Optimization, AIMS CDT

Francesco Capuano

3rd November, 2025

1 Convex sets, functions and problems

1.1 An Optimization Problem

$$\min_{x\in\mathbb{R}^2} \; x_1^2 + x_2^2 \quad \text{s.t.} \quad f_1(x) = \frac{x_1}{1+x_2^2} \leq 0, \;\; h_1(x) = (x_1+x_2)^2 = 0.$$

(a) Nonconvex formulation. For an optimization problem to be *convex*, the objective function and constraints must be convex as well. The function $f(x_1, x_2) = x_1^2 + x_2^2$ is convex everywhere, as the Hessian of f, $H_f(x) = 2I$). The equality constraint $h_1(x) = (x_1 + x_2)^2$ is a convex function as well, as its Hessian $H_{h_1}(x) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$ is positive semi-definite ($\lambda_1 = 0, \lambda_2 = 4$). Conversely, $H_{f_1}(x) = \begin{pmatrix} 0 & -\dfrac{2x_2}{(x_2^2+1)^2} \\ -\dfrac{2x_2}{(x_2^2+1)^2} & \cdots \end{pmatrix}$ is not positive semi-definite. Indeed, the determinant is

$$\det H_{f_1}(x) = -\frac{4x_2^2}{(x_2^2+1)^2} \leq 0 \, \forall x \in \mathbb{R}^2,$$

which in turn makes the whole problem non-convex.

(b) Feasible set & equivalent convex problem. As $1 + x_2^2 > 0 \, \forall x \in \mathbb{R}^2$, the inequality $\frac{x_1}{(1+x_2^2)} \leq 0$ coincides with $x_1 \leq 0$. Moreover, $(x_1 + x_2)^2 = 0$ is equivalent to the affine equality $x_1 + x_2 = 0$. Thus the feasible set is

$$\mathcal{X} = \{x \in \mathbb{R}^2 \mid x_1 \leq 0, \; x_1 + x_2 = 0\} \equiv \{t \in \mathbb{R}^+ : (-t, t)\}.$$

On $\mathcal{X}$ the objective function coincides with $x_1^2 + x_2^2 = 2t^2$, which is minimized at $t^\star = 0$, i.e. $x^\star = (0, 0)$. A *convex* reformulation that is equivalent to the original problem therefore is:

$$\min_{x\in\mathbb{R}^2} \; x_1^2 + x_2^2 \quad \text{s.t.} \quad x_1 + x_2 = 0, \;\; x_1 \leq 0,$$

or, in one-dimension only, $\min_{t\geq 0} t^2$. Both formulations are convex.

1.2 Hyperbolic constraints

For $x \in \mathbb{R}^n$, $y, z \in \mathbb{R}$ with $y \geq 0$, $z \geq 0$, prove that:

$$\| \begin{pmatrix} 2x, & y - z \end{pmatrix}^\top \|_2^2 \leq (y+z)^2 \tag{1}$$

Proof. Expanding the right-hand side,

$$\| \begin{pmatrix} 2x, & y - z \end{pmatrix}^\top \|_2^2 \leq (y+z)^2 \tag{2}$$

$$\begin{pmatrix} 2x, & y - z \end{pmatrix} \begin{pmatrix} 2x, \\ y - z \end{pmatrix} \leq (y+z)^2 \tag{3}$$

$$4x^\top x + (y-z)^2 \leq y^2 + 2yz + z^2 \tag{4}$$

$$4x^\top x \leq 4yz \tag{5}$$

$$x^\top x \leq yz \tag{6}$$

□

(a) Maximizing the harmonic mean. Let $t_i \geq (a_i^\top x - b_i)^{-1} > 0$ (from $\mathcal{X} = \{x \in \mathbb{R}^n : Ax > b\}$). Then, the optimization problem can be rewritten as:

$$\begin{aligned} &\min_t \mathbf{1}^\top t \\ &t_i \geq (a_i^\top x - b_i)^{-1}, \quad \forall i \end{aligned} \tag{7}$$

The constraint:

$$t_i(a_i^\top x + b_i) \geq 1, \qquad t_i \geq 0, \quad a_i^\top x + b_i > 0,$$

is referred to as *hyperbolic* because, if we let $u = a_i^\top x + b_i$, $v = t_i$, the *equality* $uv = 1$ defines a hyperbola in the (u, v)-plane, and the feasible set:

$$\mathcal{X} = \{(u, v) \mid uv \geq 1\}$$

corresponds to the region on (or above) the hyperbola, thus the name *hyperbolic constraint.*

Clearly, $\mathcal{X}$ is *not* convex—the segment between two points on opposite branches do not satisfy $uv \geq 1$. However, one can handle hyperbolic constraints (i.e., hyperbolic feasible sets) within a convex optimization problem by rewriting them in terms of a *second-order cone* (SOC), representing a convex set in a higher-dimensional space.

From eq. 1, by choosing $w = 2$, we obtain the equivalence

$$uv \geq 1 \quad \Longleftrightarrow \quad \left\| \begin{bmatrix} 2 \\ u - v \end{bmatrix} \right\|_2 \leq u + v, \quad u, v \geq 0.$$

Substituting back $u = a_i^\top x + b_i$ and $v = t_i$, the hyperbolic constraint can thus be expressed as the SOC (convex) constraint

$$\left\| \begin{bmatrix} 2 \\ a_i^\top x + b_i - t_i \end{bmatrix} \right\|_2 \leq a_i^\top x + b_i + t_i, \qquad t_i \geq 0, \quad a_i^\top x + b_i \geq 0.$$

This result allows to tackle problem 7 using convex optimization, as both the objective function and constraints are convex.

3. Support functions

Let $S_C(y) := \sup\{y^\top x | x \in C\}$ (possibly $+\infty$).

(a) Convexity of S_C. The support function of a set $C \subseteq \mathbb{R}^n$ is

$$S_C(y) = \sup_{x \in C} y^\top x.$$

Each function $f_x(y) = y^\top x$ is linear (hence convex) in y, and S_C is their pointwise supremum:

$$S_C(y) = \sup_{x \in C} f_x(y).$$

Since the supremum of convex functions is convex (§3.2.3, B&V), S_C is in turn convex for any set C.

(b) $S_C = S_{\text{conv}(C)}$. Clearly, $\text{conv}(C) \supseteq C$, which results in $S_{\text{conv}(C)} \geq S_C$, in keeping with the geometric interpretation of S_C. Conversely, for any given $x \in \text{conv}(C) : x = \sum_i \theta_i x_i$ ($\theta_i \geq 0$, $\sum \theta_i = 1$, $x_i \in C$),

$$y^\top x = \sum_i \theta_i \, y^\top x_i \leq \sum_i \theta_i \, S_C(y) \leq S_C(y).$$

The equality holds for any $x \in \text{conv}(C)$, and thus it must hold for the supremum over $\text{conv}(C)$ too, yielding $S_{\text{conv}(C)} \gtreqless S_C$.

4. Largest-L norm

For $x \in \mathbb{R}^n$, sort $|x|$ in non-increasing order, and define $\|x\|_{[L]} = \sum_{i=1}^{L} |x|_{[i]}$.

(a) Convexity. The largest-L norm can alternatively be represented as:

$$\|x\|_{[L]} = \max_y \Big\{ y^\top x \ : \ \|y\|_\infty \leq 1, \ \|y\|_1 \leq L \Big\}, \tag{8}$$

similarly to how support functions are represented too. The first constraint on y ensures the norm computed solving $\max_y y^\top x$ is not larger than the real norm, while the second ensures it is not smaller. Being a (support) linear function defined on a convex set—intersecting sets preserves convexity, and "all norms on $\mathbb{R}^n$ are convex functions", §3.1.5, B&V)—$\|x\|_{[L]}$ is convex.

(b) Integer programming (IP) formulation. In keeping with the result just presented, one can write the largest-L norm as an integer programming (i.e., combinatorial optimization) problem, where the variables $z_i \in \{0, 1\}$ select which indices in x to use to form the norm. This results in:

$$\|x\|_{[L]} = \max_z \Big\{ \sum_{i=1}^{n} |x_i| \, z_i \ : \sum_{i=1}^{n} z_i = L, \ z_i \in \{0, 1\} \Big\}.$$

(c) Linear programming formulation. Given in (a).

2 Duality

1. Projection onto the ℓ_1-ball

$$\min_{x} \ \tfrac{1}{2}\|x-a\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \leq c. \tag{9}$$

(a) Derive the dual The dual problem to 9 can be defined starting from the Lagrangian function associated with the problem, $\mathcal{L}(x,\lambda) = f(x) + \lambda g(x)$, for problems written in standard form (i.e., using only $g(x) \leq 0$ and $h(x) = 0$ constraints.

In this particular case, the Lagrangian function is:

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|x-a\|_2^2 + \lambda(\|x\|_1 - c), \tag{10}$$

which can be rewritten in terms of the individual x_i variables, yielding

$$\mathcal{L}(x,\lambda) = \sum_{i=1}^{n} \Bigg(\underbrace{\frac{1}{2}(x_i - a_i)^2 + \lambda|x_i|}_{l_i(x,\lambda)} \Bigg) - \lambda c. \tag{11}$$

Therefore, obtaining the dual function $g(\lambda) = \inf_x \mathcal{L}(x,\lambda)$ corresponds to minimizing each individual $l_i(x,\lambda)$, resulting in:

$$\frac{\partial l_i(x,\lambda)}{\delta x} = x_i - a_i + \lambda \operatorname{sign}(x_i) = 0 \iff \begin{cases} x_i = a_i - \lambda, \ x_i > 0 \\ x_i = a_i + \lambda, \ x_i < 0 \end{cases}, \tag{12}$$

implying:

$$g(\lambda) = \begin{cases} \frac{1}{2}\lambda^2 + a_i - \lambda, & \lambda < a_i \\ \frac{1}{2}\lambda^2 - a_i - \lambda, & \lambda > a_i \end{cases} \tag{13}$$

With the dual problem being $\max_{\lambda \geq 0} g(\lambda)$.

(b) Retrieving x^* from λ^* One can use the KKT conditions to obtain the primal solution x^* straight from λ^*, and in particular use the KKT conditions as a set of conditions to be solved for x^* given the optimal dual λ^*. In particular,

$$\text{KKT}: \begin{cases} 0 \in x^* - a + \lambda^* \, \partial\|x\|_1 & \text{(stationarity, with sub-gradients)} \\ \|x\|_1 \leq c & \text{(primal feasibility)} \\ \lambda \geq 0 & \text{(dual feasibility)} \\ \lambda(\|x\|_1 - c) = 0 & \text{(complementary slackness)} \end{cases} \tag{14}$$

2. SVM duality

(a) SVM as a QP Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1,+1\}$, the hard-margin SVM primal problem is:

$$\begin{aligned} \min_{\mathbf{w},\, b} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1, \quad i = 1,\ldots,n. \end{aligned} \tag{15}$$

We seek to represent this problem as a Quadratic Problem (QP) of the type:

$$\begin{aligned} \min_{\mathbf{z}\in\mathbb{R}^{d+1}} \quad & \tfrac{1}{2}\,\mathbf{z}^\top Q\,\mathbf{z} + c^\top \mathbf{z} \\ \text{s.t.} \quad & A\,\mathbf{z} \le \mathbf{b}, \end{aligned}$$

Let us start by stacking both $\mathbf{w}$ and b variables, which results into:

$$\mathbf{z} \;=\; \begin{bmatrix} \mathbf{w} \\ b \end{bmatrix} \in \mathbb{R}^{d+1}.$$

Only $\mathbf{w}$ is penalized, hence

$$\frac{1}{2}\,\mathbf{z}^\top \underbrace{\begin{bmatrix} I_d & 0 \\ 0 & 0 \end{bmatrix}}_{Q} \mathbf{z} \;+\; \underbrace{\mathbf{0}^\top}_{c^\top} \mathbf{z}.$$

Each margin constraint is

$$y_i(\mathbf{w}^\top \mathbf{x}_i - b) \;\ge\; 1 \;\iff\; -\,y_i\,\mathbf{x}_i^\top\,\mathbf{w} \;+\; y_i\,b \;\le\; -1.$$

Representing all the n inequality constraints in matrix form:

$$A\,\mathbf{z} \;\le\; \mathbf{b},$$

one obtains

$$A \;=\; \left[\begin{array}{c|c} -y_1\mathbf{x}_1^\top & y_1 \\ \vdots & \vdots \\ -y_n\mathbf{x}_n^\top & y_n \end{array}\right] \;=\; \big[\, -\operatorname{diag}(\mathbf{y})\,X \;\big|\; \mathbf{y} \,\big] \in \mathbb{R}^{n\times(d+1)}, \qquad \mathbf{b} \;=\; -\mathbf{1} \in \mathbb{R}^n,$$

resulting in a QP problem with:

$$Q = \begin{bmatrix} I_d & 0 \\ 0 & 0 \end{bmatrix}, \quad c = \mathbf{0}, \quad A = \big[\, -\operatorname{diag}(\mathbf{y})\,X \;\big|\; \mathbf{y} \,\big], \quad \mathbf{b} = -\mathbf{1}.$$

(b) Dual of the QP formulation. Here, $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is a symmetric, positive semi-definite matrix by construction ($\operatorname{null}(\mathbf{Q}) \neq \{0\}$), $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$.

Introducing a vector of non-negative Lagrange multipliers (dual variables) $\lambda \in \mathbb{R}^m$ the Lagrangian function is:

$$\mathcal{L}(\mathbf{z},\lambda) = \frac{1}{2}\mathbf{z}^\top Q\mathbf{z} + c^\top\mathbf{z} + \lambda^\top(A\mathbf{z} - \mathbf{b}) = \frac{1}{2}\mathbf{z}^\top Q\mathbf{z} + \mathbf{z}^\top(c + A^\top\lambda) - \mathbf{b}^\top\lambda,$$

with the dual function $g(\lambda)$ defined as the infimum of the Lagrangian over $\mathbf{z} \in \mathbb{R}^{d+1}$. However, the singularity of $\mathbf{Q}$ implies that the minimum of $\mathcal{L}(\mathbf{z},\lambda)$ over $\mathbf{z}$ is only finite if the term $v = (c + A^\top\lambda)$ is orthogonal to every vector in the *null space* of $\mathbf{Q}$, $\operatorname{null}(\mathbf{Q})$.

Indeed, let $\mathbf{z}$ be any non-zero vector in the null-space of $\mathbf{Q}$, i.e. $(\mathbf{z} : \mathbf{Q}\mathbf{z} = 0)$. Consider $\mathcal{L}(\mathbf{z},\lambda)$ moving along the direction $\mathbf{z}$. Formally,

$$\mathcal{L}(\mathbf{x} + \alpha\mathbf{z},\lambda) = \frac{1}{2}(\mathbf{x} + \alpha\mathbf{z})^\top\mathbf{Q}(\mathbf{x} + \alpha\mathbf{z}) + (\mathbf{x} + \alpha\mathbf{z})^\top v - \mathbf{b}^\top\lambda$$

Expanding and using $\mathbf{z} \in \operatorname{null}(\mathbf{Q})$, we find the quadratic term to α to vanish:

$$\mathcal{L}(\mathbf{x} + \alpha\mathbf{z}, \lambda) = \mathcal{L}(x, \lambda) + \alpha(\mathbf{z}^\top v).$$

For the function to be bounded below (have a finite minimum), one must prevent the $\alpha(\mathbf{z}^\top v)$ term from driving the value to $-\infty$ as $\alpha \to \pm\infty$. This requires that $\mathbf{z}^\top v = 0$ for all $\mathbf{z} \in \text{null}(\mathbf{Q})$.

This condition is equivalent to v being orthogonal to the entire null-space of $\mathbf{Q}$, i.e.

$$c + A^\top \lambda \perp \text{null}(Q).$$

This is a necessary condition for λ to be a feasible point in the dual problem, i.e., for $g(\lambda) > -\infty$ (bounded).

In practice, orthogonality with respect to the null-space is converted into explicit linear equality constraints.

Having Z as the matrix whose columns span the nullspace of $\mathbf{Q}$, the condition $c + A^\top \lambda \perp \text{null}(Q)$ is enforced by requiring that v is orthogonal to all columns of Z. Formally,

$$Z^\top v = 0 \implies Z^\top(c + A^\top \lambda) = 0 \tag{16}$$

Rearranging gives the explicit linear equality constraints:

$$Z^\top A^\top \lambda = -Z^\top c.$$

As per $g(\lambda)$, it can be derived starting from $\nabla_{\mathbf{z}}\mathcal{L}(\mathbf{z}, \lambda)$. Since $\mathbf{Q}$ is singular, a solution $\mathbf{z}^*$ exists if and only if v is in the column space (range) of $\mathbf{Q}$, a condition which needs to also be satisfied for the minimum to exist.

Whenever the Lagrangian is bounded from below, then it can be minimized imposing stationarity of the first-order derivative, resulting in:

$$\nabla_{\mathbf{z}}\mathcal{L}(\mathbf{z}, \lambda) = \mathbf{Q}\mathbf{z} + c + A^\top \lambda = 0 \implies \mathbf{z}^* = -\mathbf{Q}^+(c - A^\top \lambda) \tag{17}$$

In turn, this result yields the dual function $g(\lambda)$ by substituting this optimal $\mathbf{z}^*$ back into the Lagrangian function:

$$g(\lambda) = \mathcal{L}(\mathbf{z}^*, \lambda) = \frac{1}{2} v^T \mathbf{Q}^+ \mathbf{Q} \mathbf{Q}^+ v - v^T \mathbf{Q}^+ v - \mathbf{b}^T \lambda \tag{18}$$

Since $\mathbf{Q}$ is symmetric, so is $\mathbf{Q}^+$, and using the pseudo-inverse property $\mathbf{Q}\mathbf{Q}^+ = I$ one can obtain the final expression for the dual objective function:

$$g(\lambda) = -\frac{1}{2} v^T \mathbf{Q}^+ v - b^T \lambda$$