Solvers Principles and Architecture (SPA)

Convex Optimization

Master Sciences Informatique (Sif)
2017–2019
Rennes

Khalil Ghorbal
khalil.ghorbal@inria.fr

- Find an **optimal value** of a function with respect to some **constraints**
- Optimum: minimum or maximum
- The function to optimize is called the **objective** or **cost** function
- The constraints form a set called the **feasible set**

$$\begin{aligned} \min / \max \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{aligned}$$

- x denotes a point in some vector space (e.g. $\mathbb{R}^n$)
- All functions are real valued: their codomain is $\mathbb{R}$
- The codomain of the constraints f_i, $1 \leq i \leq m$, will be generalized later, together with the order relation $(\leq)$

Optimal value:

$$\mathfrak{p}^\star = \inf / \sup \left\{ f_0(x) \mid \bigwedge_{i=1}^{m} f_i(x) \leq 0 \wedge \bigwedge_{j=1}^{p} h_j(x) = 0 \right\}$$

- Let $\Diamond$ and $\Box$ be elements of some vector space V
- $\mathbb{R}^n$, $\mathcal{M}^n$, $\mathcal{S}^n$, etc.
- An inner product is a bilinear function from $V \times V$ to $\mathbb{R}$

$$\Diamond \cdot \Box$$

Outline

$$\begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & Ax \leq b \end{aligned} \qquad\qquad \begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- $c, x \in \mathbb{R}^n$
- $c \cdot x$ is the **inner product** of c and x
- A an $m \times n$ matrix (over the reals)
- $b \in \mathbb{R}^m$
- $x, y \in \mathbb{R}^k$, $x \leq y$ means $y - x \in \mathbb{R}^k_+$ (**non negative orthant**)

$$\exists x \in \mathbb{R}^n.\, Ax \leq b \quad \longleftrightarrow \quad \exists s \in \mathbb{R}^k_+.\, A's = b$$

Saturation Procedure

- add 2 fresh variables for each variable
- add a fresh variable for each row of A
- $k = 2n + \#\text{rows of } A$

Example

$$\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} \leq 1,\ \text{in } \mathbb{R}^2 \xleftarrow[x_2=s_3-s_4]{x_1=s_1-s_2}\rightarrow \begin{pmatrix}1 & -1 & 1\end{pmatrix}\begin{pmatrix}s_1\\s_2\\s_5\end{pmatrix} = 1,\ \text{in } \mathbb{R}^5_+$$

$x \in \mathbb{R}^{\mathbf{n}}_{+}$, $Ax = b$, $\text{rank}(A) = m \leq n$ (empty polyhedron otherwise).

Base (algebraic vertex)

Let $\{\mathfrak{B}, \mathfrak{N}\}$ be a partition of $\{1, \ldots, n\}$. $\mathfrak{B}$ is a *base* if and only if $|\mathfrak{B}| = \text{rank}(A_{\mathfrak{B}})$ where $A_{\mathfrak{B}}$ is the submatrix of A with columns in $\mathfrak{B}$. $\mathfrak{B}$ is *non-degenerate* if $|\mathfrak{B}| = m$, and *degenerate* otherwise ($|\mathfrak{B}| < m$).

Example

For $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$, $\{1\}$ and $\{3\}$ are degenerate bases while $\{i, j\}$, $1 \leq i < j \leq 3$, are non-degenerate.

Proposition

Let $\mathfrak{B}$ be a base. The unique point v (if any) in the polyhedron such that $v_i = 0$ for all $i \in \mathfrak{N}$ (i.e. $i \notin \mathfrak{B}$) is a **vertex** (**facet of dimension zero**). (Such a point may not exist since $A_{\mathfrak{B}}^{-1} b$ has to be non-negative.)

(Weak) **Correspondence**

- Each vertex has **at least** one base.
- Each base has **at most** one vertex.

Examples

- The polyhedron $x_1, x_2 \in \mathbb{R}_+, -x_1 + x_2 = 1$ has no vertex associated with the (non-degenerate) base $\mathfrak{B} = \{1\}$ because $A_{\mathfrak{B}}^{-1} b < 0$.
- The polyhedron $x_1, x_2 \in \mathbb{R}_+, x_1 + x_2 = 0$ has the same vertex, $(0, 0)$ associated with two (non-degenerate) bases: $\mathfrak{B} = \{1\}$ and $\mathfrak{B}' = \{2\}$.

Let $\mathfrak{B}$ be a base associated with the vertex v. For simplicity, suppose that $\mathfrak{B}$ is non-degenerate so that $A_\mathfrak{B}$ is invertible. Thus, for all $x = (x_\mathfrak{B}\ x_\mathfrak{N})^t$:

$$Ax = \begin{pmatrix} A_\mathfrak{B} & A_\mathfrak{N} \end{pmatrix} \begin{pmatrix} x_\mathfrak{B} \\ x_\mathfrak{N} \end{pmatrix} = A_\mathfrak{B} x_\mathfrak{B} + A_\mathfrak{N} x_\mathfrak{N} = b \implies x_\mathfrak{B} = A_\mathfrak{B}^{-1}(b - A_\mathfrak{N} x_\mathfrak{N})$$

The above equation has a solution in the non-negative orthant, namely v. Suppose that the polyhedron is not reduced to a point. Then, there exists a positive real number ϵ such that:

$$\forall x_\mathfrak{N} \in \mathbb{R}_+^{|\mathfrak{N}|} \quad \|x_\mathfrak{N}\|_\infty \leq \epsilon \implies x_\mathfrak{B} = A_\mathfrak{B}^{-1}(b - A_\mathfrak{N} x_\mathfrak{N}) \geq 0$$

We next **solve** the original optimization problem **locally** around v.

$$\begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \min \quad & r \cdot x_{\mathfrak{N}} + a \\ s.t. \quad & x_{\mathfrak{N}} \geq 0 \\ & \|x_{\mathfrak{N}}\|_\infty \leq \epsilon \end{aligned}$$

$$c \cdot x = \begin{pmatrix} c_{\mathfrak{B}} \\ c_{\mathfrak{N}} \end{pmatrix} \cdot \begin{pmatrix} A_{\mathfrak{B}}^{-1}(b - A_{\mathfrak{N}} x_{\mathfrak{N}}) \\ x_{\mathfrak{N}} \end{pmatrix} = \underbrace{\left(c_{\mathfrak{N}} - A_{\mathfrak{N}}^t {A_{\mathfrak{B}}^{-}}^t c_{\mathfrak{B}}\right)}_{r} \cdot x_{\mathfrak{N}} + \underbrace{c_{\mathfrak{B}} \cdot A_{\mathfrak{B}}^{-1} b}_{a}$$

- As long as $\|x_{\mathfrak{N}}\|_\infty \leq \epsilon$, the point $(A_{\mathfrak{B}}^{-1}(b - A_{\mathfrak{N}} x_{\mathfrak{N}}), x_{\mathfrak{N}})$ is **feasible**
- $r \cdot x_{\mathfrak{N}}$ is called the **reduced cost function**

- We seek a displacement that **locally** decreases $r \cdot x_{\mathfrak{N}}$
- Suppose that there exists a index j such that $r_j < 0$
- Consider a displacement along this jth coordinate
- Let e_j denote the jth vector of the canonical orthonormal basis of $\mathbb{R}^{|r|}$
- Let ρ be a positive real number: $x_{\mathfrak{N}} \leftarrow v_{\mathfrak{N}} + \rho e_j$

$$r \cdot x_{\mathfrak{N}} = r \cdot (v_{\mathfrak{N}} + \rho e_j) = r \cdot v_{\mathfrak{N}} + \rho r \cdot e_j = r \cdot v_{\mathfrak{N}} + \rho r_j < r \cdot v_{\mathfrak{N}}$$

Optimality criterion: $r \geq 0$

- If $r \geq 0$: no possible minimization for $r \cdot x_{\mathfrak{N}}$ since $x_{\mathfrak{N}} \geq 0$
- The only **local minimum** is $x_{\mathfrak{N}} = v_{\mathfrak{N}} = 0$
- which is also **global by convexity**

- Recall that locally $x_{\mathfrak{B}} = A_{\mathfrak{B}}^{-1}(b - A_{\mathfrak{N}} x_{\mathfrak{N}})$
- So the update $x_{\mathfrak{N}} \leftarrow v_{\mathfrak{N}} + \rho e_j$ leads to

$$x_{\mathfrak{B}} \leftarrow A_{\mathfrak{B}}^{-1}(b - A_{\mathfrak{N}}(v_{\mathfrak{N}} + \rho e_j)) = \underbrace{A_{\mathfrak{B}}^{-1} b}_{v_{\mathfrak{B}}} - A_{\mathfrak{B}}^{-1} A_{\mathfrak{N}} \underbrace{v_{\mathfrak{N}}}_{0} - \rho \underbrace{A_{\mathfrak{B}}^{-1} A_{\mathfrak{N}} e_j}_{\delta_{\mathfrak{B}}}$$

- Since $x_{\mathfrak{B}} \geq 0$, we get $v_{\mathfrak{B}} \geq \rho \delta_{\mathfrak{B}}$
- This gives an upper bound for ρ:

$$\rho \leq \min_i \left\{ \frac{(v_{\mathfrak{B}})_i}{(\delta_{\mathfrak{B}})_i} \mid (\delta_{\mathfrak{B}})_i > 0 \right\}$$

Unboundedness criterion: $\delta_{\mathfrak{B}} \leq 0$
ρ can be chosen arbitrarily big and the minimum is $-\infty$

When $x_{\mathfrak{N}} \leftarrow v_{\mathfrak{N}} + \rho e_j$:

- The jth component of $x_{\mathfrak{N}}$ becomes strictly positive
- When ρ increases, x **moves along an edge** (a facet of dimension 1)
- If ρ is unbounded, the minimum is $-\infty$ (**halt**)
- If ρ is bounded, one component (say the ith) of $x_{\mathfrak{B}}$ vanishes when ρ reaches its upper bound: we **reach a new vertex**.
- **update the base**: let $(\mathfrak{B}', \mathfrak{N}') = ((\mathfrak{B} \setminus \{i\}) \cup \{j\}, (\mathfrak{N} \setminus \{j\}) \cup \{i\})$
- If $\text{rank}(A_{\mathfrak{B}'}) = m$, then $\mathfrak{B}'$ is a new non-degenerate base
- Otherwise, $\text{rank}(A'_{\mathfrak{B}}) < m$, and we can remove some elements from $\mathfrak{B}'$ (other than j) to make it a non-degenerate base
- **repeat** if the optimality criterion ($r \geq 0$) is not met.

1. Start at a vertex (base)
2. If the optimality criterion is satisfied, halt: the problem is solved
3. Otherwise, move along an edge that minimizes the reduced cost function
4. If the unboundedness criterion is satisfied, halt: the problem is unbounded
5. Otherwise, we reach a new vertex and we loop back to the first step

Does it always terminate?

$$\begin{aligned} \min \quad & x_1 - x_2 \\ s.t. \quad & x_1 + x_2 = 0 \\ & x \geq 0 \end{aligned}$$

- Start with the base $\mathfrak{B} = \{1\}$, $\mathfrak{N} = \{2\}$
- $v = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $A_{\mathfrak{B}} = A_{\mathfrak{N}} = \begin{pmatrix} 1 \end{pmatrix}$
- $r = c_{\mathfrak{N}} - A_{\mathfrak{N}}^t A_{\mathfrak{B}}^{-t} c_{\mathfrak{B}} = \begin{pmatrix} -2 \end{pmatrix}$ and $\delta_{\mathfrak{B}} = A_{\mathfrak{B}}^{-1} A_{\mathfrak{N}} e_j = \begin{pmatrix} 1 \end{pmatrix}$
- update $x_{\mathfrak{N}} \leftarrow 0 + \rho$, $x_{\mathfrak{B}} \leftarrow 0 - \rho$ $(\rho = 0)$
- So the algorithm is **updating the base without changing the vertex**

Outline

The **primal** problem is the **minimization** problem (by convention).

$$\begin{aligned} \text{min} \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \qquad (\mathfrak{p}) \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{aligned}$$

Intuition: inject the constraint into the objective function.
The Lagrangian associated to $(\mathcal{P})$ is defined by:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x),$$

- No **extra constraints** for x (as long as the functions are defined)
- λ_i, $i = 1, \ldots, m$, are **non negative real numbers**
- μ_j, $j = 1, \ldots, p$, are **unconstrained real numbers**

$$L(x,\lambda,\mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \ .$$

- If there exists an $\bar{x}$ and an index i such that $f_i(\bar{x}) > 0$, then $L(\bar{x},\lambda,\mu)$ is unbounded since λ_i can be chosen arbitrarily big.
- If there exists an $\bar{x}$ and an index j such that $h_j(\bar{x}) \neq 0$, then $L(\bar{x},\lambda,\mu)$ is also unbounded since μ_j can be chosen arbitrarily big or small depending on the sign of $h_j(\bar{x})$.

$$\sup_{\lambda\geq 0,\mu} L(x,\lambda,\mu) = \begin{cases} f_0(x) & \text{if } \bigwedge_i f_i(x) \leq 0 \wedge \bigwedge_j h_j(x) = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Solving $(\mathfrak{p})$ is then equivalent to minimizing $\sup_{\lambda\geq 0,\mu} L(x,\lambda,\mu)$ over x:

$$\mathfrak{p}^\star = \inf_x \sup_{\lambda\geq 0,\mu} L(x,\lambda,\mu)$$

In general, if L is a real valued function defined over the product $X \times Y$, then

$$\sup_y \inf_x L(x,y) \leq \inf_x \sup_y L(x,y)$$

Proof. Let $(\bar{x},\bar{y}) \in X \times Y$, then, by definition of inf and sup

$$\inf_x L(x,\bar{y}) \leq L(\bar{x},\bar{y}) \leq \sup_y L(\bar{x},y)$$

So $\sup_y L(\bar{x},y)$ is an upper bound of $\inf_x L(x,\bar{y})$. Since the sup is the smallest upper bound by definition, one gets

$$\sup_{\bar{y}} \inf_x L(x,\bar{y}) \leq \sup_y L(\bar{x},y)$$

But then $\sup_{\bar{y}} \inf_x L(x,\bar{y})$ is a lower bound for $\sup_y L(\bar{x},y)$. Since, dually, the inf is the biggest lower bound, one gets the desired result:

$$\sup_{\bar{y}} \inf_x L(x,\bar{y}) \leq \inf_{\bar{x}} \sup_y L(\bar{x},y) \quad .$$

By the weak duality, we get a **lower bound** of the optimal value $\mathfrak{p}^\star$:

$$\mathfrak{d}^\star := \sup_{\lambda \geq 0, \mu} \inf_x L(x, \lambda, \mu) \quad \leq \quad \inf_x \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu) = \mathfrak{p}^\star$$

where $\mathfrak{d}^\star$ denotes the objective value of a distinct, yet related, optimization problem, $(\mathfrak{d})$, called the **dual problem**, and defined by $\sup_{\lambda \geq 0, \mu} \inf_x L(x, \lambda, \mu)$, for the **exact same Lagrangian** L of $(\mathfrak{p})$.

$$\begin{aligned} \max \quad & g(\lambda, \mu) := \inf_x L(x, \lambda, \mu) \\ s.t. \quad & \lambda_i \geq 0, \quad i = 1, \ldots, m \qquad (\mathfrak{d}) \end{aligned}$$

- The evaluation of the dual cost function on **any** feasible point of the dual problem bounds from below $\mathfrak{p}^\star$ (primal optimum):

$$\forall(\lambda, \mu) \in \mathbb{R}_+^m \times \mathbb{R}^p. \quad g(\lambda, \mu) \leq \mathfrak{p}^\star$$

- If the primal is unbounded ($\mathfrak{p}^\star = -\infty$) then the dual is unfeasible
- If the dual is unbounded ($\mathfrak{d}^\star = +\infty$) then the primal is unfeasible
- The primal and dual **cannot be unbounded simultaneously**
- The primal and the dual **can be both unfeasible** ($-\infty \leq +\infty$)

$$\begin{aligned} \min \quad & -x \\ s.t. \quad & 0x + 1 \leq 0 \end{aligned} \qquad (\mathfrak{p}) \qquad\qquad \begin{aligned} \max \quad & \lambda \\ s.t. \quad & 0\lambda - 1 = 0 \end{aligned} \qquad (\mathfrak{d})$$

Weak duality: Always true

$$\mathfrak{d}^\star \leq \mathfrak{p}^\star$$

Strong duality: Not true in general

$$\mathfrak{d}^\star = \mathfrak{p}^\star$$

Sufficient conditions under which the strong duality holds are known as **constraint qualifications**.

- $f_0(x) = c \cdot x$ for some fixed vector $c \in \mathbb{R}^n$
- $f_i(x) = -x_i$, $i = 1, \ldots, n$ ($m = n$ in this case)
- $h_j(x) = A_j \cdot x - b_j$, $j = 1, \ldots, p$, for some fixed $A_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$

$$L(x, \lambda, \mu) = c \cdot x + \underbrace{\sum_{i=1}^{n} \lambda_i(-x_i)}_{-\lambda \cdot x} + \underbrace{\sum_{j=1}^{p} \mu_j(A_j \cdot x - b_j)}_{\mu \cdot (Ax - b)}$$

The Lagrangian L could be rearranged as follows (recall that $Ax \cdot y = x \cdot A^t y$, where A^t denotes the transpose of the matrix A):

$$L(x, \lambda, \mu) = -b \cdot \mu + x \cdot (A^t \mu + c - \lambda)$$

and we get:

$$\inf_x L(x, \lambda, \mu) = \begin{cases} -b \cdot \mu & \text{if } A^t \mu + c - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

$$\begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & Ax = b \qquad (\mathfrak{p}) \\ & x \geq 0 \end{aligned} \qquad \begin{aligned} \max \quad & -b \cdot \mu \\ s.t. \quad & A^t \mu + c - \lambda = 0 \qquad (\mathfrak{d}) \\ & \lambda \geq 0 \end{aligned}$$

There are **several possible formulations**, for instance:

$$\begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & Ax \leq b \qquad (\mathfrak{p}) \end{aligned} \qquad \begin{aligned} \max \quad & -b \cdot \lambda \\ s.t. \quad & A^t \lambda + c = 0 \qquad (\mathfrak{d}) \\ & \lambda \geq 0 \end{aligned}$$

In this case (everything is linear), they are all **dual** of each other!

The **reduced problem** has the form ($\epsilon > 0$, $|\mathfrak{N}| = k$):

$$\begin{aligned} \min \quad & r \cdot x_{\mathfrak{N}} \\ s.t. \quad & \begin{pmatrix} -I_k \\ I_k \end{pmatrix} x_{\mathfrak{N}} \leq \begin{pmatrix} 0 \\ \epsilon \end{pmatrix} \end{aligned} \qquad (\mathfrak{p})$$

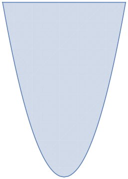
$$\begin{aligned} \max \quad & -\begin{pmatrix} 0 \\ \epsilon \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = -\epsilon \cdot \lambda_2 \\ s.t. \quad & \begin{pmatrix} -I_k & I_k \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} + r = -\lambda_1 + \lambda_2 + r = 0 \\ & \lambda_1 \geq 0, \lambda_2 \geq 0 \end{aligned} \qquad (\mathfrak{d})$$

So $\lambda_2^* = 0$ and $r = \lambda_1^*$. Thus $r \geq 0$ which is the **optimality criterion**.

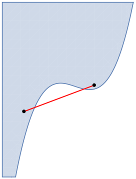
- The objective function $g(\lambda, \mu)$ is **concave** (to be proven later)
- The feasible set is **convex**
  - λ belongs to the non negative orthant $\mathbb{R}_+^m$
  - μ is unconstrained

What is convexity?

- **Intuition**: A set C is convex if and only if, for any two points in C, the shortest path that links these two points is also entirely in C.
- A point in a vector space is a vector and one can define scalar multiplication, addition etc.
- In these settings, C is convex if and only if, for all $c_1, c_2 \in C$, for all $\lambda \in [0, 1], \lambda c_1 + (1 - \lambda) c_2$ is also in C.



Convex



Non convex

Definition: The **epigraph** of a function $f : \mathcal{D} \to \mathbb{R}$ is defined by

$$\text{epi}(f) := \{(x, y) \mid f(x) \leq y\} \subset \mathcal{D} \times \mathbb{R}$$

- f is **convex** if and only if its epigraph is a convex set
- f is **concave** if and only if $-f : x \mapsto -f(x)$ is convex

Examples:

- $f : x \mapsto x^2$ is convex (cf. left figure in the previous slide)
- $f : x \mapsto x^3 + x^2$ is not convex (cf. right figure in the previous slide)

- $\forall \lambda \in [0,1].\ \forall x, y. \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
- **Intuition**: the image of a point in the segment joining x and y is somewhere below the segment joining $f(x)$ and $f(y)$
- **Any local minimum of f is also a global minimum**
- One can define a weak notion of differentiability over convex functions
- The **sub-differential** of f at x is defined by the following set:

$$\partial f(x) := \{z \in \mathbb{R}^n \mid \forall t \in \mathbb{R}^n. \quad f(t) \geq f(x) + z \cdot (t - x)\}$$

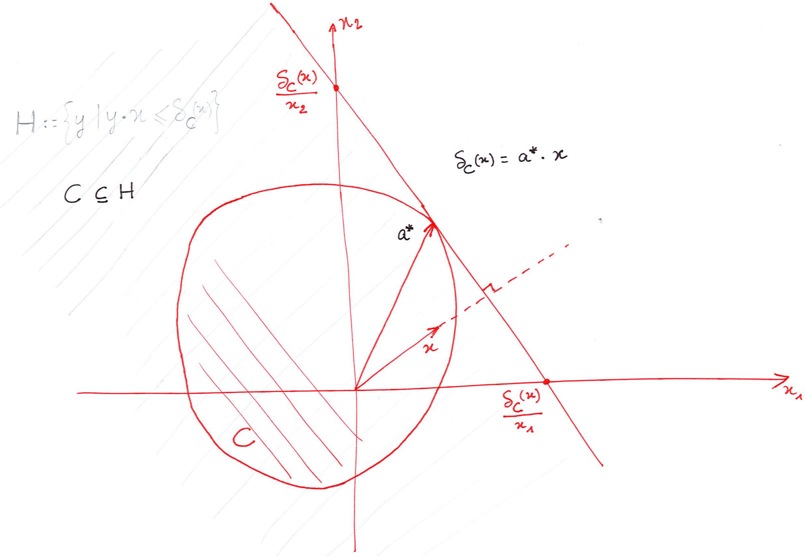
  where $x \cdot y$ denotes the usual scalar product over $\mathbb{R}^n$
- **Intuition**: the sub-differential at x is the set of all affine functions that touches the graph of f only at x
- **Example**: the absolute value function is non-differentiable at 0 in the usual sense, but it is sub-differentiable, $\partial f(0) = [-1, 1]$

Let C be **any non-empty** subset of a vector space equipped with an **inner product** denoted by $(\cdot)$.

Support function of a set

$$\delta_C(x) := \sup_{a \in C}\{x \cdot a\}$$

- δ_C is defined for any vector x
- δ_C, as a function of x, is **convex**


x_2
$\frac{\delta_C(x)}{x_2}$
$H := \{y \mid y \cdot x \leq \delta_C(x)\}$
$\delta_C(x) = a^* \cdot x$
$C \subseteq H$
a^*
x
x_1
$\frac{\delta_C(x)}{x_1}$
C

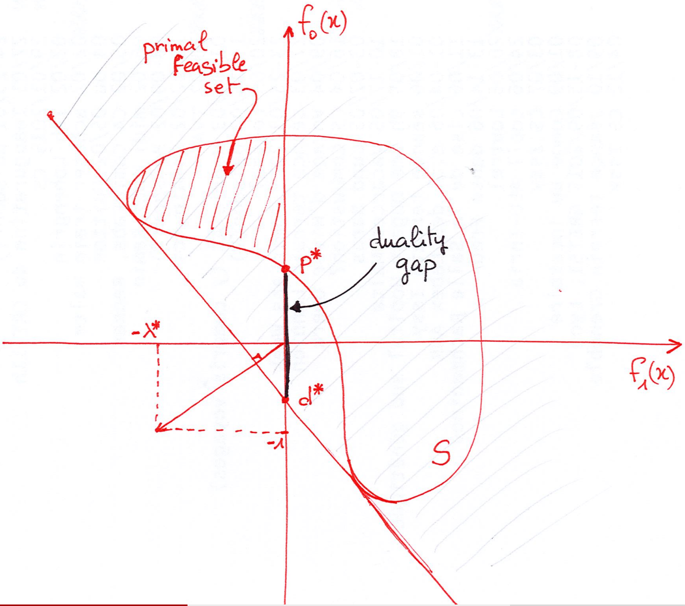
- Let $\nu := (\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p, 1) \in \mathbb{R}^{m+p+1}$
- Let $u_x := (f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x)) \in \mathbb{R}^{m+p+1}$
- Let $S := \{u_x \mid f_i, h_j \text{ are defined } \} \subseteq \mathbb{R}^{m+p+1}$

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) = \nu \cdot u_x$$

The **objective function** g is **concave** (opposite of a support function):

$$\begin{aligned} g(\lambda, \mu) &= \inf_x L(x, \lambda, \mu) \\ &= \inf_x \{\nu \cdot u_x\} \\ &= -\sup_x \{(-\nu) \cdot u_x\} \\ &= -\delta_S(-\nu) \end{aligned}$$

Geometrical intuition: weak vs strong duality



Outline

- f_0 is convex
- $f_i,\ i = 1, \ldots, m$ are convex
- $h_j,\ j = 1, \ldots, p$ are linear in x: $h_j(x) = A_j \cdot x - b_j$

$$
\begin{array}{ll}
\min & f_0(x) \\
s.t. & f_i(x) \leq 0, i = 1, \ldots, m \\
 & A_j \cdot x - b_j = 0, j = 1, \ldots, p
\end{array}
\qquad (\mathfrak{p})
$$

Slater's condition (constraint qualifications for convex problems)

If the **primal is strictly feasible** (i.e. there exists an x such that $f_i(x) < 0,\ i = 1, \ldots, m$, and $A_j \cdot x - b_j = 0,\ j = 1, \ldots, p$), then **strong duality holds** $\mathfrak{d}^\star = \mathfrak{p}^\star < +\infty$.

Let (λ^*, μ^*) be the optimum dual and x^* be the optimum primal:

- x^* is feasible: $\begin{cases} f_i(x^*) \leq 0 & i = 1, \ldots, m \\ A_j \cdot x^* - b_j = 0 & j = 1, \ldots, p \end{cases}$
- (λ^*, μ^*) is feasible: $\lambda^* \geq 0$

As a consequence of the strong duality, we have in addition:

$$\mathfrak{d}^\star = g(\lambda^*, \mu^*) = \inf_x L(x, \lambda^*, \mu^*) = f_0(x^*) = p^\star$$

Therefore, by definition of the infimum

$$\begin{aligned} f_0(x^*) = \inf_x L(x, \lambda^*, \mu^*) &\leq L(x^*, \lambda^*, \mu^*) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{p} \mu_j (A_j \cdot x^* - b_j) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) \end{aligned}$$

$$\left.\begin{array}{l} 0 \leq \sum_{i=1}^{m} \lambda_i^* f_i(x^*) \\ \lambda_1^*, \ldots, \lambda_m^* \geq 0 \\ f_1(x^*), \ldots, f_m(x^*) \leq 0 \end{array}\right\} \iff \left\{\begin{array}{l} \lambda_i^* f_i(x^*) = 0 \\ \lambda_i^* \geq 0 \\ -f_i(x^*) \geq 0 \end{array}\right. \quad i = 1, \ldots, m$$

Complementarity conditions

$$0 \leq \lambda_i^* \perp -f_i(x^*) \geq 0, \quad i = 1, \ldots, m$$

When f_0, $f_1, \ldots, f_m$ are continuously differentiable (i.e. C^1), the optimum x^* has also to satisfy the following condition:

$$\nabla_x L(x^*, \lambda, \mu) = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) + \sum_{j=1}^{p} \mu_j A_j = 0$$

Recall that

$$\nabla_x L = \left(\frac{\partial L}{\partial x_1}, \ldots, \frac{\partial L}{\partial x_m} \right)$$

Definition

For an optimization problem $(\mathfrak{p})$ with Lagrangian L and such that f_0, $f_1, \ldots, f_m$, $h_1, \ldots, h_p$ are C^1, x^* verify the KKT conditions if and only if there exists some $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that:

1. **Primal feasibility**: $\begin{cases} f_i(x^*) \leq 0 & i = 1, \ldots, m \\ h_j(x^*) = 0 & j = 1, \ldots, p \end{cases}$
2. **Dual feasibility**: $\lambda \geq 0$
3. **Complementarity** $\lambda_i f_i(x^*) = 0, \quad i = 1, \ldots, m$
4. **Stationarity**: $\nabla_x L(x^*, \lambda, \mu) = 0$

Under constraint qualifications, KKT conditions are **only necessary**.

Convex problems

Under Slater's condition, KKT conditions are also sufficient: x^* is optimum if and only if KKT conditions hold.

Outline

$$\begin{aligned} \min \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0, i = 1, \dots, m \qquad (\mathfrak{p}) \\ & A_j \cdot x - b_j = 0, j = 1, \dots, p \end{aligned}$$

- $f_0, f_1, \dots, f_m$ are **convex** and **twice continuously differentiable**
- Slater's condition holds: the problem is strictly feasible
- Thus, strong duality holds and $\mathfrak{p}^\star$ is finite and attained for some x^* that satisfy KKT conditions

Examples: Linear, Quadratic, Geometric Programming (LP, QP, GP)

KKT conditions

x^* is an optimum for $(\mathfrak{p})$ if and only if

- $A_j \cdot x^* - b_j = 0$, $j = 1, \ldots, p$
- $0 \leq \lambda_i^* \perp -f_i(x^*) \geq 0, \quad i = 1, \ldots, m$
- $\nabla_x L(x^*, \lambda, \mu) = 0$

We cannot solve such system numerically as it combines equality and inequality constraints.

Main idea

Design a sequence of optimization problems that we can solve and such that their solutions converges towards the optimum of the original problem.

To get rid of the (problematic) inequality constraints $f_i(x) \leq 0$, one can *hide* them inside indicator functions.

Indicator function

The indicator function of $\mathbb{R}_-$ is a **convex** function defined as follows:

$$\mathcal{I}(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ +\infty & \text{otherwise} \end{cases}$$

The problem $(\mathfrak{p})$ becomes then equivalent to

$$\begin{aligned} \min \quad & f_0(x) + \sum_{i=1}^{m} \mathcal{I}(f_i(x)) \\ s.t. \quad & A_j \cdot x - b_j = 0, \quad j = 1, \dots, p \qquad (\mathfrak{p}_{\mathcal{I}}) \end{aligned}$$

we can approximate the indicator function $\mathcal{I}$ **smoothly** using a sequence of **logarithmic barriers**:

$$\varphi_t : \mathbb{R} \to \mathbb{R}, \quad u \mapsto \begin{cases} -\frac{1}{t}\log(-u) & \text{if } u < 0 \\ +\infty & \text{otherwise} \end{cases}$$

As t increases, $\varphi_t(u)$ remains close to 0 for a fixed $u < 0$; as u gets close to 0 (from the left), $\varphi_t(u)$ diverges to $+\infty$ for any arbitrarily big fixed t. Let

$$\phi_t(x) = \sum_{i=1}^{m} \varphi_t(f_i(x)) = -\frac{1}{t}\sum_{i=1}^{m} \log(-f_i(x))$$

Logarithmic barrier approximation

The idea is to approximate $\mathfrak{p}^\star$ using the sequence $\mathfrak{p}_t^\star$ $(t > 0)$:

$$\begin{aligned} \min \quad & f_0(x) + \phi_t(x) \\ s.t. \quad & A_j \cdot x - b_j = 0, \quad j = 1, \dots, p \qquad (\mathfrak{p}_t) \end{aligned}$$

Fix a positive t.

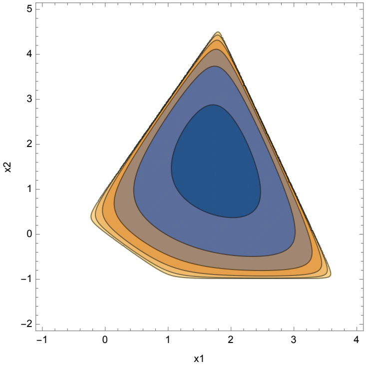
$$\phi_t(x) = -\frac{1}{t}\sum_{i=1}^{m} \log(-f_i(x)), \quad \mathrm{dom}_t\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

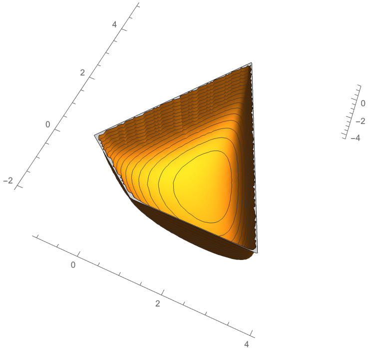
- ϕ_t is **convex** as a function of x (composition rule applied to φ_t and f_i)
- ϕ_t **twice continuously differentiable** (with respect to x)

$$\nabla\phi_t(x) = \sum_{i=1}^{m} \frac{1}{-tf_i(x)}\nabla f_i(x)$$

$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{-tf_i(x)^2}\nabla f_i(x)\nabla f_i(x)^t + \frac{1}{t}\sum_{i=1}^{m} \frac{1}{-tf_i(x)}\nabla^2 f_i(x)$$

$$\phi(x) = -\log(-(-x_1 - x_2)) - \log(-(-2x1 + x2 - 1))$$
$$- \log(-(3x1 + x2 - 10)) - \log(x2 + 1)$$





Since $\mathfrak{p}$ satisfies Slater's condition, so does $\mathfrak{p}_t$ for any $t > 0$: strong duality holds ($\mathfrak{d}_t^\star = \mathfrak{p}_t^\star < +\infty$).

KKT conditions

Fix $t > 0$. $x^*(t)$ is an optimum for $(\mathfrak{p}_t)$ if and only if

- $x^*(t) \in \mathrm{dom}\phi_t$
- $A_j \cdot x^*(t) - b_j = 0,\ j = 1, \ldots, p$
- $\nabla_x L_t(x^*(t), \mu(t)) = 0$

Observe that, by construction, the system has **no complementarity conditions** since the feasible set of $(\mathfrak{p}_t)$ has no inequality constraints.

$$\nabla_x L(x^*, \lambda, \mu) = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^*) + \sum_{j=1}^{p} \mu_j A_j = 0$$

For $x \in \mathrm{dom}\phi_t$:

$$\begin{aligned}
\nabla_x L_t(x^*(t), \mu(t)) &= \nabla f_0(x^*(t)) + \nabla \phi_t(x^*(t)) + \sum_{j=1}^{p} \mu_j(t) A_j \\
&= \nabla f_0(x^*(t)) + \sum_{i=1}^{m} \underbrace{\frac{1}{-t f_i(x^*(t))}}_{\lambda_i(t)} \nabla f_i(x^*(t)) + \sum_{j=1}^{p} \mu_j(t) A_j \\
&= 0
\end{aligned}$$

$\lambda_i(t)$ and $\mu_j(t)$ seem to be natural candidates for λ_i and μ_j respectively.

Checking KKT conditions of $\mathfrak{p}$

Consider $(x^*(t), \lambda(t), \mu(t))$ as potential candidates for (x^*, λ, μ). We need to check whether they satisfy the KKT conditions of $\mathfrak{p}$.

- $A_j \cdot x^*(t) - b_j = 0$ holds thanks to the primal feasibility of $x^*(t)$ as an optimal solution of $\mathfrak{p}_t$
- $f_i(x^*(t)) \leq 0$ holds thanks to the strong duality of $\mathfrak{p}_t$, in particular $\mathfrak{p}_t^\star < +\infty$
- $0 \leq \lambda_i^*(t)$ holds by definition (recall that $t > 0$)
- $\nabla_x L(x^*(t), \lambda(t), \mu(t)) = 0$ holds also by definition of $\lambda(t)$ and $\mu(t)$

Only the complementarity is missing and we have

$$-\lambda_i(t) f_i(x^*(t)) = \frac{1}{t}, \quad i = 1 \ldots, m$$

As t increases the product tends towards zero, fulfilling the complementarity at infinity.

$$\mathfrak{d}^\star = g(\lambda, \mu) = \inf_x L(x, \lambda, \mu) = f_0(x^*) = \mathfrak{p}^\star$$

$$\mathfrak{d}_t^\star = g_t(\mu(t)) = \inf_x L_t(x, \mu(t)) = f_0(x^*(t)) + \phi_t(x^*(t)) = \mathfrak{p}_t^\star$$

$$\begin{aligned} L(x^*(t), \lambda(t), \mu(t)) &= f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-t} + \sum_{j=1}^{p} \mu_j(t)(A_j \cdot x^*(t) - b_j) \\ &= f_0(x^*(t)) - \frac{m}{t} \end{aligned}$$

$$\begin{aligned} f_0(x^*(t)) \geq \mathfrak{p}^\star = \mathfrak{d}^\star \geq g(\lambda(t), \mu(t)) &= \inf_x L(x, \lambda(t), \mu(t)) \\ &=? L(x^*(t), \lambda(t), \mu(t)) \\ &= f_0(x^*(t)) - \frac{m}{t} \end{aligned}$$

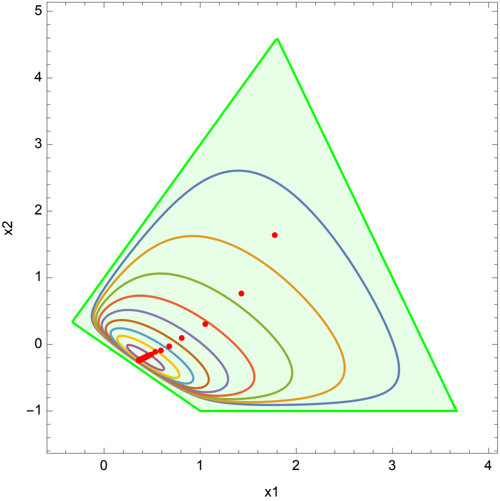
Start with a strictly feasible x, $t > 0$, $\alpha > 1$, and $\epsilon > 0$

1. Numerically compute $x^*(t)$ by solving the KKT conditions for $\mathfrak{p}_t$ (Newton-based techniques)
2. Update: $x \leftarrow x^\star(t)$
3. If $\frac{m}{t} < \epsilon$, halt (Stopping criterion)
4. Otherwise, increase $t \leftarrow \alpha t$ and repeat

- Halts with $f_0(x^*(\bar{t})) \sim \mathfrak{p}^\star \pm \epsilon$
- Several heuristics exist for the choice of α and the initial t

Central path: $\{x^\star(t) \mid t > 0\}$

$$\min x_1 + x_2 + \frac{1}{t}\phi(x)$$



Linear programming

$$\begin{aligned} \min \quad & c \cdot x \\ s.t. \quad & A_j \cdot x = b_j, \quad (\mathfrak{p}) \\ & 1 \leq j \leq p \\ & x \in \mathbb{R}^n_+ \end{aligned}$$

$$\begin{aligned} \max \quad & -b \cdot \mu \\ s.t. \quad & A_i^t \cdot \mu + c_i \geq 0 \quad (\mathfrak{d}) \\ & 1 \leq i \leq n \end{aligned}$$

Semidefinite programming

$$\begin{aligned} \min \quad & C \cdot X \\ s.t. \quad & A_j \cdot X = b_j, \quad (\mathfrak{p}) \\ & 1 \leq j \leq p \\ & X \in \mathcal{S}^n_+ \end{aligned}$$

- $\mathcal{S}^n$: set of $n \times n$ **symmetric matrices**
- $C, A_j \in S^n$, $b_j \in \mathbb{R}$, $1 \leq j \leq p$
- $\mathcal{S}^n_+$: **positive semidefinite** matrices
- $X \in \mathcal{S}^n_+$ also denoted as $X \succeq 0$
- $(\cdot)$: **Frobenius inner product** over $\mathcal{S}^n$
- $A \cdot B = \mathrm{tr}(A^t B)$ (tr for the trace)

SDP generalizes LP in the following sense: instead of linear combinations of real variables (x_i), $1 \leq i \leq n$, seen as coordinates of one vector x, SDP allows **linear combinations of inner products** $(X_i \cdot X_j)$, $1 \leq i, j \leq n$, seen as components of one symmetric matrix X (where $X_1, \ldots, X_n$ are vectors of $\mathbb{R}^n$).

Two equivalent definitions for $M \in \mathcal{S}^n$ to be **positive semidefinite**:

(i) M **is a Gramian matrix**: $\exists u \in \mathbb{R}^n. \quad M = uu^t$

(ii) **Non negative quadratic form**: $\forall v \in \mathbb{R}^n. \quad v \cdot Mv = M \cdot vv^t \geq 0$

The Frobenius inner product has a related norm:

$$\|M\|^2 = M \cdot M = \sum_{1 \leq i,j \leq n} m_{i,j}^2$$

Let $X, M \in \mathcal{S}^n$, then

$$\inf_X X \cdot M = \begin{cases} 0 & \text{if } M = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- If $M \succ 0$ or $M \prec 0$, then take $X = -tM$. Then, $X \cdot M = -t\|M\|^2$ and make t goes towards $+\infty$
- If M is undefinite, then there exists $v \in \mathbb{R}^n$ such that $v \cdot Mv < 0$. Then take $X = tvv^t$, thus:

$$M \cdot X = M \cdot (tvv^t) = t(v \cdot Mv) < 0,$$

  and make t goes towards $+\infty$.

So the only choice left is $M = 0$, in which case the inf is trivial.

Lagrangian ($\Lambda \in \mathcal{S}^n_+$)

$$L(X, \Lambda, \mu) = C \cdot X + \Lambda \cdot (-X) + \sum_{j=1}^{p} \mu_j (A_j \cdot X - b_j)$$

$$g(\Lambda, \mu) = \inf_{X \in \mathcal{S}^n} L(X, \Lambda, \mu) = -b \cdot \mu + \inf_{X \in \mathcal{S}^n} X \cdot \left(C - \Lambda + \sum_{j=1}^{p} \mu_j A_j \right)$$

$$\begin{aligned} \max \quad & -b \cdot \mu \\ s.t. \quad & C - \Lambda + \sum_{j=1}^{p} \mu_j A_j = 0, \quad (\mathfrak{d}) \\ & \Lambda \succeq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & -b \cdot \mu \\ s.t. \quad & C + \sum_{j=1}^{p} \mu_j A_j \succeq 0, \quad (\mathfrak{d}) \end{aligned}$$

Linear Matrix Inequality

- SDP is a convex problem
- Strong duality holds under Slater's condition
- $\nabla_X C \cdot X = C$

X^* satisfy the KKT conditions for the primal SDP if and only if there exists $\Lambda \in \mathcal{S}^n$, $\mu \in \mathbb{R}^p$ such that:

1. Primal feasibility: $A_j \cdot X^* = b_j$, $1 \leq j \leq p$
2. Primal feasibility: $X^* \succeq 0$
3. Dual feasibility: $\Lambda \succeq 0$
4. Complementarity: $\Lambda \cdot X^* = 0$
5. Stationarity: $\nabla_X L(X^*, \Lambda, \mu) = C - \Lambda + \sum_{j=1}^{p} \mu_j A_j = 0$

Logarithmic barrier for the positive orthant of $\mathbb{R}^n$

For $x > 0$: $\phi(x) = -\sum_{i=1}^{n} \log(x_i)$

Logarithmic barrier for the positive orthant of $\mathcal{S}^n$

For $X \succ 0$: $\phi(X) = -\log(\det X)$

Central path

$\{X^*(t) \mid t > 0\}$, where $x^*(t)$ is the optimum of the following parametric convex problem:

$$\begin{aligned} \min \quad & C \cdot X + \frac{1}{t}\phi(X) \\ s.t. \quad & A_j \cdot X - b_j = 0, 1 \leq j \leq p \quad (\mathfrak{p}_t) \end{aligned}$$

$$\begin{aligned} \min \quad & f_0(x) \\ s.t. \quad & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & A_j \cdot x - b_j = 0, j = 1, \ldots, p \end{aligned}$$

- x in a vector space V equipped with an inner product
- $f_0 : V \to \mathbb{R}$ convex and real valued
- $f_i : V \to V$, $i = 1, \ldots, m$, convex
- $f_i(x) \preceq_{K_i} 0$ means that $-f_i(x) \in K_i$ for some proper cone K_i of V
- f_0, $f_1, \ldots, f_m$ twice continuously differentiable (possibly in a weak sense)
- $A_j \in V$, $b_j \in \mathbb{R}$
- Under Slater's condition strong duality holds

$\phi : V \to \mathbb{R}$ is a generalized logarithm for the proper cone $K \subseteq V$ if:

- ϕ is defined over the interior of K
- $\nabla^2\phi(x) \prec_{\mathcal{S}^n_+} 0$ for $0 \prec_K x$
- $\phi(sx) = \phi(x) + rlog(s)$ for all $0 \prec_K x$ and $s > 0$
- r is the degree of ϕ

Examples:

- $K = \mathbb{R}_+$, $\phi(x) = \log(x)$ (classical logarithm)
- $K = \mathbb{R}^n_+$, $\phi(x) = \sum_{i=1}^n log(x_i)$ $(r = n)$
- $K = S^n_+$: $\phi(x) = \log(\det x)$ $(r = n)$

Observe that $-\phi$ is convex (ϕ is concave)

Solvers

- Matlab packages: SeDuMi, SDPT3
- Open source: CSDP

Environment

- Matlab software: CVX, YALMIP, SoSTools
- Open source: coin-or.org

Let $h \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial over the reals.
h is **non-negative** if and only if $\forall x \in \mathbb{R}^n.\ h(x) \geq 0$

Sum-of-squares (SoS)

A polynomial h is a sum of squares if and only if there exists polynomials g_i, $1 \leq i \leq m$, such that:

$$h = \sum_{i=1}^{m} g_i^2$$

A SoS polynomial is necessarily non-negative. The converse does not hold in general (Motzkin polynomial):

$$h(x_1, x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$$

h is non-negative and is not a SoS.

Take a polynomial $h \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $\leq 2d$.

- We can write h as a **scalar product** $H \cdot X$
- H is a symmetric matrix (not unique)
- X is symmetric and **semidefinite positive** (not unique)

X can be seen as a Gramian matrix formed as the (matrix) product of the vector χ and its transpose, where χ denote a vector of monomials of n variables of total degree less than d.

Example

$$x_1^4 - x_1^2 x_2^2 + x_2^4 = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{H} \cdot \left(\underbrace{\begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}}_{\chi} \begin{pmatrix} x_1^2 & x_1 x_2 & x_2^2 \end{pmatrix} \right)$$

Proposition

A polynomial h is SoS if and only if $H \succeq 0$.

Proof. If $H \succeq 0$ then there exists a matrix U such that $H = U^t U$. Thus

$$h = H \cdot X = (U^t U) \cdot (\chi \chi^t) = (U \chi) \cdot (U \chi) = \|U \chi\|^2 \ .$$

If h is SoS, then there exist a list of polynomials g_i such that $h = \sum_i g_i^2$. The monomials vector χ is then formed by all the (distinct) monomials appearing in all the g_i. The rows of the matrix U are formed by the coefficients of the polynomials g_i.

Example (cont'd)

$$x_1^4 - x_1^2x_2^2 + x_2^4 = \underbrace{\begin{pmatrix} 1 & 0 & \mu_1 \\ 0 & -2\mu_1 - 1 & 0 \\ \mu_1 & 0 & 1 \end{pmatrix}}_{H} \cdot \underbrace{\begin{pmatrix} x_1^4 & x_1^3x_2 & x_1^2x_2^2 \\ x_1^3x_2 & x_1^2x_2^2 & x_1x_2^3 \\ x_1^2x_2^2 & x_1x_2^3 & x_2^4 \end{pmatrix}}_{X}$$

Thus, h is SoS if and only if

$$\exists \mu_1. \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \mu_1 \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \succeq 0$$

which is an **LMI** problem: **dual feasibility** of the a **dual SDP** problem.

h is SoS is equivalent to solving the following dual SDP problem:

$$\begin{aligned} \max \quad & 0 \\ s.t. \quad & \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \mu_1 \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \succeq 0 \quad (\mathfrak{d}) \end{aligned}$$

For a fixed degree d, the size of χ is

$$\binom{n+d}{d}$$

The size of the unknown vector of the LMI reformulation is

$$\frac{1}{2}\binom{n+d}{d}\left(\binom{n+d}{d}+1\right)-\binom{n+2d}{2d}$$

The choice of the monomials list is important:

$$x_1^4 - x_1^2 x_2^2 + x_2^4 = \underbrace{\begin{pmatrix} 1 & 0 & \mu_1 \\ 0 & -2\mu_1 - 1 & 0 \\ \mu_1 & 0 & 1 \end{pmatrix}}_{H} \cdot \left(\underbrace{\begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}}_{\chi} \begin{pmatrix} x_1^2 & x_1 x_2 & x_2^2 \end{pmatrix} \right)$$

$$= \underbrace{\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}}_{H'} \cdot \left(\underbrace{\begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}}_{\chi'} \begin{pmatrix} x_1^2 & x_2^2 \end{pmatrix} \right)$$

$$\begin{aligned} \min \quad & p(x) \\ s.t. \quad & h_j(x) = 0, \\ & \quad 1 \leq j \leq p \end{aligned}$$

- **non-convex**
- size of x: n

$$\begin{aligned} \min \quad & C \cdot X \\ s.t. \quad & A_j \cdot X = b_j, \quad (\mathfrak{p}) \\ & \quad 1 \leq j \leq p \\ & X \in \mathcal{S}_+^n \end{aligned}$$

- **convex**
- size of X: $\binom{n+d}{d} \times \binom{n+d}{d}$
- $\mathfrak{p}^\star \leq p(x)$

Lasserre hierarchy

Increasing d gives tighter and tighter approximations for the optimal value of the original non-convex problem.

Max-cut problem

Let $G = (V, E)$ be a graph. The max-cut problem is the following **discrete** optimization problem

$$\begin{aligned} \max \quad & \sum_{(i,j)\in E} \frac{1 - v_i v_j}{2} \\ s.t. \quad & v_i = \{-1, 1\} \quad (v_i \in V) \end{aligned}$$

SDP relaxation (Goemans and Williamson 95)

v_i are now considered vectors, and $v_i v_j$ becomes $v_i \cdot v_j$. Let $X = vv^t$.

$$\begin{aligned} -\min \quad & C \cdot X \\ s.t. \quad & \mathrm{diag}(X) = 1, \quad (\mathfrak{p}) \\ & X \succeq 0 \end{aligned}$$

- *Convex Optimization*. Stephen Boyd and Lieven Vandenberghe. Cambridge University Press
- *Recherche Opérationnelle: aspects mathématiques et applications*. Frédéric Bonnans and Stéphane Gaubert. Ecole Polytechnique
- EE364A (Stephen Boyd, Stanford), EE236B (UCLA), Convex Optimization
  - www.stanford.edu/class/ee364a
  - www.ee.ucla.edu/ee236b/