Convex Optimization in Machine Learning and Computational Finance
Lecture 4: Optimality conditions Lagrangian Duality

Dr. Michel Baes, Pr. Patrick Cheridito

RiskLab / ETH Zürich

Quick recall of last week's lecture

- ▶ Duality is a clever way to play with constraints.
- ▶ When can we play optimally with these constraints?

Strong duality holds:

  - ▷ always when you can do anything
  - ▷ sometimes when you only take affine combinations of constraints.

- ▶ Another view on convex cones: generalized orders.
- ▶ Dual cones and $K^{**} = \mathrm{cl}(\mathrm{cone}(K))$.
- ▶ Farkas Lemma, Tucker-Key's Theorem
- ▶ No arbitrage iff there exists a risk-neutral measure

A brief recall on continuity

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

- The function f is *continuous in $x \in \text{dom} f$* if for every sequence $(x_i)_{i \geq 0} \subseteq \mathbb{R}^n$ converging to x, we have $\lim_{i\to\infty} f(x_i) = f(x)$.
- The function f is *continuous on its domain* if continuous in every $x \in \text{dom} f$.

Here, we call such f *continuous*. Optimizers have another term for a continuous f with $f(x) \to +\infty$ when $x \to \partial\text{dom}(f)$.

A continuous function with a closed domain is LSC

Proof: Let f be a continuous function and $(x_i, t_i)_{i \geq 0}$ be a sequence of epi$f$ converging to (x, t). As $\lim_{i\to\infty} x_i = x$ and $\lim_{i\to\infty} t_i = t$,

$$f(x) = \lim_{i\to\infty} f(x_i) \leq \lim_{i\to\infty} t_i = t.$$

A convex function f is continuous in $\text{int}\,\text{dom} f$.

Convex functions are continuous on the interior of their domain

Proof: (for finite dimension; the fact is also true in infinite dimension)

1. f is locally bounded on $\operatorname{int}\operatorname{dom} f$**:** let $x \in \operatorname{int}\operatorname{dom} f$, $\Delta := \operatorname{conv}(P)$ with $P := \{x_0, \dots, x_n\} \in\subset \operatorname{dom} f$ such that $x \in \operatorname{int}\Delta$, $M := \max_{0\le i\le n} f(x_i)$. Then $f(y) \le M$ for all $y \in \Delta$ by 4onvexity of f. Let $B := B(x,\epsilon) \subseteq \Delta$. If $y \in B$, then $2x-y \in B$, and $f(x) \le [f(2x-y)+f(y)]/2 \le [M+f(y)]/2$. Thus $m := 2f(x) - M \le f(y) \le M$ for every $y \in B$.

2. f is locally Lipschitz continuous on $\operatorname{int}\operatorname{dom} f$**:** i.e.

$$\forall x \in \operatorname{int}\operatorname{dom} f,\ \exists U \ni x \text{ open}, L > 0 : \forall y, z \in U\ |f(y) - f(z)| \le L||y - z||.$$

Take B as above and $U := B(x, \epsilon/2)$, $y \ne z \in U$ with $f(y) \ge f(z)$, and $y' := y + \epsilon/4 \cdot (y - z)/||y - z||$, so that $y = (1 - \lambda)z + \lambda y'$ with $\lambda \le 4||y - z||/\epsilon$. Now, with $\Gamma := \max\{m, M\}$

$$|f(y)-f(z)| \le |\lambda f(y')+(1-\lambda)f(z)-f(z)| = \lambda|f(y')-f(z)| \le 8\Gamma||y-z||/\epsilon.$$

3. Locally Lipschitz continuous functions are continuous. Let $(x_i)_{i\ge0} \subseteq U$ converging to $x \in U$. As $|f(x_i) - f(x)| \le L||x_i - x|| \to 0$, $f(x_i) \to f(x)$ and f is continuous in x. ■

A brief reminder on differentiability

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ continuous, $d \in \mathbb{R}^n$.

- The function f is *differentiable at $x \in \mathrm{dom} f$ in the direction d* if the following limit exists:

$$\nabla f(x)[d] = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

- The function f is *differentiable at $x \in \mathrm{dom} f$* if differentiable at x in every direction $d \in \mathrm{aff}(\mathrm{dom} f)$, and if $d \mapsto \nabla f(x)[d]$ is linear. (aka *Gâteaux differentiability*)
- The *gradient* of f at x is the vector $f'(x)$ satisfying $\langle f'(x), d\rangle = \nabla f(x)[d]$ for the dot scalar product. Note that $f'_i(x) = \lim_{t \to 0}(f(x + te_i) - f(x))/t$.
- f is *continuously differentiable* if $x \mapsto f'(x)$ is continuous.
- **Chain rule:** If $g(t) := f(x + th)$ for a differentiable f, then $g'(t) = \langle f'(x + th), h\rangle$.

The mean value theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ differentiable on its domain.

- **Mean value theorem**. Let $x \in \mathrm{dom} f$, $h \in \mathbb{R}^n$ with $[x, x+h] \in \mathrm{dom} f$. There is a $t \in]0,1[$ for which:

$$f(x+h) = f(x) + \langle f'(x+th), h \rangle.$$

Proof: follows from Rolle's Theorem: *Let $g : [a,b] \to \mathbb{R}$ differentiable such that $g(a) = g(b)$. Then there is a $c \in]a,b[$ such that $g'(c) = 0$.*

Proof of Rolle's Theorem:

Assume wlog that there is a $t \in]a,b[$ for which $g(t) > g(a)$.

Take $c := \arg\max\{g(t) : t \in [a,b]\} = \arg\max\{g(t) : t \in]a,b[\}$. Then:

$$\frac{g(c+t) - g(c)}{t} \leq 0 \text{ for } t \in]0, b-c] \text{ and } \frac{g(c+t) - g(c)}{t} \geq 0 \text{ for } t \in [a-c, 0[.$$

By differentiability of g, the limit exists when $t \to 0$, and $g'(c) = 0$. ■

Set now $g(t) := f(x+th) + t(f(x) - f(x+h))$, $a := 0$, $b := 1$. ■

First order optimality conditions

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ differentiable on its domain.

- **Mean value theorem**. Let $x \in \mathrm{dom} f$, $h \in \mathbb{R}^n$ with $[x, x+h] \in \mathrm{dom} f$. There is a $t \in]0,1[$ for which:

$$f(x+h) = f(x) + \langle f'(x+th), h \rangle.$$

- **Necessary optimality conditions**:

Suppose that f is continuously differentiable, and there exists a minimizer $x^* \in \mathrm{int}\,\mathrm{dom} f$. Then $f'(x^*) = 0$.

Proof: Assume $f'(x^*) \neq 0$. Let $g(t) := f(x^* - tf'(x^*))$, so that $g'(t) = -\langle f'(x^* - tf'(x^*)), f'(x^*) \rangle$. Then $g'(0) < 0$. By continuity of f', there exists a $T > 0$ such that $g'(t) < 0$ for all $t \in [0, T]$.
By the mean value theorem, there is a $t \in]0, T[$ such that:

$$f(x^* - Tf'(x^*)) - f(x^*) = -T\langle f'(x^* - tf'(x^*)), f'(x^*) \rangle = Tg'(t) < 0,$$

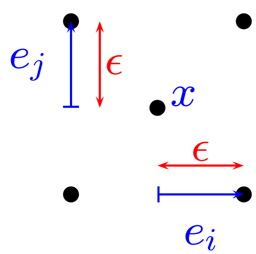
contradicting the minimality of $f(x^*)$.

Second-order differentiability

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ differentiable on its domain.

- If f_i' are themselves differentiable at x for each i,
  their gradients at x form the *Hessian* $f''(x)$ of f at x:
  $f''_{ij}(x) = \lim_{t\to 0}(f'_j(x + te_i) - f'_j(x))/t$ for all i, j.
- If $x \mapsto f''(x)$ is continuous,
  $f''(x)$ is **symmetric** for each $x \in \operatorname{int}\operatorname{dom} f$

Proof: Fix $1 \le i < j \le n$. Is $f''_{ij}(x) = f''_{ji}(x)$?



Let $x \in \operatorname{int}\operatorname{dom} f$ and $\epsilon > 0$ such that $x \pm \epsilon e_i \pm \epsilon e_j \in \operatorname{int}\operatorname{dom} f$. Let $F(s,t) := f(x + se_i + te_j)$ for $|s|, |t| \le \epsilon$. We will write $\Delta(s,t) := F(s,t) - F(0,t) - F(s,0) + F(0,0)$ in two ways. Denote $G_s(t) := F(s,t) - F(0,t)$. Then

$$\Delta(s,t) = G_s(t) - G_s(0) = G'_s(\tau)t = (f'_j(x+se_i+\tau e_j) - f'_j(x+\tau e_j))t,$$

and $\Delta(s,t) = f''_{ij}(x + \sigma e_i + \tau e_j)ts$ for a $|\tau| \in]0, |t|[$ and a $|\sigma| \in]0, |s|[$ (use twice the mean value thm.) Similarly, $\Delta(s,t) = f''_{ji}(x + \sigma' e_i + \tau' e_j)ts$ for a $|\tau'| \in]0, |t|[$ and a $|\sigma'| \in]0, |s|[$. Letting $\epsilon \to 0$, $\sigma, \tau, \sigma', \tau' \to 0$, and by continuity of $f''(x)$, $f''_{ij}(x) = \lim_{s,t\to 0} \Delta(s,t)/(st) = f''_{ji}(x)$.

Second-order conditions 1

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ twice continuously differentiable.

▶ **Taylor expansion theorem**. Let $x \in \operatorname{dom} f$, $h \in \mathbb{R}^n$ with $[x, x+h] \in \operatorname{dom} f$. There is a $t \in]0,1[$ for which:

$$f(x+h) = f(x) + \langle f'(x), h\rangle + \frac{1}{2}\langle f''(x+th)h, h\rangle.$$

Proof: Let $R := f(x+h) - f(x) - \langle f'(x), h\rangle$ and $g(t) := f(x+th) + (1-t)\langle f'(x+th), h\rangle + Rt^2$.
Then $g(1) - g(0) = 2R = g'(t)$ for a $t \in]0,1[$ by the mean value theorem.
Thus $g'(t) = (1-t)\langle f''(x+th)h, h\rangle + 2tR = 2R$. It remains to simplify.

▶ **More necessary optimality conditions**:
Suppose that $x^* \in \operatorname{int} \operatorname{dom} f$ is a local minimum of f.
Then $f'(x^*) = 0$, and $f''(x^*) \in \mathbb{S}^n_+$.

Indeed, if $\langle f''(x^*)h, h\rangle < 0$, we let $g(t) := f(x^* + th)$. As $g''(0) < 0$, there is a $T > 0$ such that $g''(t) < 0$ for all $t \in [0, T]$. Thus $g(T) = g(0) + g'(0)T + g''(t)T^2/2 = f(x^*) + g''(t)T^2/2 < f(x^*)$, contradiction.

Second-order conditions 2

We assume that f is twice continuously differentiable.

▶ **Sufficient optimality conditions**:

Suppose that $x^* \in \operatorname{int}\operatorname{dom} f$ satisfies $f'(x^*) = 0$, and $f''(x^*) \in \mathbb{S}^n_{++}$. Then x^* is a local minimum of f.

Assume that x^* is not a local minimizer of f.
For every $\epsilon > 0$, there is $x^*_\epsilon \in B(x^*, \epsilon)$ for which:

$$f(x^*) > f(x^*_\epsilon) = f(x^*) + \frac{1}{2}\langle f''(x^* + t_\epsilon(x^*_\epsilon - x^*))(x^*_\epsilon - x^*), x^*_\epsilon - x^*\rangle.$$

Thus $\lambda_{\min}(f''(x^* + t_\epsilon(x^*_\epsilon - x^*))) < 0$ for every $\epsilon > 0$.

By continuity of $\lambda_{\min}$ and of f'', we deduce that $\lambda_{\min}(f''(x^*)) \leq 0$,

contradicting the positive definiteness of $f''(x^*)$.

Lagrangian duality:

another view on strong duality

An alternative way to get conditions for strong duality

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g(x) \succeq_K b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x)\ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

The *Lagrangian* of the above problem is defined as:

$$L : X \times \mathcal{F} \to \mathbb{R}, \quad (x, F) \mapsto L(x, F) := f(x) + F(b) - F(g(x))$$

$$\text{If} \quad \sup_{F \in \mathcal{F}} \inf_{x \in X} L(x, F) = \inf_{x \in X} \sup_{F \in \mathcal{F}} L(x, F),$$

and under some conditions on $\mathcal{F}$, **strong duality holds**.
The proof spans 4 slides, and each step is important.

Looking between p^* and $d^*(\mathcal{F})$ # 1/4

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g(x) \succeq_K b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x),\ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

Step 1

$$\begin{array}{ll} \inf\limits_{x \in X} & f(x) \\ \text{s.t.} & g(x) \succeq_K b \end{array} \quad \geq \quad \begin{array}{ll} \inf\limits_{x \in X} & f(x) \\ \text{s.t.} & F(g(x)) \geq F(b) \\ & \forall F \in \mathcal{F} \end{array}$$

Equality satisfied if:

When $K = \mathbb{R}^m_+$, $\mathcal{F}$ contains $x \mapsto x_i$ for $1 \leq i \leq m$.

For a general K, $g(x) \succeq_K b$ iff $g(x) - b \in K$ iff $\langle u, g(x) - b \rangle \geq 0$ for all $u \in K^*$. Thus, it suffices that $\mathcal{F}$ contains all linear functions $x \mapsto \langle u, x \rangle$ with $u \in K^*$.

Looking between p^* and $d^*(\mathcal{F})$ # 2/4

Recall the definition of the *Lagrangian*:

$$L : X \times \mathcal{F} \to \mathbb{R}, \quad (x, F) \mapsto L(x, F) := f(x) + \underbrace{F(b) - F(g(x))}_{\text{make it nonpositive}}.$$

Step 2

$$\begin{array}{ll} \inf_{x \in X} & f(x) \\ \text{s.t.} & F(g(x)) \geq F(b) \\ & \forall F \in \mathcal{F} \end{array} \quad \geq \quad \inf_{x \in X} \sup_{F \in \mathcal{F}} \; L(x, F)$$

$$\text{Left-hand side} = \begin{array}{ll} \inf_{x \in X} & f(x) \\ \text{s.t.} & 0 \geq \sup_{F \in \mathcal{F}} F(b) - F(g(x)) \end{array}$$

Equality satisfied if: $F \in \mathcal{F} \Rightarrow \alpha F \in \mathcal{F}$ for all $\alpha \geq 0$.

If $F(b) > F(g(x))$ for an $x \in X$ and an $F \in \mathcal{F}$, then $L(x, \alpha F) \to \infty$ when $\alpha \to \infty$, and that x is not taken in the infimum.

Looking between p^* and $d^*(\mathcal{F})$ # 3/4

$$\begin{array}{ll} \text{inf} & f(x) \\ \text{s.t.} & g(x) \succeq_K b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \text{sup} & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x) \; \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

$$L(x, F) := f(x) + F(b) - F(g(x))$$

Step 3

$$\inf_{x \in X} \sup_{F \in \mathcal{F}} L(x, F) \geq \sup_{F \in \mathcal{F}} \inf_{x \in X} L(x, F)$$

Equality condition: there is a whole theory for it (*minimax theorems*), initiated by John von Neumann.

Theorem (Rockafellar, Cor. 37.3.2): The equality holds

if L is continuous, convex in x and concave in F,

$\mathcal{F}$ is finite-dimensional, convex and closed,

X is convex and if either X or $\mathcal{F}$ is compact.

We'll show another minimax statement later.

Looking between p^* and $d^*(\mathcal{F})$ # 4/4

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g(x) \succeq_K b \\ & x \in X \end{array} \geq \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x) \ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

$$L(x,F) := f(x) + F(b) - F(g(x))$$

Step 4

$$\sup_{F \in \mathcal{F}} \inf_{x \in X} \ L(x,F) \geq \begin{array}{ll} \sup\limits_{F \in \mathcal{F}} & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x) \ \forall x \in X \end{array}$$

Equality satisfied if: $F \in \mathcal{F}$, $a \in \mathbb{R}$ implies $a + F \in \mathcal{F}$.

Let $h(F) := \inf_{x \in X} L(x,F)$, and let $(F_i)_{i \geq 0} \subseteq \mathcal{F}$ be a *maximizing sequence*, i.e. $h(F_i) \uparrow \sup_{F \in \mathcal{F}} h(F)$. Set $a_i := \inf_{x \in X} f(x) - F_i(g(x))$. Then $h(F_i + a_i) = h(F_i)$, and $(F_i + a_i)_{i \geq 0}$ is also a maximizing sequence, but with $F_i(g(x)) + a_i \leq f(x) \ \forall x \in X$.

Strong duality – recap with $K = \mathbb{R}^m_+$

$$\begin{array}{ll} \text{inf} & f(x) \\ \text{s.t.} & g(x) \geq b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \text{sup} & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x)\ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

We have strong duality if:

$$\sup_{F \in \mathcal{F}} \inf_{x \in X} L(x, F) = \inf_{x \in X} \sup_{F \in \mathcal{F}} L(x, F),$$

and when $\mathcal{F}$ satisfies:

1. $x \mapsto 0$ and $x \mapsto x_i$ belong to $\mathcal{F}$ for all i.

2. $F \in \mathcal{F}$ and $\alpha \geq 0$ implies $\alpha F \in \mathcal{F}$.

3. $F \in \mathcal{F}$ and $a \in \mathbb{R}$ implies $a + F \in \mathcal{F}$.

In particular, $\mathcal{F} := \{x \mapsto \langle u, x \rangle + u_0 : u \geq 0\}$ satisfies **1**,**2**,**3**, where the scalar product is here the dot product.

Strong duality – recap with $K = \mathbb{R}^m_+$

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g(x) \geq b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x) \ \forall x \in X \\ & F \in \{u^T x + u_0 : u \geq 0\} \end{array}$$

The Lagrangian becomes: $L(x, u) = f(x) + u^T(b - g(x))$ (note the cancellation of u_0's).

Theorem: With such an $\mathcal{F}$, strong duality holds iff

$$\sup_{u \geq 0} \inf_{x \in X} L(x, u) = \inf_{x \in X} \sup_{u \geq 0} L(x, u).$$

Moreover, the above "sup-inf" equals the primal/dual optimal value.

What happens with the complementarity conditions?

Recall:

Theorem 1 (Complementarity conditions) *Suppose that x^* and F^* are feasible for their respective problems, and that $f(x^*) = F^*(b)$. Then*

$$p^* = f(x^*) = F^*(g(x^*)) = F^*(b) = d^*(\mathcal{F}).$$

Here $F^*(y) = \langle u^*, y \rangle + u_0$ for a $u^* \geq 0$ and a $u_0 \in \mathbb{R}$.
The condition $F^*(g(x^*)) = F^*(b)$ reads as:

$$\langle u^*, g(x^*) \rangle + u_0 = \langle u^*, b \rangle + u_0 \quad \Leftrightarrow \quad \sum_{i=1}^{m} u_i^*(g_i(x^*) - b_i) = 0$$

As $u_i^* \geq 0$ and $g_i(x^*) \geq b_i$, we have $u_i^*(g_i(x^*) - b_i) = 0 \ \forall\, i$.

Strong duality – recap with $K = \mathbb{R}^m_+$

With the usual order $K := \mathbb{R}^m_+$, we have strong duality if:

$$\sup_{F \in \mathcal{F}} \inf_{x \in X} L(x, F) = \inf_{x \in X} \sup_{F \in \mathcal{F}} L(x, F),$$

$$\text{and } \mathcal{F} := \{x \mapsto \langle u, x \rangle + u_0 : u \geq 0\}$$

The Lagrangian is then $L(x, u) = f(x) + \langle u, b - g(x) \rangle$.
Connection with the perturbation function of last week:

If $\langle u^*, x \rangle + u_0$ is the optimal dual function around b,
then u_i^* is the **price** (or the **sensitivity**)
of the constraint $g_i(x) \geq b_i$ around b.

Recall the electricity producer problem

In the long run, electricity producers might be submitted to CO_2 emission constraints. Legislators have to set fines.

Let x_c be the energy to be produced in coal plants [MWh]

x_g be the energy to be produced in natural gas plants [MWh]

x_n be the energy to be produced in nuclear plants [MWh]

d be the demand to satisfy [MWh]

E be the maximal CO_2 emission allowed [Tons]

c_i be the capacity of production of mean i [MWh]

The electricity producer problem

Recall the electricity producer problem of last week:

$$
\begin{array}{lrcrcrcll}
\text{min} & 22x_c & + & 75x_g & + & 17x_n & & & \\
\text{s.t.} & x_c & + & x_g & + & x_n & \geq & d & \\
 & -0.9x_c & - & 0.6x_g & & & \geq & -E & \color{red}{u_E} \\
 & -x_c & & & & & \geq & -c_c & \color{red}{u_c} \\
 & & & -x_g & & & \geq & -c_g & \color{red}{u_g} \\
 & & & & & -x_n & \geq & -c_n & \color{red}{u_n} \\
 & & & & & x_i & \geq & 0 &
\end{array}
$$

$\alpha \cdot u_E$ is by how much the cost decreases
if the allowed pollution level E goes to $E + \alpha$

$\rightarrow$ the tax level must be above u_E

$\alpha \cdot u_n$ is by how much the cost decreases
if the nuclear production c_n goes to $c_n + \alpha$

$\rightarrow$ the production increase cost must be below u_n/unit

Strong duality – recap with convex K

inf	$f(x)$
s.t.	$g(x) \succeq_K b$
	$x \in X$

$\geq$

sup	$F(b)$
s.t.	$F(g(x)) \leq f(x) \ \forall x \in X$
	$F \in \mathcal{F}$

We have strong duality if:

$$\sup_{F \in \mathcal{F}} \inf_{x \in X} L(x, F) = \inf_{x \in X} \sup_{F \in \mathcal{F}} L(x, F),$$

$$g(x) \succeq_K b \quad \Rightarrow \quad F(g(x)) \geq F(b) \quad \forall F \in \mathcal{F}$$

and when $\mathcal{F}$ satisfies:

1. $x \mapsto \langle u, x \rangle$ belong to $\mathcal{F}$ for all $u \in K^*$.

2. $F \in \mathcal{F}$ and $\alpha \geq 0$ implies $\alpha F \in \mathcal{F}$.

3. $F \in \mathcal{F}$ and $a \in \mathbb{R}$ implies $a + F \in \mathcal{F}$.

In particular, $\mathcal{F} := \{x \mapsto \langle u, x \rangle + u_0 : u \in K^*\}$ works for **1**,**2**,**3**.

A first interesting particular case

- $K = \{0\}$ (equality constraints).
  As $K^* = \mathbb{R}^m$, the Lagrangian is:
  $L(x, u) = f(x) + \langle u, b - g(x) \rangle$, and u is unconstrained.

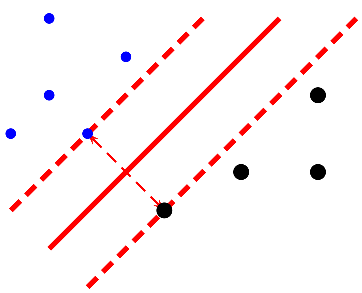
An application in classification

Replacing many constraints by a simple one

Support Vector Machines

We have points $a_1, \ldots, a_N$ in $\mathbb{R}^n$, belonging to two classes -1 and **1**: p_i :=Class(a_i). We want to separate them strictly with a hyperplane $H = \{x : \langle w, x \rangle + w_0 = 0\}$.

Wlog, it means that $p_i(\langle w, a_i \rangle + w_0) \geq 1$.



1. Distance from a point a_i to H:

$\min\{||x - a_i||_2^2 : \langle w, x \rangle + w_0 = 0\}$

$= \min\{||x - a_i||_2^2 : \pm\langle w, x \rangle \pm w_0 \geq 0\}$.

Here, $L(x, u) = ||x - a_i||_2^2 + u(\langle w, x \rangle + w_0)$;

let us compute $\max_u \min_x L(x, u)$:

using optimality conditions, we get

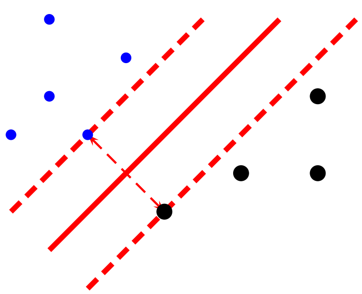
$2||x^* - a_i||_2 = |\langle w, a_i \rangle + w_0| / ||w||_2$.

An application in classification

Replacing many constraints by a simple one

Support Vector Machines

We have points $a_1, \ldots, a_N$ in $\mathbb{R}^n$, belonging to two classes -1 and 1: p_i :=Class(a_i). We want to separate them strictly with a hyperplane $H = \{x : \langle w, x\rangle + w_0 = 0\}$.
Wlog, it means that $p_i(\langle w, a_i\rangle + w_0) \geq 1$.



2. The best hyperplane:

$\max_{w,w_0} \min_i p_i(\langle w, a_i\rangle + w_0)/||w||_2$
$= \max_{w,w_0}\{1/||w||_2 : p_i(\langle w, a_i\rangle + w_0) \geq 1\}$,
or $\min_{w,w_0}\{||w||_2^2/2 : p_i(\langle w, a_i\rangle + w_0) \geq 1\}$.

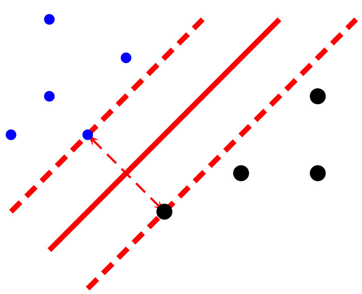
Here, $L(w, w_0, \alpha) = ||w||_2^2/2 - \sum_i \alpha_i(p_i(\langle w, a_i\rangle + w_0) - 1)$

An application in classification

Replacing many constraints by a simple one

Support Vector Machines

We have points $a_1, \ldots, a_N$ in $\mathbb{R}^n$, belonging to two classes -1 and 1: p_i :=Class(a_i). We want to separate them strictly with a hyperplane $H = \{x : \langle w, x\rangle + w_0 = 0\}$.
Wlog, it means that $p_i(\langle w, a_i\rangle + w_0) \geq 1$.



3. Playing with the Lagrangian:

$L(w, w_0, \alpha)$
$= ||w||_2^2/2 - \sum_i \alpha_i(p_i(\langle w, a_i\rangle + w_0) - 1)$

Again, we $\max_{\alpha\geq 0} \min_{w,w_0} L(w, w_0, \alpha)$:
then $\sum_i \alpha_i p_i = 0$ (As if not, the min is $-\infty$),

and we get $\max_{\alpha\geq 0}\{\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j p_i p_j \langle a_i, a_j\rangle : \sum_i \alpha_i p_i = 0\}$.

Finally, we get to convex things!

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g(x) \succeq_K b \\ & x \in X \end{array} \quad \geq \quad \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(g(x)) \leq f(x)\ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

From now on, f is convex, X is convex, and g *K-concave*:

$$g(\lambda x + (1 - \lambda)y) \succeq_K \lambda g(x) + (1 - \lambda)g(y) \quad \forall x, y,\ \forall \lambda \in [0, 1]$$

- Usual concavity is just $\mathbb{R}_+$-concavity.
- $\{0\}$-concavity (or convexity) is linearity.
- It's the **most general** convex setting in finite dimension.
- Usually, considering only linear g, or $K = \mathbb{R}^m_+$ is general enough for most applications. However, duality theory is powerful enough for allowing K-convexity.

Strong duality holds for this convex problem

Slater's condition holds when $\exists\, \hat{x} \in \text{int}\, X$
such that $\hat{x} \in \text{relintdom} f$ and $g(\hat{x}) - b \in \text{relint} K$

Theorem 2 *Under convexity and Slater's condition,*

$$p^* = \inf_{x\in X} \sup_{u\in K^*} f(x) + \langle u, b - g(x)\rangle \leq \sup_{u\in K^*} \inf_{x\in X} f(x) + \langle u, b - g(x)\rangle$$

Proof:

Find a $u^* \in K^*$ for which
$p^* \leq f(x) + \langle u^*, b - g(x)\rangle$ for all $x \in X$

Strong duality for convex problem holds

Here's the proof

Find a $u^* \in K^*$ for which
$p^* \leq f(x) + \langle u^*, b - g(x) \rangle$ for all $x \in X$

Let $\bar{C} = \{(b - g(x), f(x)) : x \in X\} + (K \times \mathbb{R}_+)$, $C = \text{relint}(\bar{C})$ and $D = \{(0, p^*)\}$. $\bar{C}$ is convex (short proof on the board — K-convexity plays a crucial role here). Also $C \cap D = \phi$. Otherwise, there exists $x \in X$ such that $b = g(x)$ and $f(x) < p^*$, impossible.

By the affine separation theorem, there exists a nonzero linear function $(s, t) \mapsto f(s, t) = \langle \lambda, s \rangle + \alpha t$ and a $\gamma \in \mathbb{R}$ such that $f(s, t) > \gamma$ for all $(s, t) \in C$ and $D \subseteq \{(s, t) : f(s, t) = \gamma\}$. Obviously $\gamma = \alpha p^*$ and $\lambda \in K^*$. If $\alpha \neq 0$, define $u^* := \lambda / \alpha$. Then $p^* \leq \langle u^*, b - g(x) \rangle + f(x)$ for any $x \in X$ and we're done. Suppose then that $\alpha = 0$, and let $\hat{x}$ be a Slater point: $g(\hat{x}) - b \in \text{relint} K$. Then $(b - g(\hat{x}), f(\hat{x}) + 1) \in \text{relint}\bar{C}$. Thus $\langle \lambda, b - g(\hat{x}) \rangle > 0$. However, since $\lambda \in K^*$ and $g(\hat{x}) - b \in K$, we must have $\langle \lambda, g(\hat{x}) - b \rangle \geq 0$, a contradiction.

Why convex optimization simplifies your life

- Strong duality holds with a very simple $\mathcal{F}$
  (just some affine functions)
  when Slater's condition holds
- The Lagrangian $f(x) + \langle u, b - g(x) \rangle$
  -is trivial to maximize in u for fixed x
  -is easy to minimize in x for fixed u by convexity

Also, we can rewrite problems in the standard *conic programming* format: (see blackboard – nice exercise for those who are not here)

$$\begin{array}{ll} \text{min} & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

The dual of conic programming problems

Let $\bar{K} = \{0\} \times K$; $\bar{K}^* = \mathbb{R}^m \times K^*$

$$\begin{array}{ll} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} = \begin{array}{ll} \min & \langle c, x\rangle \\ \text{s.t.} & \begin{pmatrix} Ax \\ x \end{pmatrix} \succeq_{\bar{K}} \begin{pmatrix} b \\ 0 \end{pmatrix} \end{array}$$

$$\geq \begin{array}{ll} \max & \langle u_{(1)}, b\rangle + u_0 \\ \text{s.t.} & \langle u_{(1)}, Ax\rangle + \langle u_{(2)}, x\rangle + u_0 \leq \langle c, x\rangle \ \forall x \in \mathbb{R}^n \\ & u_{(1)} \text{ free, } u_{(2)} \in K^* \end{array}$$

Note that $u_0 \leq 0$ (take $x = 0$) and if $u_0 < 0$, the max can be increased. So $u_0 = 0$. Denote by A^* the adjoint of A

$$= \begin{array}{ll} \max & \langle y, b\rangle \\ \text{s.t.} & A^* y + s = c \\ & s \in K^* \end{array}$$

The dual of the dual

Exercise: What is the dual of

$$\begin{array}{rl} \max & \langle y, b \rangle \\ \text{s.t.} & A^* y + s = c \\ & s \in K^* \end{array}$$

Recall that $K^{**} = K$ since K is closed and convex

And when you don't have Slater

Things can go very wrong !

We use Frobenius scalar product

$$\begin{array}{ll} \min & 2x_3 - 2x_4 \\ \text{s.t.} & x_3 + x_4 = 1, x_2 = 0 \\ & \begin{pmatrix} x_1 & x_4 & x_6 \\ x_4 & x_2 & x_5 \\ x_6 & x_5 & x_3 \end{pmatrix} \in \mathbb{S}^3_+ \end{array} = \begin{array}{ll} \min & 2 \\ \text{s.t.} & x_3 = 1 \\ & \begin{pmatrix} x_1 & 0 & x_6 \\ 0 & 0 & 0 \\ x_6 & 0 & 1 \end{pmatrix} \in \mathbb{S}^3_+ \end{array}$$

$$\begin{array}{ll} \max & y_1 \\ \text{s.t.} & \begin{pmatrix} 0 & -1 - y_1/2 & 0 \\ -1 - y_1/2 & -y_2 & 0 \\ 0 & 0 & 2 - y_1 \end{pmatrix} \in \mathbb{S}^3_+ \end{array}$$

$y_1 = -2$ (look at the 2nd principal minor),

and $p^* = 2 > d^* = -2$

What you have to keep in mind

$$\begin{array}{ll} \text{min} & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \quad \geq \quad \begin{array}{ll} \text{max} & \langle y, b\rangle \\ \text{s.t.} & A^*y + s = c \\ & s \in K^* \end{array}$$

- In case of strong duality, the optimal dual function describes the sensitivity of each constraint.
- When Slater condition holds, strong duality holds for conic programming with linear dual functions.
- Complementarity conditions $(f(x^*) = F^*(b))$ is an easy way of proving optimality.
- The dual cone K^* represents the nondecreasing linear functions for $\succeq_K$.
- (Still conjectural, but many are convinced about it) A problem is tractable iff it has an easy dual.

For next week

- (Sub)differentiability of convex functions
- Conjugate functions: relating duality and differentiability