Convex Optimization and Duality

François Durand

Nokia Bell Labs France

Based on the Online Course "Convex Optimization" by Stephen Boyd
and the Book "Convex Optimization" by Stephen Boyd and Lieven Vandenberghe

Reading Group
"Network Theory"

Lincs
24 September 2018

Sources

Book:

http://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

MOOC:

https://lagunita.stanford.edu/courses/Engineering/CVX101/Winter2014/about

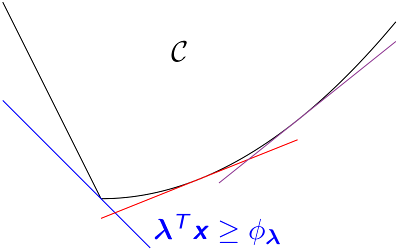
- Theory: 8h40 + exercises.
  - Convex sets,
  - Convex functions,
  - Convex optimization problems,
  - Duality.
- Applications: 3h15 + exercises.
  - Approximation and fitting, Statistical estimation, Geometric problems.
- Algorithms: 5h15 + exercises.
  - Numerical linear algebra, Unconstrained minimization, Equality constrained minimization, Interior point methods.

Plan

Plan

Fundamental Idea

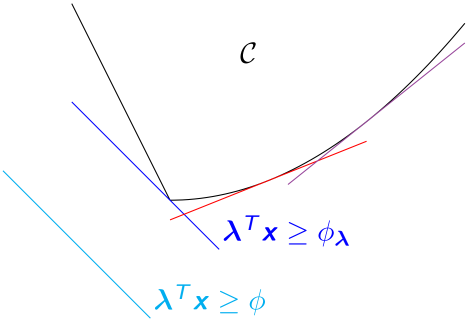
A convex set $\mathcal{C}$ can be fully described by its supporting hyperplanes.



Fundamental Idea

A convex set $\mathcal{C}$ can be fully described by its supporting hyperplanes.

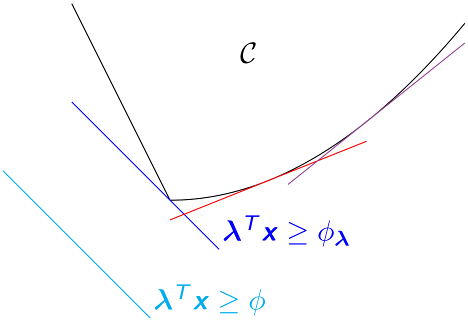
$$\phi_{\boldsymbol{\lambda}} = \inf\{\boldsymbol{\lambda}^T \boldsymbol{x}, \boldsymbol{x} \in \mathcal{C}\}$$
$$= \sup\{\phi \mid \forall \boldsymbol{x} \in \mathcal{C}, \boldsymbol{\lambda}^T \boldsymbol{x} \geq \phi\}.$$



Fundamental Idea

A convex set $\mathcal{C}$ can be fully described by its supporting hyperplanes.

$$\phi_{\boldsymbol{\lambda}} = \inf\{\boldsymbol{\lambda}^T \boldsymbol{x}, \boldsymbol{x} \in \mathcal{C}\}$$
$$= \sup\{\phi \mid \forall \boldsymbol{x} \in \mathcal{C}, \boldsymbol{\lambda}^T \boldsymbol{x} \geq \phi\}.$$



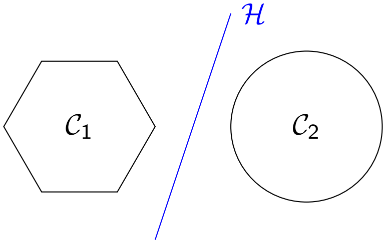
What is the dual space? Example:

- $x_1, x_2 \dots$ in (say) kg,
- ϕ in (say) \$,
- Then $\lambda_1, \lambda_2, \ldots$ in \$.kg^{-1}.

$\Rightarrow$ It is a space of unit prices.
$\Rightarrow$ If the unit prices were $\boldsymbol{\lambda}$, then each bundle in $\mathcal{C}$ would cost at least $\phi_{\boldsymbol{\lambda}}$.

Separating Hyperplanes Theorem

If two convex sets do not intersect, then they can be separated by (at least) a hyperplane.



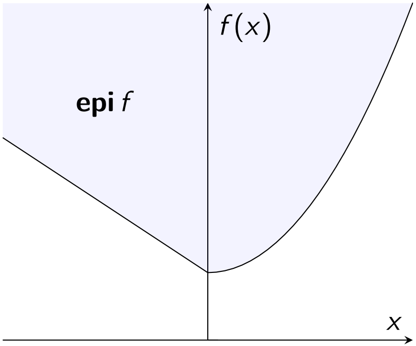
Plan

Convex Functions $\mathbb{R}^n \to \mathbb{R}$

Definition:

- $f\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$
- And **dom** f is a convex set.

I.e. the *epigraph* of f is a convex set.



Concave: $-f$ is convex. I.e. the *hypograph* of f is a convex set.

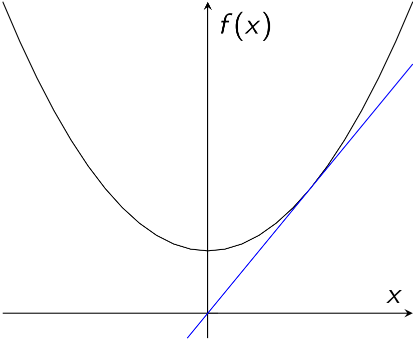
Affine $\Leftrightarrow$ convex and concave.

Convex Functions $\mathbb{R}^n \to \mathbb{R}$: Differentiable Case

If differentiable:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\nabla} f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}).$$

$\Rightarrow$ First-order Taylor gives a global underestimator.
$\Rightarrow$ For example, if $\boldsymbol{\nabla} f(\boldsymbol{x}) = \mathbf{0}$...



If differentiable twice: $\boldsymbol{\nabla}^2 f(\boldsymbol{x}) \succeq 0$ (the Hessian is semidefinite positive).

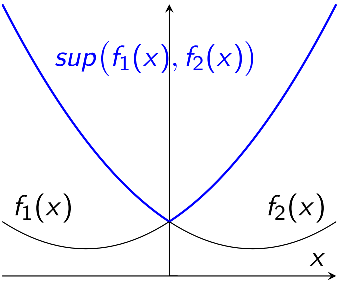
Examples of Convex Functions

- Any norm on $\mathbb{R}^n$,
- $\boldsymbol{x} \to \boldsymbol{x}^T P \boldsymbol{x}$, where P is semidefinite positive,
- $\boldsymbol{x} \to \max_i(x_i)$,
- $\boldsymbol{x} \to \log \sum_i \exp(x_i)$ ("LogSumExp"),
- $\boldsymbol{x} \to \left(\prod_i x_i\right)^{\frac{1}{n}}$ (geometric mean),
- $\boldsymbol{x} \to \infty_{\boldsymbol{x} \notin \mathcal{C}}$, where $\mathcal{C}$ is a convex set,
- $X \to \log \det X$ over the set of definite positive matrices,
- $X \to \text{eigenvalue}_{\max}(X)$ over the set of symmetric matrices.

Convex Functions Calculus

Are convex:

- $\boldsymbol{x} \to f(\boldsymbol{x}) + g(\boldsymbol{x})$,
- $\boldsymbol{x} \to \lambda f(\boldsymbol{x})$, for $\lambda \geq 0$,
- $\boldsymbol{x} \to f(A\boldsymbol{x} + \boldsymbol{b})$,
- $\boldsymbol{x} \to sup_{\theta \in \Theta} f_\theta(\boldsymbol{x})$,



Composition rule. Assume:

- Functions v_i are convex, c_i concave, a_i affine,
- f convex, nondecreasing in each argument v_i, nonincreasing in each argument c_i.

Then $\boldsymbol{x} \to f\left(v_1(\boldsymbol{x}), \ldots, c_1(\boldsymbol{x}), \ldots, a_1(\boldsymbol{x}), \ldots\right)$ is convex.

[Mnemonic: write second derivative. But still true if not differentiable.]

Prove that a Function is Convex in Practice

- Apply the definition (extremely rare),
- Compute the Hessian / second derivative (to be avoided if possible),
- Prove that any restriction to a line, i.e. $t \to f(\boldsymbol{x} + t\boldsymbol{v})$, is convex (sometimes),
- Apply the rules of the previous slide (laziest hence preferred method).

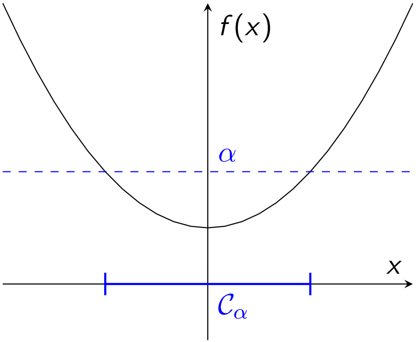
Sublevel Sets

If f is convex, then:

$$\mathcal{C}_\alpha = \{\boldsymbol{x} \in \textbf{dom}\, f \mid f(\boldsymbol{x}) \leq \alpha\}$$

is convex.

$\Rightarrow$ Often used to prove that a set is convex.



Plan

Optimization Problem in the Standard Form

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \le 0, \quad i = 1, \dots, r, \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, s. \end{aligned}$$

- $f(\boldsymbol{x})$: cost (e.g. in \$).
- Example of an inequality constraint:

$$v_1 x_1 + v_2 x_2 - S \le 0$$

  where x_i in kg, v_i in $\text{m}^3.\text{kg}^{-1}$ and S is the volume of my warehouse in m^3.
- (Theoretical) remark: any equality constraint can be seen as two opposite inequality constraints.

minimize $f(\boldsymbol{x})$
subject to $g_i(\boldsymbol{x}) \le 0, \quad i = 1, \dots, r,$
$h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, s.$

A Bit of Vocabulary

$\mathcal{D} = \mathbf{dom}\, f \ \cap \ \bigcap_i \mathbf{dom}\, g_i \ \cap \ \bigcap_j \mathbf{dom}\, h_j$: *domain* of the problem.

$\mathcal{F} = \{\boldsymbol{x} \in \mathcal{D} \mid \forall i, g_i(\boldsymbol{x}) \le 0 \text{ and } \forall j, h_j(\boldsymbol{x}) = 0\}$: set of *feasible points*.

$p^\star = \inf_{\boldsymbol{x} \in \mathcal{F}} f(\boldsymbol{x})$: *optimal value* of the problem.

- $p^\star = \infty$: the problem is *infeasible*, i.e. $\mathcal{F} = \varnothing$.
- $p^\star = -\infty$: the problem is *unbounded below*.
- $p^\star$ is finite:
  - If $f(\boldsymbol{x}^\star) = p^\star$, then $\boldsymbol{x}^\star$ is an *optimal point* or *solution*.
  - If no such $\boldsymbol{x}^\star$, then the optimal value $p^\star$ is *not attained.*

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \le 0, & i = 1, \dots, r, \\ & h_j(\boldsymbol{x}) = 0, & j = 1, \dots, s. \end{array}$$

Convex Optimization Problem

We say that the problem is convex if:

- f is convex,
- All g_i are convex,
- And all h_j are affine.

Motivation: this ensures that f restricted to $\mathcal{F}$ is a convex function.

Then:

- Any local minimum is a global optimum, i.e. a solution.
- (Generally) efficient algorithms to find a solution.

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \dots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \dots, s.$

Lagrangian: Motivation

We do **not** assume that the problem is convex.

The problem is equivalent to:

$$\text{minimize} \;\; f(\boldsymbol{x}) + \sum_i \infty_{g_i(\boldsymbol{x})>0} + \sum_j \infty_{h_j(\boldsymbol{x})\neq 0} \qquad (x \in \mathcal{D})$$

Example with constraint $v_1x_1 + v_2x_2 - S \leq 0$: using more than S has infinite cost, using less than S is costless (but does not generate an income).

Relaxation of the problem: fix a unit price $\lambda \geq 0$ (in \$.m^{-3}). Use more than S: buy space at price λ. Use less: sell the extra space at price λ.

$$\text{minimize} \;\; f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x}) \qquad (x \in \mathcal{D})$$

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \ldots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \ldots, s.$

Lagrangian and Dual Lagrangian

We do **not** assume that the problem is convex.

Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$$

Dual Lagrangian:

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

- It is always concave (even if the original problem is not convex).
- It provides a *parametrized family of lower bounds* for f. Indeed, for any feasible $\boldsymbol{x}$ and any $\boldsymbol{\lambda} \geq 0$ (and without any requirement on $\boldsymbol{\nu}$):

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \underbrace{\lambda_i g_i(\boldsymbol{x})}_{\leq 0} + \sum_j \nu_j \underbrace{h_j(\boldsymbol{x})}_{=0} \leq f(\boldsymbol{x}).$$

Dual Problem

We do **not** assume that the problem is convex.

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0, & i = 1, \ldots, r, \\ & h_j(\boldsymbol{x}) = 0, & j = 1, \ldots, s. \end{array}$$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$$

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

We have $\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$. To find the best lower bound, let us maximize ϕ. This is the dual problem:

$$\begin{array}{lll} \text{maximize} & \phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) & \\ \text{subject to} & \lambda_i \geq 0, & i = 1, \ldots, r. \end{array}$$

This is a convex problem, hence (generally) convenient to solve, at least numerically.

Its optimal value is denoted by $d^\star$. By construction, we have $d^\star \leq p^\star$.
Duality gap: $p^\star - d^\star$.

- If $= 0$: "strong duality" situation.
- If > 0: "weak duality" situation.

Slater's Conditions

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \le 0, & i = 1, \dots, r, \\ & h_j(\boldsymbol{x}) = 0, & j = 1, \dots, s. \end{array}$$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

$$\begin{array}{lll} \text{maximize} & \phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) & \\ \text{subject to} & \lambda_i \ge 0, & i = 1, \dots, r. \end{array}$$

If:

- The primal problem is convex,
- And the constraints are "strictly feasible" (i.e. with strict inequalities),

Then:

- $p^\star = d^\star$,
- If $p^\star = d^\star > -\infty$, then the dual optimum is attained, i.e. $\exists(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ s.t. $\phi(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = d^\star = p^\star$.

Remark: for *affine* inequality constraint, the "strictness" condition can be dropped.

minimize $f(\boldsymbol{x})$
subject to $g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, r,$
$h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize $\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$
subject to $\lambda_i \geq 0, \quad i = 1, \ldots, r.$

Geometric Interpretation: Set of Values

We do **not** assume that the problem is convex.

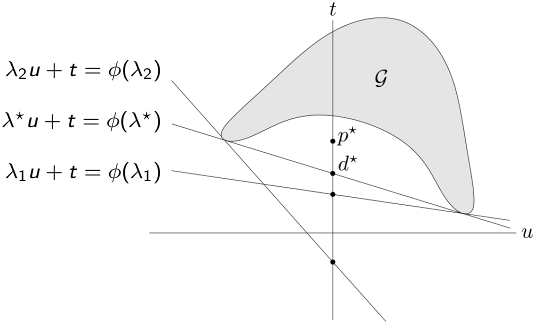
Example: only one constraint $g(\boldsymbol{x}) \leq 0$.

$$\mathcal{G} = \left\{ (g(\boldsymbol{x}), f(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathcal{D} \right\}.$$

$$p^\star = \inf\{t \mid (u, t) \in \mathcal{G} \text{ and } u \leq 0\}.$$

$$\begin{aligned} \phi(\lambda) &= \inf\{f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{D}\} \\ &= \inf\{\lambda u + t \mid (u, t) \in \mathcal{G}\} \\ &= \sup\{b \mid \forall (u, t) \in \mathcal{G}, \lambda u + t \geq b\}. \end{aligned}$$

$\Rightarrow$ $\lambda u + t \geq \phi(\lambda)$ is a supporting hyperplane of $\mathcal{G}$ (of slope $-\lambda$).



Geometric Interpretation: Epigraph variation

We do **not** assume that the problem is convex.

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \ldots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \ldots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize	$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$	
subject to	$\lambda_i \geq 0,$	$i = 1, \ldots, r.$

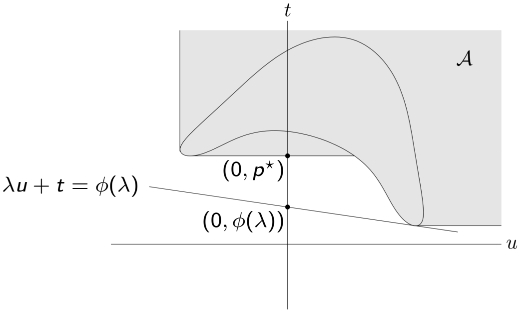
Example: only one constraint $g(\boldsymbol{x}) \leq 0$.

$$\mathcal{A} = \{(u, t) \mid \exists \boldsymbol{x} \in \mathcal{D}, u \geq g(\boldsymbol{x}), t \geq f(\boldsymbol{x})\}$$
$$= \mathcal{G} \cup \text{ points that are worse.}$$

$$p^\star = \inf\{t \mid (0, t) \in \mathcal{A}\}.$$

$$\begin{aligned}\phi(\lambda) &= \inf\{f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{D}\} \\ &= \inf\{\lambda u + t \mid (u, t) \in \mathcal{A}\} \\ &= \sup\{b \mid \forall (u, t) \in \mathcal{A}, \lambda u + t \geq b\}.\end{aligned}$$

$\Rightarrow \lambda u + t \geq \phi(\lambda)$ is a supporting hyperplane of $\mathcal{A}$ (of slope $-\lambda$).



minimize $f(\boldsymbol{x})$
subject to $g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, r,$
$h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

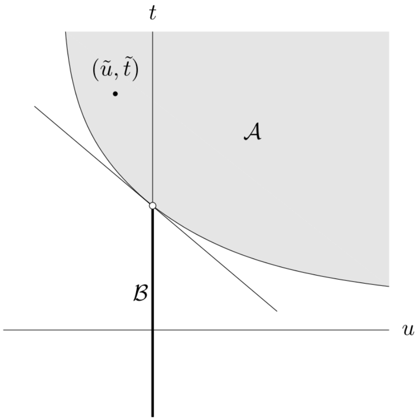
maximize $\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$
subject to $\lambda_i \geq 0, \quad i = 1, \ldots, r.$

Convex Problems: Why There is (Usually) Strong Duality

If the optimization problem is convex, then $\mathcal{A}$ is convex.
There is a separating hyperplane between $\mathcal{A}$ and $\mathcal{B} = \{(0, t) \mid t < p^\star\}$.

[Here, constraints qualification such as Slater's conditions ensure that the hyperplane is not vertical.]

This hyperplane gives the good λ and $\phi(\lambda)$.



Example of Convex Problem Without Strong Duality

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \ldots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \ldots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

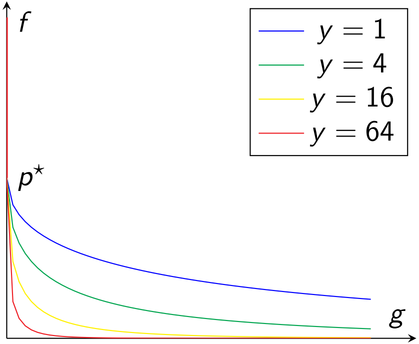
$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize	$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$	
subject to	$\lambda_i \geq 0,$	$i = 1, \ldots, r.$

minimize e^{-x}
subject to $x^2/y \leq 0$
with $\mathcal{D} = \{(x, y) \mid y > 0\}$.

This is a convex problem.
Constraint means $x = 0$.
N.B.: Slater's conditions are violated.

$p^\star = 1$
$\phi(\lambda) = \inf_{x,y} e^{-x} + \lambda \frac{x^2}{y} = 0$
$d^\star = 0$
$\Rightarrow$ No strong duality.



minimize $f(\boldsymbol{x})$
subject to $g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, r,$
$h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize $\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$
subject to $\lambda_i \geq 0, \quad i = 1, \dots, r.$

Saddle-point interpretation

We do **not** assume that the problem is convex.

$$\sup_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \sup_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\nu}} \left(f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x}) \right)$$

$$= \left| \begin{array}{l} f(\boldsymbol{x}) \text{ if } \boldsymbol{x} \in \mathcal{F}, \\ \infty \text{ otherwise.} \end{array} \right.$$

$$\Rightarrow p^\star = \inf_{\boldsymbol{x} \in \mathcal{D}} \sup_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\nu}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

And by definition: $d^\star = \sup_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\nu}} \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$

Hence weak duality is simply a particular case of the max-min inequality: $\sup_y \inf_x f(x, y) \leq \inf_x \sup_y f(x, y)$. And:

$(\boldsymbol{x}, (\boldsymbol{\lambda}, \boldsymbol{\nu}))$ saddle-point of $L \Leftrightarrow \boldsymbol{x} = \boldsymbol{x}^\star, \boldsymbol{\lambda} = \boldsymbol{\lambda}^\star, \boldsymbol{\nu} = \boldsymbol{\nu}^\star$ and $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = p^\star = d^\star$.

Plan

Complementary slackness

We do **not** assume that the problem is convex.

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \dots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \dots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize	$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$	
subject to	$\lambda_i \geq 0,$	$i = 1, \dots, r.$

Assume strong duality: $d^\star = p^\star$. Assume these optimal values are reached at $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ and $\boldsymbol{x}^\star$ respectively. Then all the following inequalities are in fact equalities:

$$\phi(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \leq L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f(\boldsymbol{x}^\star) + \sum_i \underbrace{\lambda_i^\star g_i(\boldsymbol{x}^\star)}_{\leq 0} + \sum_j \nu_j^\star \underbrace{h_j(\boldsymbol{x}^\star)}_{=0} \leq f(\boldsymbol{x}^\star).$$

Hence:

- $L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f(\boldsymbol{x}^\star) = \phi(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star),$
- $\forall i, \lambda_i^\star g_i(\boldsymbol{x}^\star) = 0$ (complementary slackness). I.e. $\lambda_i^\star = 0$ or $g_i(\boldsymbol{x}^\star) = 0.$

minimize	$f(\boldsymbol{x})$	
subject to	$g_i(\boldsymbol{x}) \leq 0,$	$i = 1, \dots, r,$
	$h_j(\boldsymbol{x}) = 0,$	$j = 1, \dots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize	$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$	
subject to	$\lambda_i \geq 0,$	$i = 1, \dots, r.$

Karush-Kuhn Tucker conditions

We do **not** assume that the problem is convex.

We now assume that f, all g_i and all h_j are differentiable.

1. If $f(\boldsymbol{x}^\star) = p^\star = d^\star = \phi(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$, then (KKT conditions):
   - $\boldsymbol{\nabla}_x L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \mathbf{0}$,
   - $\forall i, g_i(\boldsymbol{x}^\star) \leq 0$ and $\forall j, h_j(\boldsymbol{x}^\star) = 0$ (primal feasibility),
   - $\forall i, \lambda_i^\star \geq 0$ (dual feasibility),
   - $\forall i, \lambda_i^\star g_i(\boldsymbol{x}^\star) = 0$ (complementary slackness).
2. If the problem is convex, then the converse is true.
3. If the problem is convex and satisfies Slater's conditions, then $\boldsymbol{x}^\star$ is optimal iff there exists $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ that meets KKT conditions.

KKT: Geometric Interpretation (One Constraint)

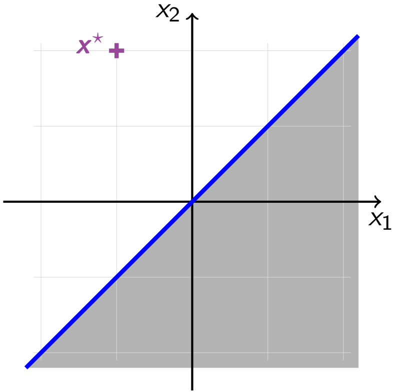
$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_1(\boldsymbol{x}) \leq 0 \end{array}$$

Example:

$$x_1 - \quad x_2 \leq 0$$

If $\boldsymbol{x}^\star$ is in the interior:

$$\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$$



KKT: Geometric Interpretation (One Constraint)

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_1(\boldsymbol{x}) \leq 0 \end{array}$$

Example:

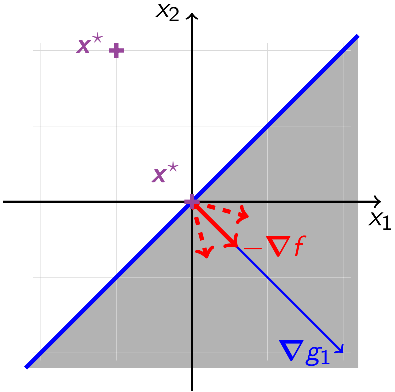
$$x_1 - \quad x_2 \leq 0$$

If $\boldsymbol{x}^\star$ is in the interior:

$$\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \boldsymbol{0}$$

If $\boldsymbol{x}^\star$ is on the frontier:

$$-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_1 \boldsymbol{\nabla} g_1(\boldsymbol{x}^\star) \quad (\text{with } \lambda_1 \geq 0)$$



KKT: Geometric Interpretation (One Constraint)

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_1(\boldsymbol{x}) \leq 0 \end{array}$$

Example:

$$x_1 - \quad x_2 \leq 0$$

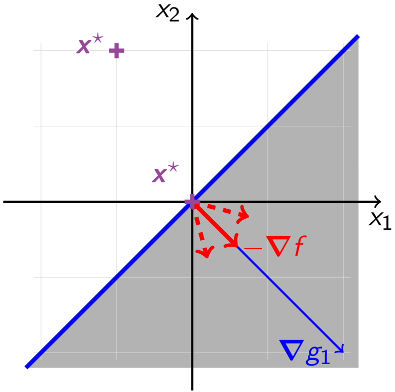
If $\boldsymbol{x}^\star$ is in the interior:

$$\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$$

If $\boldsymbol{x}^\star$ is on the frontier:

$$-\nabla f(\boldsymbol{x}^\star) = \lambda_1 \nabla g_1(\boldsymbol{x}^\star) \quad (\text{with } \lambda_1 \geq 0)$$

Anyway, the second condition is met.
Moreover, if $\lambda_1 > 0$, then $g_1(\boldsymbol{x}) = 0$.



KKT: Geometric Interpretation (Several Constraints)

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_1(\boldsymbol{x}) \leq 0 \\ & g_2(\boldsymbol{x}) \leq 0 \end{array}$$

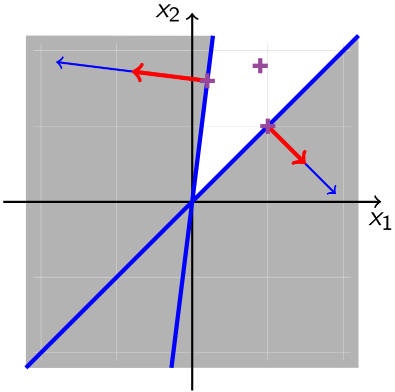
Example:

$$\begin{aligned} x_1 - \quad x_2 &\leq 0 \\ -2x_1 + 0.25x_2 &\leq 0 \end{aligned}$$

$\boldsymbol{x}^\star$ interior: $\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \mathbf{0}$.

$\boldsymbol{x}^\star$ on first frontier: $-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_1 \boldsymbol{\nabla} g_1(\boldsymbol{x}^\star)$.

$\boldsymbol{x}^\star$ on second frontier: $-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_2 \boldsymbol{\nabla} g_2(\boldsymbol{x}^\star)$.

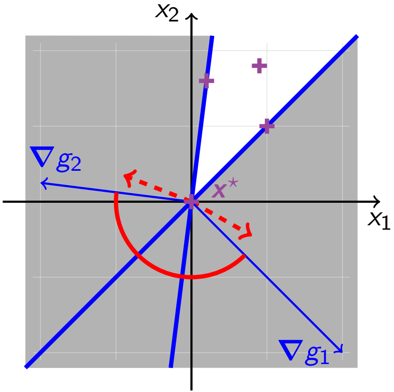


KKT: Geometric Interpretation (Several Constraints)

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_1(\boldsymbol{x}) \leq 0 \\ & g_2(\boldsymbol{x}) \leq 0 \end{aligned}$$

Example:

$$\begin{aligned} x_1 - \quad x_2 &\leq 0 \\ -2x_1 + 0.25x_2 &\leq 0 \end{aligned}$$



$\boldsymbol{x}^\star$ interior: $\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \boldsymbol{0}$.
$\boldsymbol{x}^\star$ on first frontier: $-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_1 \boldsymbol{\nabla} g_1(\boldsymbol{x}^\star)$.
$\boldsymbol{x}^\star$ on second frontier: $-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_2 \boldsymbol{\nabla} g_2(\boldsymbol{x}^\star)$.
If $\boldsymbol{x}^\star$ is on the intersection of frontiers:

$$-\boldsymbol{\nabla} f(\boldsymbol{x}^\star) = \lambda_1 \boldsymbol{\nabla} g_1(\boldsymbol{x}^\star) + \lambda_2 \boldsymbol{\nabla} g_2(\boldsymbol{x}^\star) \quad (\text{with } \lambda_i \geq 0).$$

In all cases, the above condition is met.
Moreover, if $\lambda_i > 0$, then $g_i(\boldsymbol{x}) = 0$.

Sensitivity Analysis (Very Quickly)

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0, & i = 1, \ldots, r, \\ & h_j(\boldsymbol{x}) = 0, & j = 1, \ldots, s. \end{array}$$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

$$\begin{array}{lll} \text{maximize} & \phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) & \\ \text{subject to} & \lambda_i \geq 0, & i = 1, \ldots, r. \end{array}$$

Consider a perturbed problem:

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) \leq u_i, & i = 1, \ldots, r, \\ & h_j(\boldsymbol{x}) = v_j, & j = 1, \ldots, s. \end{array}$$

Denote $p^\star(\boldsymbol{u}, \boldsymbol{v})$ its optimal value.
If strong duality holds and if $p^\star$ is differentiable at $(\mathbf{0}, \mathbf{0})$, then:

$$\lambda_i^\star = -\frac{\partial p^\star}{\partial u_i}(\mathbf{0}, \mathbf{0}) \text{ and } \nu_i^\star = -\frac{\partial p^\star}{\partial v_i}(\mathbf{0}, \mathbf{0}).$$

Take-aways

minimize $f(\boldsymbol{x})$
subject to $g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, r,$
$h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, s.$

$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x})$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\nu}) = inf_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$

maximize $\phi(\boldsymbol{\lambda}, \boldsymbol{\nu})$
subject to $\lambda_i \geq 0, \quad i = 1, \ldots, r.$

- The **Lagrangian** describes a *relaxation* of the problem with a *unit price* for each constraint (Lagrange multiplier).
- The **dual Lagrangian** provides a *parametrized family of lower bounds* for the primal problem.
- The **dual problem** is *always convex*, even if the primal problem is not.
- When the primal problem is convex, there is usually **strong duality** (with mild additional assumptions such as *Slater's conditions*).
- For differentiable problems, think of **KKT conditions** (*necessary* if there is strong duality, *sufficient* if the problem is convex).

Thank you!

Questions?