_μ^*

Supposons que (SDP) ait une solution intérieure $\hat{X} \succ 0$ et que son dual (DSDP) ait aussi une solution intérieure $\hat{\beta}$ telle que la matrice d'écart \hat{S} satisfaisant $\hat{S} = \sum_{r=1}^m \beta_r A_r - Q$ soit définie positive (i.e. $S \succ 0$).

De plus si les A_r sont linéairement indépendantes, alors $\forall \mu \geq 0$, (SDP_μ) a une unique solution $X^* = X_\mu^*$, $\beta^* = \beta_\mu^*$ et $S^* = S_\mu^*$, et X_μ^* est l'unique maximum de f_μ sous les contraintes $g_r(X) = 0$ et $X \succ 0$

Méthode des points intérieurs : algorithme

Soient un point strictement réalisable $(\mathbf{X}, \beta, \mathbf{S}) = (\mathbf{X}^0, \beta^0, \mathbf{S}^0)$, et un $\epsilon > 0$

- 1 Calculer le μ courant : $\mu = \frac{\mathbf{S}\mathbf{X}}{n}$
- 2 Calculer les nouvelles directions de recherche pour le nouveau μ , i.e. calculer $(\Delta\mathbf{X}, \Delta\beta, \Delta\mathbf{S})$ en résolvant (OPT_μ) :

$$(OPT_\mu) \begin{cases} \langle \mathbf{A}_r, \mathbf{X} + \Delta\mathbf{X} \rangle = b_r & \forall r, \mathbf{X} \succ 0 \\ \sum_{r=1}^m (\beta_r + \Delta\beta_r) \mathbf{A}_r - (\mathbf{S} + \Delta\mathbf{S}) = \mathbf{Q} & \beta \in \mathbb{R}^m, \mathbf{S} \succ 0 \\ (\mathbf{S} + \Delta\mathbf{S})(\mathbf{X} + \Delta\mathbf{X}) = \mu \mathbf{I}_n \end{cases}$$

- 3 Calcul de la distance de recherche : Calculer les distances α_p et α_d telles que :

$$\mathbf{X} + \alpha_p \Delta\mathbf{X} \succ 0 \quad \mathbf{S} + \alpha_d \Delta\mathbf{S} \succ 0$$

- 4 Mise à jour : $\mathbf{X} = \mathbf{X} + \alpha_p \Delta\mathbf{X}$, $\beta = \beta + \alpha_d \Delta\beta$ et $\mathbf{S} = \mathbf{S} + \alpha_d \Delta\mathbf{S}$
- 5 Si $\mathbf{S}\mathbf{X} < \epsilon$, s'arrêter.

Etape 2 : approximation des directions de recherche

Résolution du système :

$$(OPT_{\mu}) \begin{cases} \langle A_r, X + \Delta X \rangle = b_r & \forall r, X \succeq 0 \\ \sum_{r=1}^m (\beta_r + \Delta\beta_r) A_r - (S + \Delta S) = Q & \beta \in \mathbb{R}^m \\ (S + \Delta S)(X + \Delta X) = \mu I_n \end{cases}$$

Les deux premières équations sont linéaires.

Approximer la troisième par une fonction linéaire (ici Helmberg et al. 96) :

$$\begin{aligned} \mu I_n &= (S + \Delta S)(X + \Delta X) \\ &= SX + S\Delta X + \Delta SX + \Delta S\Delta X \approx SX + S\Delta X + X\Delta S \end{aligned}$$

Remarque : même si X et S sont symétriques, le produit $X\Delta S$ ne l'est pas forcément. Lorsque ça arrive, on considère la matrice $\hat{X}\hat{S} = \frac{SX + (SX)^T}{2}$.