Kernel Machines

Alain Celisse

SAMM

Paris 1-Panthéon Sorbonne University

alain.celisse@univ-paris1.fr

Lecture 3: Duality gap and KKT conditions

—

Master 2 Data Science – Centrale Lille, Lille University
Fall 2022

Outline of the lecture

- Convexity
- Fenchel-Legendre transform
- Lagrange Multipliers
- Duality gap
- KKT conditions

Convexity

Convex function

Definition (Convex set)

$D \subset \mathbb{R}^d$ is convex if

$$\forall x, y \in D, \forall \theta \in [0, 1], \qquad \theta x + (1 - \theta) y \in D$$

Definition (Convex function)

$f : \mathbb{R}^d \to \mathbb{R}$ is convex if

1. $Dom(f)$ is a convex set
2. For all $x, y \in Dom(f)$ and $\theta \in [0, 1]$

$$f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y)$$

Remark: : Strict convexity if strict ineq. for all $x \neq y \in Dom(f)$ and $\theta \notin \{0, 1\}$

Examples of convex functions

Examples

- Hyperplane: $f(x) = \alpha^\top x + b$, with $\alpha \in \mathbb{R}^d$ and $b \in \mathbb{R}$
- ℓ_p-Norm: $f(x) = \|x\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, with $p \geq 1$
- With A, X are a $n \times d$ matrix:
  - $f(X) = \mathrm{Tr}\left(A^\top X\right) + b$
  - Operator norm:

$$f(X) = \|X\|_{op} = \sup_{\|u\|_{\mathbb{R}^d} \leq 1} \|Xu\|_{\mathbb{R}^n}$$
$$= \sigma_{\max}(X) = \sqrt{\lambda_{\max}\left(X^\top X\right)}$$

Convexity characterization

Theorem

$f : \mathbb{R}^d \to \mathbb{R}$ is convex iff the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(t) = f(x + tv), \ \textit{and} \quad Dom(g) = \{t \in \mathbb{R} \mid x + tv \in Dom(f)\}$$

satisfies that $t \mapsto g(t)$ is convex for any $x \in Dom(f)$ and $v \in \mathbb{R}^d$

Example:

$f(X) = \log(\det(X))$ is concave on $Dom(f) = S_d^{++}(\mathbb{R})$

Hints:

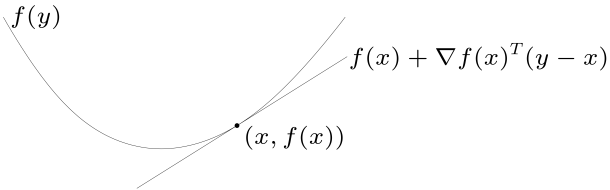
- $g(t) = f(X + tV) = f(X^{1/2} \cdot \left[I_d + tX^{-1/2}VX^{-1/2} \right] X^{1/2})$
- $g(t) = f(X) + \sum_{i=1}^d \log(1 + t\lambda_i)$, with $\lambda_i \geq 0$ eigenvalues of $I_d + tX^{-1/2}VX^{-1/2}$ (symmetric)

First-order characterization

Theorem

f: real-valued and differentiable function on the open set $Dom(f)$ is convex iff

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y \in Dom(f)$$



Second-order characterization

Theorem

f: real-valued and twice-differentiable function on the open set $Dom(f) \subset \mathbb{R}^d$ is convex iff

$$\nabla^2 f(x) \in S_d^+(\mathbb{R}), \quad \forall x \in Dom(f)$$

Strict convexity iff

$$\nabla^2 f(x) \in S_d^{++}(\mathbb{R}), \quad \forall x \in Dom(f)$$

Strong convexity iff there exists some $\mu > 0$ s.t.

$$\nabla^2 f(x) - \mu I_d \in S_d^+(\mathbb{R}), \quad \forall x \in Dom(f)$$

Example: $f(x) = \log\left(\sum_{i=1}^d e^{x_i}\right)$ is convex

Operations preserving convexity

Nonnegative sum and Pointwise maximum

- Nonnegative sum, and composition with affine function ($f(Ax+b)$ convex iff f convex)
- Pointwise maximum:
  $x \mapsto f(x) = \max\{f_1(x), \ldots, f_n(x)\}$ if $f_1, \ldots, f_n$: convex

Example:
For all $x \in \mathbb{R}^d$, $f(x) = \sum_{i=1}^k x_{(i)}$ is convex

$$(x_{(1)} \geq X_{(2)} \geq \cdots \geq x_{(k)})$$

Hint:

$$\begin{aligned} f(x) &= \sum_{i=1}^{k} x_{(i)} \\ &= \max\{x_{i_1} + \cdots + x_{i_k} \mid 1 \leq i_1 < i_2 < \cdots < i_k \leq d\} \end{aligned}$$

Operations preserving convexity (Cont'd)

Pointwise supremum

$x \mapsto f(x) = \sup_{y \in B} \{f(x,y)\}$ is convex if $x \mapsto f(x,y)$: convex for all $y \in B$

Examples:

- $f(x) = \sup_{a \in B} \left\{a^\top x\right\}$: convex
- $f(x) = \sup_{a \in B} \{\|x - a\|\}$: convex
- For all $X \in S_d(\mathbb{R})$,

$$f(X) = \sup_{\|u\|_{\mathbb{R}^d}} \left\{u^\top X u\right\} = \lambda_{\max}(X): \quad \text{convex}$$

Operations preserving convexity (Cont'd)

Minimization

- $x \mapsto f(x) = \inf_{y \in B} \{f(x, y)\}$ is convex if
  - B: convex
  - $(x, y) \mapsto f(x, y)$ is convex

Example:

- Distance to a set C (within $\mathbb{R}^d$):

$$x \mapsto dist(x, C) = \inf_{y \in C} \|x - y\| \quad \text{is convex if } C \text{ is convex}$$

Perspective

- $(x, t) \mapsto g(x, t) = t \cdot f(x/t)$ is convex if f is convex with $Dom(g) = \{(x, t) \mid x/t \in Dom(f), t > 0\}$

Example:

- $g(x, t) = t \log(t) - t \log(x)$ is convex on $(\mathbb{R}_+^*)^2$

$$\text{since } x \mapsto -\log(x) \quad \text{is convex}$$

Fenchel-Legendre transform

Fenchel-Legendre transform

Definition (Conjugate function)

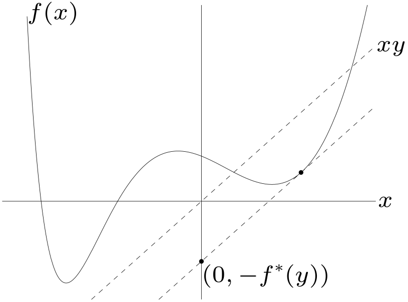
For any real-valued function f, the Fenchel-Legendre transform (also called the conjugate) of f is given by

$$f^\star(y) = \sup_{x \in Dom(f)} \{\langle y, x\rangle - f(x)\}$$

Remark: No convexity requirement on f!

Proposition

- *$f^\star$ is convex*



Examples

First case: $f(x) = -\log(x)$

Prove that:

$$f^{\star}(y) = \begin{cases} -1 - \log(-y), & y < 0 \\ +\infty, & y \geq 0 \end{cases}$$

Second case: $f(x) = \frac{1}{2}x^{\top}Ax$, with $A \in S_d^{++}(\mathbb{R})$

Prove that:

$$f^{\star}(y) = \frac{1}{2}y^{\top}A^{-1}y$$

Optimization problem

Standard Optimization problem

Classical formulation

$$
\begin{aligned}
\text{Minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\
& g_j(x) = 0, \quad \forall j = 1, \ldots, q
\end{aligned}
$$

Vocabulary

- x: optimization variable
- f_0: objective function (cost function)
- f_i: inequality constraint function
- g_j: equality constraint function

Feasible point and optimal value

Optimal value

$$p^\star = \inf \{ f_0(x) \mid f_i(x) \leq 0, 1 \leq i \leq p, \quad g_j(x) = 0, 1 \leq j \leq q \}$$

- $p^\star = +\infty$, if infeasible problem
  (no x satisfies the constraints)
- $p^\star = -\infty$, if unbounded below problem

Feasible point

- x feasible: $x \in Dom(f_0)$ and satisfies the constraints
- x optimal: x feasible and $f_0(x) = p^\star$

Example
With $f_0(x) = 1/x$ on $\mathbb{R}_+^*$, $p^\star = 0$ but no optimal point!

Convex optimization

Convex optimization problem

$$\begin{aligned} \text{Minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\ & a_j^\top x = b_j, \quad \forall j = 1, \ldots, q \end{aligned}$$

with $f_0, f_1, \ldots, f_p$ are convex

equivalent to

$$\begin{aligned} \text{Minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\ & Ax = b \qquad (A : \; q \times d \text{ matrix}) \end{aligned}$$

with $f_0, f_1, \ldots, f_p$ are convex

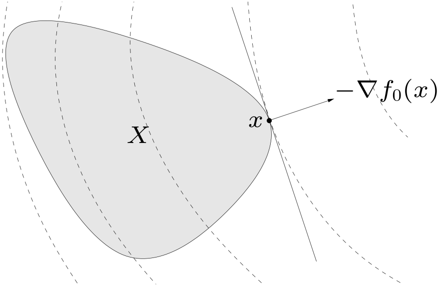
Remark:

The feasible set of convex optimization problem is convex

Optimiality criterion with differentiable f_0

With f_0 convex, x is optimal if

- x: feasible
- $\nabla f_0(x)^\top (y - x) \geq 0$, for all feasible y

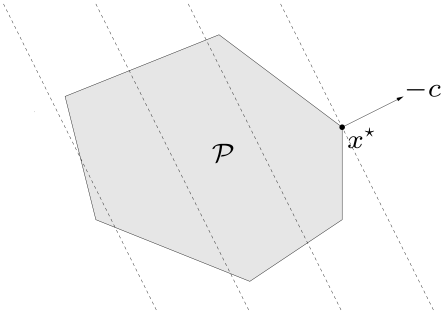


Hint:
f_0 convex implies $f_0(y) \geq f_0(x) + \nabla f_0(x)^\top (y - x)$,
for all feasible y

Linear program

$$
\begin{aligned}
\text{Minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\
& a_j^\top x = b_j, \quad \forall j = 1, \ldots, q
\end{aligned}
$$

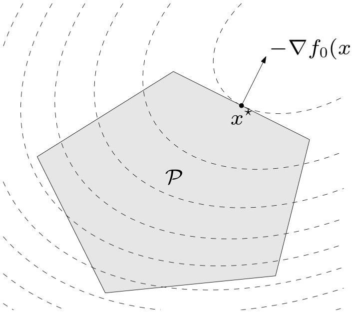
with f_0 and $f_1, \ldots, f_p$ are affine functions



Quadratic program

$$\begin{aligned} \text{Minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\ & a_j^\top x = b_j, \quad \forall j = 1, \ldots, q \end{aligned}$$

with $f_0(x) = \frac{1}{2} x^\top A x + B^\top x + C$ (where $A \in S_d^+(\mathbb{R})$)
and $f_1, \ldots, f_p$ are affine functions



Duality gap

Lagrangian

Classical formulation

$$\begin{array}{rll} \text{Minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & \forall i = 1, \ldots, p \\ & g_j(x) = 0, & \forall j = 1, \ldots, q \end{array}$$

$\mathcal{D}$: feasible set

Lagrangian

$L : \mathcal{D} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ s.t.

$$L(x; \lambda, \nu) = f_0(x) + \sum_{i=1}^{p} \lambda_i f_i(x) + \sum_{j=1}^{q} \nu_j g_j(x)$$

with $\lambda = (\lambda_1, \ldots, \lambda_p)^\top \geq 0$: Lagrange multipliers for negative constraints
and $\nu = (\nu_1, \ldots, \nu_q)^\top$ are Lagrange multipliers for equality constraints

Lagrange Dual function

Dual function

$L_D : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ s.t.

$$\begin{aligned} L_D(\lambda, \nu) &= \inf_{x \in \mathcal{D}} \{ L(x; \lambda, \nu) \} \\ &= \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^p \lambda_i f_i(x) + \sum_{j=1}^q \nu_j g_j(x) \right\} \end{aligned}$$

$(\lambda, \nu) \mapsto L_D(\lambda, \nu)$ is concave (can be $-\infty$ for some (λ, ν))

Remark:
The dual function provides a tool for deriving non-trivial lower bounds on $p^\star$

Lower bounding the optimal value

Proposition

For all $\lambda \in \mathbb{R}_+^p$ and $\nu \in \mathbb{R}^q$,

$$L_D(\lambda, \nu) \leq p^\star$$

Proof.

For any feasible $\tilde{x} \in \mathcal{D}$,

$$\begin{aligned} L_D(\lambda, \nu) &\leq f_0(\tilde{x}) + \sum_{i=1}^{p} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{q} \nu_j \overbrace{g_j(\tilde{x})}^{=0} \\ &\leq f_0(\tilde{x}) + \sum_{i=1}^{p} \overbrace{\lambda_i f_i(\tilde{x})}^{\leq 0} \leq f_0(\tilde{x}) = p^\star \end{aligned}$$

if $\tilde{x}$ is one optimal point. $\square$

Example:

Equality constrained norm minimization

$$\begin{aligned} \text{Minimize} \quad & \|x\| \\ \text{subject to} \quad & Ax = b \in \mathbb{R}^q \end{aligned}$$

Proposition

$$L_D(\nu) = \begin{cases} b^\top \nu, & \left\|A^\top \nu\right\|_* \leq 1 \\ -\infty, & \left\|A^\top \nu\right\|_* > 1 \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leq 1} v^\top u$

Proof.

If $\sup_{\|u\| \leq 1} y^\top u \leq 1$, then $\|y\| \leq 1$, $\sup_x \left\{x^\top y - \|x\|\right\} \leq 0$
Otherwise, $\sup_x \left\{x^\top y - \|x\|\right\} = +\infty$
Notice that

$$\begin{aligned} L_D(\nu) &= \inf_x \left\{\|x\| - \nu^\top A x + \nu^\top b\right\} \\ &= \nu^\top b - \sup_x \left\{(A^\top \nu)^\top x - \|x\|\right\} \end{aligned}$$

Example (Cont'd)

Equality constrained norm minimization

$$\begin{aligned} \text{Minimize} \quad & \|x\| \\ \text{subject to} \quad & Ax = b \in \mathbb{R}^q \end{aligned}$$

Proposition

$$L_D(\nu) = \begin{cases} b^\top \nu, & \left\|A^\top \nu\right\|_* \leq 1 \\ -\infty, & \left\|A^\top \nu\right\|_* > 1 \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leq 1} v^\top u$

Theorem

$$p^\star = \|x^\star\| \geq b^\top \nu, \quad \forall \nu \text{ s.t. } \left\|A^\top \nu\right\|_* \leq 1$$

Example (Cont'd)

Equality constrained squared-norm minimization

$$\begin{aligned} \text{Minimize} \quad & \|x\|^2 \\ \text{subject to} \quad & Ax = b \in \mathbb{R}^q \end{aligned}$$

Proposition

For all $\nu \in \mathbb{R}^q$,

$$L_D(\nu) = -\frac{1}{4}\nu^\top A^\top A\nu - b^\top \nu$$

Theorem

$$(p^\star)^2 = \|x^\star\|^2 \geq -\frac{1}{4}\nu^\top A^\top A\nu - b^\top \nu, \quad \forall \nu$$

Remark: Which bound is the best?

Weak duality: Tightest lower bound

Lagrange dual problem

$$\begin{aligned} \text{Maximize} \quad & L_D(\lambda, \nu) \\ \text{subject to} \quad & \lambda \in \mathbb{R}_+^p \end{aligned}$$

- Optimal value denoted by $d^\star$
- Best lower bound on $p^\star$
- (λ, ν) called dual feasible if $\lambda \in \mathbb{R}_+^p$ and $(\lambda, \nu) \in Dom(g)$
- Convex optimization problem (why?)

Weak duality

$$d^\star \leq p^\star$$

holds true for convex and non-convex problems

Strong duality

Strong duality

$$d^\star = p^\star$$

does not hold in general

- Often (not always!) holds true for convex problems
- Conditions for strong duality are called constraint qualifications (see Slater's constraint qualifications for instance)

Remark:
Strong duality can sometimes hold with non-convex problems

Slater's constraint qualifications for strong duality

Theorem

For a convex optimization problem

$$\begin{aligned} \textit{Minimize} \quad & f_0(x) \\ \textit{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, \ldots, p \\ & Ax = b \end{aligned}$$

strong duality holds true if it is strictly feasible that is,

$$\exists x \in Int(\mathcal{D}), \quad Ax = b, \quad \textit{and} \quad f_i(x) < 0, \quad \forall 1 \leq i \leq p$$

KKT Conditions

Strong duality and Lagrange multipliers

Assumption, notations

- Strong duality holds true ($d^\star = p^\star$)
- Optimal primal point $x^\star$, and optimal dual point $(\lambda^\star, \nu^\star)$

$$
\begin{aligned}
f_0(x^\star) &= L_D(\lambda^\star, \nu^\star) \\
&= \inf_x \left\{ f_0(x) + \sum_{i=1}^p \lambda_i^\star f_i(x) + \sum_{j=1}^q \nu_j^\star g_j(x) \right\} \\
&= \inf_x \left\{ f_0(x) + \sum_{i=1}^p \lambda_i^\star f_i(x) \right\} \\
&\leq f_0(x^\star) + \sum_{i=1}^p \underbrace{\lambda_i^\star f_i(x^\star)}_{\leq 0} \leq f_0(x^\star)
\end{aligned}
$$

Complementary slackness

For all $1 \leq i \leq p$,

$$\lambda_i^\star f_i(x^\star) = 0$$

Karush-Kuhn-Tucker (KKT) conditions

KKT conditions (with differentiable f_is and g_js)

1. Primal feasability (there exists x satisfying the constaints)
2. Dual feasability ($\lambda \in \mathbb{R}^p$)
3. Complementary slackness
4. First-order condition on the Lagrangian w.r.t. x:

$$\nabla f_0(x) + \sum_{i=1}^{p} \lambda_i \nabla f_i(x) + \sum_{j=1}^{q} \nu_j \nabla g_j(x) = 0$$

Necessary conditions

Strong duality holds true with differentiable constraints and x, λ, ν are optimal points imply that KKT conditions are fulfilled

KKT conditions are sufficient

Convexity and KKT conditions

- Convex optimization problem
- x, λ, ν satisfy the KKT conditions

Then, x, λ, ν are optimal and strong duality holds true

Proof.

- First-order condition implies $L(x; \lambda, \nu) = L_D(\lambda, \nu)$
- Complementary slackness implies $L(x; \lambda, \nu) = f_0(x)$
- Hence $L_D(\lambda, \nu) \leq d^\star \leq p^\star \leq f_0(x) = L_D(\lambda, \nu)$

$\square$

Remark:
Under Slater's constraint qualifications, the first-order condition can be relaxed since then:
x is optimal iff λ, ν satisfy the KKT conditions