Master in
Computer Vision
Barcelona

UAB UOC UPC upf.

T6: Convex Optimization (II). Duality principles and methods

Coloma Ballester - UPF

October 26, 2021

Optimization and inference techniques for Computer Vision

Previously on...

Unconstrained and constrained optimization
Do we still have many open questions?

Does our problem have a solution?
(Existence) ✓

Does our problem have an unique solution?
(Uniqueness) ✓

How do we know if a point x is a solution?
(Optimality conditions) ✓

Is it possible to find the solution?
(Convexity) ✓

Does our problem still have solutions if we have restrictions on them?
(Constrained Optimization) ✓

Convex problems are easy

In the previous lectures you have worked with problems in where we can answer some of the previous questions.

In convex problems we can assure the existence of solutions.

In strictly convex problems we can assure the uniqueness of their solutions.

If the general problem is (strictly) convex and the restrictions on our solutions enclose them in convex sets then we can assure the (uniqueness and) existence of solutions.

Convex constrained minimization

Consider the **constrained minimization problem**

$$\min_{x \in C} f(x).$$

Theorem

Assume that C is a **convex subset** of $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a **convex function**.

Then, **a local minimum of f over C is also a global minimum over C**.

Moreover, if f is **strictly convex**, then **any global minimum in C is unique**.

Convex problems

Corollary

If C is a **closed convex subset** of $\mathbb{R}^n$ and $f : C \to \mathbb{R}$ is **convex and continuous** on C, then f **attains its infimum**.

That is, if we solve

$$\inf_{x \in C} f(x)$$

there is a point $x_0 \in C$ such that

$$f(x_0) = \min_{x \in C} f(x).$$

In other words, **convex functions on (closed) convex sets have minima.**

Convex constrained optimization

How to compute the minimum of a convex function with convex restrictions on its variables?

The solutions will satisfy the so-called the **Karush-Kuhn-Tucker (KKT) optimality conditions.**

The KKT optimality conditions are the necessary and sufficient conditions of a minimum. They allow to write equations to compute the solution to the problems.

General case: equality and inequality constraints

Consider the smooth functions $f, c_1, \cdots, c_k, d_1, \ldots, d_r : \mathbb{R}^n \to \mathbb{R}$.
Usually, f is called the **objective function**.

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} f(x) \\
&\text{s.t.} \qquad c_1(x) \geq 0, \ldots, c_k(x) \geq 0 \qquad \text{(inequality constraints)} \\
&\text{and} \qquad d_1(x) = 0, \ldots, d_r(x) = 0 \qquad \text{(equality constraints)}
\end{aligned}
\tag{1}
$$

Define the Lagrange function for the problem as Lagrange multipliers λ_i, $i = 1, \ldots, k$, ν_j, $j = 1, \ldots, r$ of the problem. Setting $\lambda = (\lambda_1, \ldots, \lambda_k)$, $\nu = (\nu_1, \ldots, \nu_r)$, the Lagrange function

$$
\mathscr{L}(x, \lambda, \nu) = f(x) - \sum_{i=1}^{k} \lambda_i c_i(x) - \sum_{j=1}^{r} \nu_j d_j(x)
$$

Karush-Kuhn-Tucker (KKT) optimality conditions

$$
\begin{aligned}
&\min_{x\in\mathbb{R}^n} f(x) \\
&\text{s.t.} \qquad c_1(x)\geq 0,\ldots,c_k(x)\geq 0 \qquad\qquad \text{(inequality constraints)} \\
&\text{and} \qquad d_1(x)=0,\ldots,d_r(x)=0 \qquad\qquad \text{(equality constraints)}
\end{aligned} \tag{2}
$$

Suppose that there is a **minimum x_0 of (2)**.
It coincides with a **saddle point** (x_0, λ^0, ν^0) **of** $\mathscr{L}(x,\lambda,\nu)$.

The **Karush-Kuhn-Tucker (KKT) optimality conditions** are

$$
\begin{aligned}
\nabla_x \mathscr{L}(x_0,\lambda^0) &= 0 && && \text{dual feasibility} \\
c_i(x_0) &\geq 0 && \forall i && \text{primal feasibility} \\
d_j(x_0) &= 0 && \forall j && \text{primal feasibility} \\
\lambda_i^0 &\geq 0 && \forall i && \text{dual positivity} \\
\lambda_i^0 c_i(x_0) &= 0 && \forall i && \text{complementary slackness}
\end{aligned}
$$

Notice that the equations $\lambda_i^0 c_i(x_0)=0$ mean that:

- either $c_i(x_0)=0$ (**active constraint**) and therefore $\lambda_i^0>0$,
- or $c_i(x_0)>0$ (x_0 **interior point**) and therefore $\lambda_i^0=0$.

Karush-Kuhn-Tucker (KKT) optimality conditions

The KKT optimality conditions are

$$
\begin{aligned}
\nabla_x \mathscr{L}(x_0, \lambda^0) &= 0 && && \text{dual feasibility}\\
c_i(x_0) &\geq 0 && \forall i && \text{primal feasibility}\\
d_j(x_0) &= 0 && \forall j && \text{primal feasibility}\\
\lambda_i^0 &\geq 0 && \forall i && \text{dual positivity}\\
\lambda_i^0 c_i(x_0) &= 0 && \forall i && \text{complementary slackness}
\end{aligned}
$$

The KKT optimality conditions are **necessary conditions**, that is, they hold for a minimum of (1).

If the constraints **$-\mathbf{c_i}(\mathbf{x})$ are convex and there is a point $\mathbf{x}$ such that $\mathbf{c_i}(\mathbf{x}) > \mathbf{0}$ for all $i = 1, \ldots, k$**, then they are also **sufficient conditions**, that is, if they hold, the point $\bar{x}$ is a minimum of (1).

Karush-Kuhn-Tucker (KKT) optimality conditions

Remark the role played by considering only $\lambda \geq 0$.

In other words, while the sign of the Lagrange multipliers in case of equality constraints is not specified, in case of inequality constraints is specified.

This is because the KKT conditions come of considering Lagrange multipliers λ_i, $i = 1, \ldots, k$, ν_j, $j = 1, \ldots, r$ of the problem

$$\mathcal{L}(x, \lambda, \nu) = f(x) - \sum_{i=1}^{k} \lambda_i c_i(x) - \sum_{j=1}^{r} \nu_j d_j(x)$$

Nevertheless, the fact that they are positive is a convention. Notice that **we put a minus sign in front of the constraint terms in the Lagrangian**, so $\mathcal{L}$ is convex as a function of x.

Convex optimization: We still have a quite big open question

Does our problem have a solution?
(Existence) ✓

Does our problem have an unique solution?
(Uniqueness) ✓

How do we know if a point x is a solution?
(Optimality conditions) ✓

Is it possible to find the solution?
(Convexity) ✓

Does our problem still have solutions if we have restrictions on them?
(Constrained Optimization) ✓

Can we still find solutions for non-differentiable problems?
(Non-smooth Optimization) ✗

In general: Optimization problems

VA:

$$\min_{\mathbf{u} \in \mathbf{C}} \mathbf{J}(\mathbf{u})$$

AA:

$$\min_{\mathbf{x} \in \mathbf{C}} \mathbf{f}(\mathbf{x})$$

$\longrightarrow$ **What happens if you can't differentiate because either u belongs to an space of non-derivable functions, or if $J(u)$ is not derivable?**

Non-differentiable case

This creates a difficulty: **the two optimization strategies** you have learned so far,

- **Euler-Lagrange equation** (which is a **extremality principle**):
  $\frac{\mathbf{dJ}}{\mathbf{du}}(\mathbf{u_0}) = \mathbf{0} \iff \nabla \mathbf{J}(\mathbf{u_0}) = \mathbf{0}$ (analogy with $\nabla f(x_0) = 0$)
- **Gradient descent**:

$$\begin{cases} \mathbf{u}^{\mathbf{k}+1} = \mathbf{u}^{\mathbf{k}} - \tau \nabla \mathbf{J}(\mathbf{u}^{\mathbf{k}}) \\ \mathbf{u}^{\mathbf{0}} = \mathbf{u_0} \end{cases}$$

use the 'derivative of the functional J(u) with respect to the function u',
$\nabla_{\mathbf{u}} \mathbf{J}$
(also denoted by $\frac{\mathbf{dJ}}{\mathbf{du}}$**).**

What happens if you can't because either u belongs to an space of non-derivable functions, or if J is not derivable??

Examples: Convex Problems you have already seen...

... written as finite dimensional problems (i.e. as matrices and vectors).

Image denoising

Given f a noisy image, recover u as the solution of

$$\min_u \|\nabla^h u\|^2 + \lambda \|u - f\|^2$$

Image inpainting

Given f an image and a mask M defining the region that should be preserved:

$$\min_u \|\nabla^h u\|^2 \quad \text{s.t. } M \odot u = f$$

Remember that approximating ∇ with finite differences it can be expressed as a matrix.

However, some non-smooth functionals yield sharper results...

... more similar to real world scenes which are made of well-contrasted objects that, in the image or video capturing the scene, frequently partially occlude other objects and the backgorund.

Total Variation image denoising (aka ROF denoising model)

Given f a noisy image, recover u as the solution of

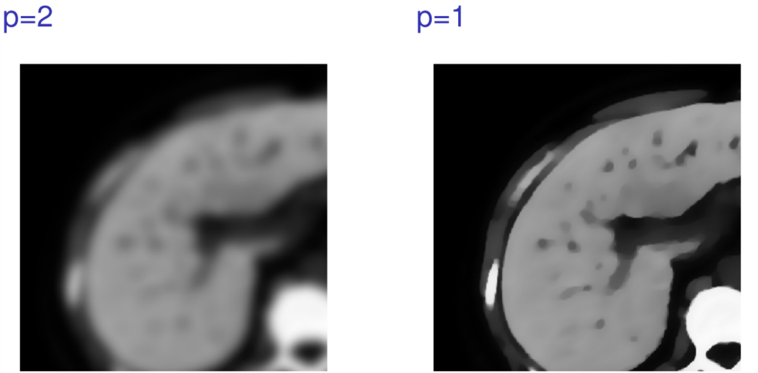
$$\min_u \|\nabla^h u\| + \lambda \|u - f\|^2$$

Image inpainting

Given f an image and a mask M defining the region that should be preserved:

$$\min_u \|\nabla^h u\| \quad \text{s.t. } M \odot u = f$$

Image denoising problem $J(u) = \frac{1}{2\lambda}\int_\Omega |u-f|^2 d\mathbf{x} + \frac{1}{2}\int_\Omega |\nabla u|^p d\mathbf{x}$



Indeed, some non-smooth functionals yield sharper results

- **Example ($p = 1$):**, Total Variation image denoising (aka ROF denoising model):

$$J(u) = \frac{1}{2\lambda}\int_\Omega |u - f|^2 d\mathbf{x} + \frac{1}{2}\int_\Omega |\nabla u| d\mathbf{x}$$

It is also a convex function. But we cannot "differentiate" as you did with before with other differenciable functionals.

- Moreover, $u \in C$ can be non-differentiable (for instance, if u is a piecewise constant image).

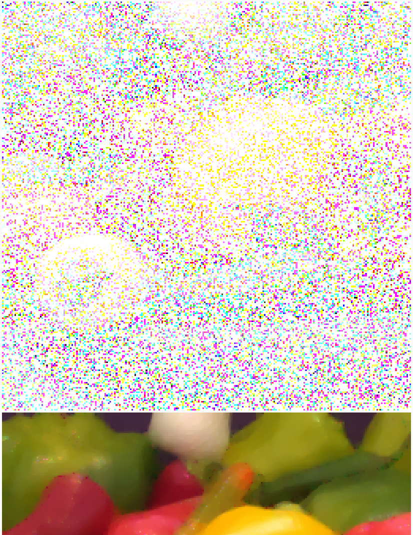
- **Another example:** optical flow estimation in video by minimization of the **TV**-$\mathrm{L}^{\mathbf{1}}$ optical flow functional

$$\min_{\mathbf{v}} \lambda \int_\Omega |(I(\mathbf{x} + \mathbf{v}, t+1) - I(\mathbf{x}, t)| d\mathbf{x} + \int_\Omega |\nabla \mathbf{v}| d\mathbf{x}$$

where $\mathbf{v}$ is the (unknown) optical flow, that is, the vector field that recovers the apparent motion of two consecutive frames of the input video.

Similar example: Inpainting with $\mathbf{p = 1}$

Recovering/interpolating an image in regions where the original information is missing. Given the color image $f \in L^{\infty}(\Omega)$, the image domain Ω an the region $D \subset \Omega$ where the image will be inpainted.



$$\begin{cases} \inf\limits_{u} \int_D |\nabla u| dx \\ u_{|\partial\Omega_I} = f \end{cases}$$

Reformulated as:

$$\begin{cases} \inf\limits_{u,v} \int_\Omega |\nabla u| dx + \frac{1}{2\lambda} \int_\Omega |u - v|^2 dx \\ v_{|\Omega \setminus D} = f \end{cases}$$

How to solve these problems?

• In this lecture we will study a trick to deal with some non-differentiable functions: **Dual and Primal-Dual methods**

• It is based in augmenting the objective function, including new variables, the (auxiliary) dual variables.

Duality. The dual problem

We will first describe how to compute the **dual problem of a given constrained optimization problem.**

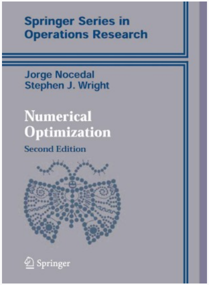
We will study primal, dual and primal-dual formulation of the problems and numerical algorithms that use those formulations to solve them.

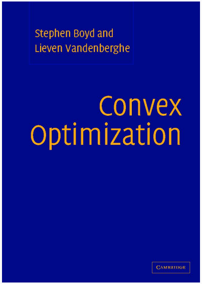
Then, we will solve **non-differentiable problems with dual and primal-dual methods.**

Primal dual methods are an example of interior point methods, that is methods that look for a solution from the interior of the set determined by the constraints.

Bibliography

- Nocedal, J., Wright, S.J., "Numerical Optimization", Springer.
- Boyd, S., Vandenberghe, L., "Convex Optimization", Cambridge University Press.
  http://www.stanford.edu/~boyd/cvxbook/
- Stamford course on Convex Optimization II
  https://web.stanford.edu/class/ee364b/lectures.html





Outline

1. Duality: Min-max Theorem.
2. Lagrangian duality.
3. Primal-dual and dual approaches.
    - Solving the primal/original problem via solving its dual problem.
4. Applications.
5. Non-convex problems and convex relaxation.

Outline

1. **Duality: Min-max Theorem.**
2. Lagrangian duality.
3. Primal-dual and dual approaches.
4. Applications.
5. Non-convex problems and convex relaxation.

Min-max theorem

Min-max theorem. Let $L : X \times Y \to \mathbb{R}$ any function of two variables, $x \in X \subset \mathbb{R}^n$, $y \in Y \subset \mathbb{R}^m$. Always we have

$$\max_y \min_x L(x,y) \leq \min_x \max_y L(x,y),$$

assuming that the minima and maxima exist, otherwise we replace them by inf and sup.

Indeed, Observe that for any x, y we have

$$\min_{\tilde{x}} L(\tilde{x}, y) \leq L(x,y)$$

Take max in y to get

$$\max_y \min_{\tilde{x}} L(\tilde{x}, y) \leq \max_y L(x,y).$$

The left hand side does not depend on x. Take min in x.

Duality gap

The difference

$$\textcolor{red}{DG} := \min_x \max_y L(x,y) - \max_y \min_x L(x,y) \geq 0$$

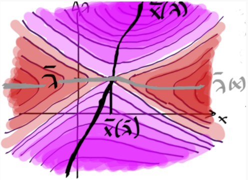
is called the **duality gap**.

If $DG = 0$ and (x_0, y_0) is such that

$$L(x_0,y_0) = \min_{x \in X} L(x, y_0) \quad \text{and} \quad L(x_0,y_0) = \max_{y \in Y} L(x_0, y)$$

then (x_0, y_0) is called a **saddle point**. It satisfies

$$L(x_0,y) \leq L(x_0,y_0) \leq L(x,y_0) \qquad \forall x, y.$$



If there exists a saddle point, then the dual gap is $DG = 0$. **(necessary condition)**

Sufficient condition for a saddle point

Theorem

Assume that X, Y are closed convex sets,

$$\mathbf{x} \in X \to L(\mathbf{x}, y) \qquad \text{is } \textbf{convex} \quad \text{for} \quad \text{all} \quad y \in Y,$$

$$\mathbf{y} \in Y \to L(x, \mathbf{y}) \qquad \text{is } \textbf{concave} \quad \text{for} \quad \text{all} \quad x \in X,$$

and

$$\text{either } X \text{ is bounded or } \exists \bar{y} \in Y \text{ such that } L(x, \bar{y}) \to \infty \text{ as } x \to \infty,$$

$$\text{either } Y \text{ is bounded or } \exists \bar{x} \in X \text{ such that } L(\bar{x}, y) \to -\infty \text{ as } x \to \infty.$$

Then $DG = 0$ and L has a **saddle point** (x_0, y_0) in $X \times Y$.

This result is the **basis of the duality theory**.

Outline

1. Duality: Min-max Theorem.
2. **Lagrangian duality (and saddle points).**
3. Primal-dual and dual approaches.
4. Applications.
5. Non-convex problems and convex relaxation.

Lagrangian Duality

Given differentiable functions $f, c_1, \cdots, c_k, d_1, \ldots, d_r : \mathbb{R}^n \to \mathbb{R}$, let us consider the general optimization problem with inequality and equality constraints

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{subject to} \quad c_1(x) \geq 0, \ldots, c_k(x) \geq 0 \quad \text{(inequality constraints)} \\ & \text{and} \quad d_1(x) = 0, \ldots, d_r(x) = 0 \quad \text{(equality constraints)} \end{aligned} \tag{3}$$

Assume that the set determined by the constraints, that is
$C = \{x \in \mathbb{R}^n \text{ such that } c_1(x) \geq 0, \ldots, c_k(x) \geq 0 \text{ and } d_1(x) = 0, \ldots, d_r(x) = 0\}$,
is non-empty, and suppose that there is a minimum $x_0 \in C$ of (3).

Consider **Lagrange multipliers** λ_i, $i = 1, \ldots, k$, v_j, $j = 1, \ldots, r$ of the problem, where $\lambda \geq 0$ (meaning $\lambda_i \geq 0, \forall i$). The Lagrange multipliers are also called the **dual variables**.

Denoting by $\lambda = (\lambda_1, \ldots, \lambda_k)$, $v = (v_1, \ldots, v_r)$, the **Lagrange function** is

$$\mathscr{L}(x, \lambda, v) = f(x) - \sum_{i=1}^{k} \lambda_i c_i(x) - \sum_{j=1}^{r} v_j d_j(x)$$

Lagrangian Duality

Given

$$\mathscr{L}(x,\lambda,\nu) = f(x) - \sum_{i=1}^{k} \lambda_i c_i(x) - \sum_{j=1}^{r} \nu_j d_j(x)$$

we have

(1) $\mathscr{L}(x,\lambda,\nu) \le f(x) \qquad \forall \lambda \ge 0\ (\lambda_i \ge 0\ \forall i),\ \nu \in \mathbb{R}^r,\ \forall x \in C.$

(2) $\max\limits_{\lambda \ge 0, \nu \in \mathbb{R}^r} \mathscr{L}(x,\lambda,\nu) = \tilde{f}(x) = \begin{cases} f(x) & \text{if}\, x \in C \\ +\infty & \text{if}\, x \notin C \end{cases}$

(3) $\min\limits_{x \in C} f(x) = \min\limits_{x \in \mathbb{R}^n} \bar{f}(x) = \min\limits_{x \in \mathbb{R}^n} \max\limits_{\lambda \ge 0, \nu \in \mathbb{R}^r} \mathscr{L}(x,\lambda,\nu).$

That is

$$\min_{x \in C} = \min_{x \in \mathbb{R}^n} \max_{\lambda \ge 0, \nu \in \mathbb{R}^r} f(x) - \sum_{i=1}^{k} \lambda_i c_i(x) - \sum_{j=1}^{r} \nu_j d_j(x),$$

and **the dual variables satisfy $\lambda \ge 0$ and there is no restriction on ν, i.e., $\nu \in \mathbb{R}^r$.**

The dual problem

We have seen that solving our mininimization problem is equivalent to solve the min-max problem:

$$\min_{x\in C} f(x) = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \mathscr{L}(x,\lambda,\nu).$$

By the Min-max theorem, we know that by changing min-max by max-min, we have

$$\max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\nu) \leq \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \mathscr{L}(x,\lambda,\nu)$$

The duality gap is

$$DG = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \mathscr{L}(x,\lambda,\nu) - \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\nu).$$

The dual problem

Let's assume that $DG = 0$ (in other words, it exists a saddle point).

This is guaranteed if f is convex, $-c_i(x)$ are convex constraints, and d_i are linear constraints: It is a consequence of the Theorem giving the **sufficient conditions**. Obviously, we need some mild assumptions to guarantee the rest of assumptions of the Theorem. Thus,

$$\min_{x\in C} f(x) = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \mathscr{L}(x,\lambda,\nu) = \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\nu).$$

The function

$$\mathbf{g_D}(\lambda,\nu) = \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\nu)$$

is called the Lagrange dual function, or simply the **dual function**.

As $0 = DG = \min_{x\in C} f(x) - \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} g_D(\lambda,\nu)$, the original problem can be re-stated as

$$\min_{x\in C} f(x) = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \mathscr{L}(x,\lambda,\nu) = \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\nu) = \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} g_D(\lambda,\nu). \quad (4)$$

That is,

$$\min_{x\in C} f(x) = \max_{\lambda\geq 0, \nu\in\mathbb{R}^r} g_D(\lambda,\nu).$$

The problem (4) is called the **dual problem** of (3), which is called the **primal problem**, $f(x)$ the **primal function** and $x \in \mathbb{R}^n$ the **primal variable**.

Some times it is much easier to solve the dual problem than the primal one.

Summary

By the *theorem with the sufficient condition*, if f is convex, $-c_i$ convex (c_i concave) and d_i linear, then $DG = 0$ and there is a saddle point (x_0, λ_0, ν_0).

In this case, the following three problems are equivalent:

(1) Primal problem: $\min_{x \in C} f(x)$.

(2) Dual problem: $\max_{\lambda \geq 0, \nu \in \mathbb{R}^r} g_D(\lambda, \nu)$.

(3) Primal-Dual problem: find a sadle point of $\mathscr{L}(x, \lambda, \nu)$.

Example: Computing the dual problem

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. For $x \in \mathbb{R}^n$, consider the problem

$$\begin{aligned} &\min \|x\|^2 \qquad\qquad \text{(P)} \\ &\text{subject to } Ax = b \end{aligned}$$

Let's compute (and solve) its dual problem.

- Let's write problem (P) as a min-max problem and define the duality gap.

$Ax = b$ gives m equality constraints on x: $(Ax)_i = b_i$, $i = 1, \cdots, m$. Therefore, we introduce m Lagrange multipliers (or dual variables), $\nu_1, \ldots, \nu_m \in \mathbb{R}$, and we construct the Lagrangian function, depending on $n+m$ variables

$$\mathscr{L}(x,\nu) = f(x) - \sum_{i=1}^{m} \nu_i((Ax)_i - b_i) = \langle x, x\rangle - \langle \nu, Ax - b\rangle = \langle x, x\rangle - \langle A^t\nu, x\rangle + \langle \nu, b\rangle,$$

where $\nu = (\nu_1, \ldots, \nu_m)^t \in R^m$. Therefore

$$\min_{\text{subject to } Ax=b} \langle x, x\rangle = \min_{x \in R^n} \max_{\nu \in R^m} \mathscr{L}(x,\nu)$$

The duality gap is the difference

$$DG = \min_{x \in R^n} \max_{\nu \in R^m} \mathscr{L}(x,\nu) - \max_{\nu \in R^m} \min_{x \in R^n} \mathscr{L}(x,\nu).$$

which is always ≥ 0.

Example: Computing the dual problem

- Let's now define and compute the dual function of problem (P).

Remember

$$\mathscr{L}(x,\nu) = f(x) - \sum_{i=1}^{m} \nu_i((Ax)_i - b_i) = \langle x, x\rangle - \langle \nu, Ax - b\rangle = \langle x, x\rangle - \langle A^t\nu, x\rangle + \langle \nu, b\rangle,$$

In our case, $DG = 0$ because

- $\mathscr{L}(x,\nu)$ is convex with respect to x (for each ν fixed). Indeed,
  - f is convex because it is a quadratic function which Hessian is equal to $2I$, a strictly positive definite matrix,
  - and $d_i(x) = (Ax)_i - b_i$ are linear constraints, thus $-d_i(x)$ is convex with respect to x.
- $\mathscr{L}(x,\nu)$ is concave with respect to ν (for each x fixed) because it is a linear function on each of the variables ν_j, thus concave with respect to ν.

Therefore, $DG = 0$, there exists a saddle point (x^*, ν^*) and we can change min-max by max-min:

Example: Computing the dual problem

Therefore, $DG = 0$, there exists a saddle point (x^*, ν^*) and we can change min-max by max-min:

$$\min_{\text{subject to } Ax=b} \langle x, x \rangle = \min_{x \in R^n} \max_{\nu \in R^m} \mathscr{L}(x,\nu) = \max_{\nu \in R^m} \min_{x \in R^n} \mathscr{L}(x,\nu) = \max_{\nu \in R^m} g_D(\nu),$$

where

$$g_D(\nu) = \mathscr{L}(x^*(\xi), \xi) \qquad \text{with} \qquad x^*(\xi) = \arg\min_{x \in R^n} \mathscr{L}(x,\nu)$$

is the dual function. We compute the dual function by solving $\min_{x \in R^n} \mathscr{L}(x,\nu)$. As $\mathscr{L}(x,\nu)$ is strictly convex on $x \in \mathbb{R}^n$, a necessary and sufficient condition of (the unique) minimum is $\nabla_x \mathscr{L}(x^*, \nu) = 0$.

In our case, $2x - A^t\nu = 0$, which gives $x^*(\nu) = \frac{1}{2}A^t\nu$. Then,

$$g_D(\nu) = \mathscr{L}(x^*(\nu), \nu) = -\frac{1}{4}\langle A^t\nu, A^t\nu \rangle + \langle \nu, b \rangle$$

Example: Computing the dual problem

- Let's write down the dual problem and solve it.

$$\max_{v\in R^m}\left(-\frac{1}{4}\langle A^t v, A^t v\rangle + \langle v, b\rangle\right) = \max_{v\in R^m}\left(-\frac{1}{4}\langle AA^t v, v\rangle + \langle v, b\rangle\right) =$$

$$= \min_{v\in R^m}\left(\frac{1}{4}\langle AA^t v, v\rangle - \langle v, b\rangle\right).$$

The last function is convex with respect to v as its Hessian is equal to $\frac{1}{2}AA^t$ which is strictly positive definite. Thus, there exists an only minimum which is found by imposing that the gradient is equal to 0. Doing the computations, we obtain the **solution of the dual problem**

$$v^* = 2(AA^t)^{-1}b.$$

Here we have assumed that AA^t is invertible, which might well not to be the case.

Finally, the **solution of the primal problem** is

$$x^* = x^*(v^*) = A^t(AA^t)^{-1}b.$$

Let's verify that satisfies the constraint of the primal problem:

$$Ax^* = AA^t(AA^t)^{-1}b = b.$$

Constrained optimization

Exercise: Let $c \in \mathbb{R}^n$ be a given vector/point of $\mathbb{R}^n$, A be a given $m \times n$ matrix, and $b \in \mathbb{R}^m$. For $x \in \mathbb{R}^n$, consider the problem

$$\begin{aligned} &\min \|x - c\|^2 \\ &\text{subject to } Ax = b \end{aligned}$$

Compute its dual problem. Solve it.

More examples: Exercises in exams of past years

Outline

1. Duality: Min-max Theorem.
2. Lagrangian duality (and saddle points).
3. **Primal-dual and dual approaches (for some non-differentiable problems).**
4. Applications.
5. Non-convex problems and convex relaxation.