# Optimization 1 — Tutorial 11

January 7, 2021

**Dual Problem**

Consider the primal problem

$$
\begin{array}{lll}
(P) & \displaystyle\min_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\
& \text{s.t.} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\
& & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p, \\
& & \mathbf{x} \in X,
\end{array}
$$

where $f, g_i, h_j, : \mathbb{R}^n \to \mathbb{R}$ are functions and $X \subseteq \mathbb{R}^n$. We define the Lagrangian $L \colon X \times \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R}$ of problem $(P)$ as

$$
L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}).
$$

The dual problem is

$$
\begin{array}{lll}
(D) & \displaystyle\max_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^p} & q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\
& \text{s.t.} & (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \operatorname{dom}(q),
\end{array}
$$

where $q \colon \mathbb{R}_+^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ is defined as $q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min\limits_{\mathbf{x}\in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and

$$
\operatorname{dom}(q) = \left\{ (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}_+^m \times \mathbb{R}^p \colon q(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty \right\}.
$$

**Weak Duality**

For $f^*$ the optimal value of $(P)$ and $q^*$ the optimal value of $(D)$ we have $q^* \leq f^*$.

**Strong Duality**

Suppose that

1. $(P)$ is a convex problem.
2. $f^* > -\infty$.
3. Generalized Slater's condition: there exists $\tilde{\mathbf{x}} \in X$ such that $g_i(\tilde{\mathbf{x}}) < 0$ for all $i = 1, 2, \ldots, m$ and $h_j(\tilde{\mathbf{x}}) = 0$ for all $j = 1, 2, \ldots, p$.

Then $f^* = q^*$ and $q^*$ is attained.

**Problem 1**

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be defined by $f(\mathbf{x}) = \sum_{i=1}^{k} \mathbf{x}_{[i]}$ for $1 \le k \le n$, where $\mathbf{x}_{[i]}$ is the $i$-th largest coordinate in the vector $\mathbf{x} \in \mathbb{R}^n$. It is easy to verify that the value $f(\mathbf{x})$ is the value of the optimization problem

$$\begin{aligned} \max_{\mathbf{y}\in\mathbb{R}^n} \quad & \mathbf{x}^T\mathbf{y} \\ \text{s.t.} \quad & \mathbf{e}^T\mathbf{y} = k \\ & 0 \le \mathbf{y} \le \mathbf{e}. \end{aligned}$$

**(a)** For any $\alpha \in \mathbb{R}$, show that $f(\mathbf{x}) \le \alpha$ if and only if there exist $\boldsymbol{\lambda} \in \mathbb{R}^n_+$ and $\mu \in \mathbb{R}$ such that $k\mu + \boldsymbol{\lambda}^T\mathbf{e} \le \alpha$ and $\mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}$.

**(b)** Let $\mathbf{Q} \succ 0$. Find a dual to the problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & f(\mathbf{x}) \le \alpha. \end{aligned}$$

**Solution**

**(a)** We will use duality:

- The primal is a maximization problem, so we look at the Lagrangian of $-f(\mathbf{x})$:

$$L(\mathbf{y}, \boldsymbol{\lambda}, \mu) = -\mathbf{x}^T\mathbf{y} + \mu\left(\mathbf{e}^T\mathbf{y} - k\right) + \boldsymbol{\lambda}^T(\mathbf{y} - \mathbf{e}), \quad \mathbf{y} \ge 0, \boldsymbol{\lambda} \ge 0.$$

- The dual function is

$$\begin{aligned} q(\boldsymbol{\lambda}, \mu) &= \min_{\mathbf{y}\ge 0} L(\mathbf{y}, \boldsymbol{\lambda}, \mu) = \min_{\mathbf{y}\ge 0}\left\{-\mathbf{x}^T\mathbf{y} + \mu\left(\mathbf{e}^T\mathbf{y} - k\right) + \boldsymbol{\lambda}^T(\mathbf{y} - \mathbf{e})\right\} \\ &= \min_{\mathbf{y}\ge 0}\left\{(-\mathbf{x} + \mu\mathbf{e} + \boldsymbol{\lambda})^T\mathbf{y} - \boldsymbol{\lambda}^T\mathbf{e} - k\mu\right\} = \begin{cases} -\boldsymbol{\lambda}^T\mathbf{e} - k\mu, & \mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}, \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

- Thus, the dual problem is

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^n, \mu\in\mathbb{R}} \quad & -\boldsymbol{\lambda}^T\mathbf{e} - k\mu \\ \text{s.t.} \quad & \mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}, \\ & \boldsymbol{\lambda} \ge 0. \end{aligned}$$

$\Longleftarrow$: Assume that there exist $\boldsymbol{\lambda} \in \mathbb{R}^n_+$ and $\mu \in \mathbb{R}$ such that $k\mu + \boldsymbol{\lambda}^T\mathbf{e} \le \alpha$ and $\mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}$. So there is a feasible solution to the dual, and from weak duality $-\alpha \le -\boldsymbol{\lambda}^T\mathbf{e} - k\mu \le -f(\mathbf{x})$.

$\Longrightarrow$: Assume that $f(\mathbf{x}) \le \alpha$. Notice that strong duality holds:

1. The primal problem is linear (thus convex).
2. The optimal value is finite since it is attained (Weierstrass theorem).
3. Generalized Slater's condition is satisfied (for example for $\mathbf{y}_i = 1$ for all $1 \le i \le k$ and 0 otherwise).

Therefore, there exist $\boldsymbol{\lambda} \in \mathbb{R}^n_+$ and $\mu \in \mathbb{R}$ such that $-k\mu - \boldsymbol{\lambda}^T\mathbf{e} = -f(\mathbf{x}) \ge -\alpha$ and $\mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}$ (optimal of the dual is attained).

**(b)** We use section (a):

- The problem is equivalent to

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n, \boldsymbol{\lambda}\in\mathbb{R}^n, \mu\in\mathbb{R}} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & k\mu + \boldsymbol{\lambda}^T\mathbf{e} \le \alpha, \\ & \mu\mathbf{e} + \boldsymbol{\lambda} \ge \mathbf{x}, \\ & \boldsymbol{\lambda} \ge 0. \end{aligned}$$

- The Lagrangian is

$$L(\mathbf{x}, \boldsymbol{\lambda}, \mu, \mathbf{y}, \eta) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{y}^T(\mathbf{x} - \mu\mathbf{e} - \boldsymbol{\lambda}) + \eta\left(k\mu + \mathbf{e}^T\boldsymbol{\lambda} - \alpha\right), \quad \mathbf{y}, \boldsymbol{\lambda} \geq 0, \eta \geq 0.$$

- The dual objective is

$$\begin{aligned} q(\mathbf{y}, \eta) &= \min_{\mathbf{x}\in\mathbb{R}^n, \boldsymbol{\lambda}\in\mathbb{R}^n_+, \mu\in\mathbb{R}} L(\mathbf{x}, \boldsymbol{\lambda}, \mu, \mathbf{y}, \eta) \\ &= \min_{\mathbf{x}\in\mathbb{R}^n, \boldsymbol{\lambda}\in\mathbb{R}^n_+, \mu\in\mathbb{R}} \left\{\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{y}^T(\mathbf{x} - \mu\mathbf{e} - \boldsymbol{\lambda}) + \eta\left(k\mu + \mathbf{e}^T\boldsymbol{\lambda} - \alpha\right)\right\} \\ &= \min_{\mathbf{x}\in\mathbb{R}^n} \left\{\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{y}^T\mathbf{x}\right\} + \min_{\boldsymbol{\lambda}\in\mathbb{R}^n_+} (\eta\mathbf{e} - \mathbf{y})^T\boldsymbol{\lambda} + \min_{\mu\in\mathbb{R}} \mu\left(k\eta - \mathbf{y}^T\mathbf{e}\right) - \alpha\eta, \end{aligned}$$

  where the last equality follows from separability w.r.t. the primal variables $\mathbf{x}, \boldsymbol{\lambda}, \mu$.
- We solve the above:
  - $\min_{\mathbf{x}\in\mathbb{R}^n} \left\{\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{y}^T\mathbf{x}\right\} = \left(-\frac{1}{2}\mathbf{Q}^{-1}\mathbf{y}\right)^T\mathbf{Q}\left(-\frac{1}{2}\mathbf{Q}^{-1}\mathbf{y}\right) + \mathbf{y}^T\left(-\frac{1}{2}\mathbf{Q}^{-1}\mathbf{y}\right) = -\frac{1}{2}\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y}$ (unconstrained minimization of a continuously differentiable convex function, thus any stationary point is a global minimizer).
  - $\min_{\boldsymbol{\lambda}\in\mathbb{R}^n_+} (\eta\mathbf{e} - \mathbf{y})^T\boldsymbol{\lambda} = \begin{cases} 0, & \eta\mathbf{e} - \mathbf{y} \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$
  - $\min_{\mu\in\mathbb{R}} \mu\left(k\eta - \mathbf{y}^T\mathbf{e}\right) = \begin{cases} 0, & k\eta - \mathbf{y}^T\mathbf{e} = 0, \\ -\infty, & \text{otherwise.} \end{cases}$
- Therefore, the dual is

$$\begin{aligned} \max_{\mathbf{y}\in\mathbb{R}^n, \eta\in\mathbb{R}} \quad & -\frac{1}{2}\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y} - \alpha\eta \\ \text{s.t.} \quad & \eta\mathbf{e} - \mathbf{y} \geq 0, \\ & k\eta - \mathbf{y}^T\mathbf{e} = 0, \\ & \mathbf{y} \geq 0, \\ & \eta \geq 0. \end{aligned}$$

**Problem 2**

Consider the primal optimization problem

$$\begin{aligned} \min_{x,y\in\mathbb{R}} \quad & x^4 - 2y^2 - y \\ \text{s.t.} \quad & x^2 + y^2 + y \leq 0. \end{aligned}$$

**(a)** Is the problem convex?

**(b)** Does there exist an optimal solution to the problem?

**(c)** Write a dual problem and solve the dual problem.

**(d)** Is the optimal value of the dual problem equal to the optimal value of the primal problem? Find the optimal solution of the primal problem.

**Solution**

**(a)** The problem is non-convex:

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 12x^2 & 0 \\ 0 & -4 \end{pmatrix} \Longrightarrow \nabla^2 f(0, y) = \begin{pmatrix} 0 & 0 \\ 0 & -4 \end{pmatrix} \prec 0.$$

Since the line $(0, y)$ is feasible when $y \in [-1, 0]$, we have that the problem is non-convex (there are feasible points with ND Hessian matrix).

**(b)** We need to show that the feasible set is bounded. We have $x^2 + y^2 + y = x^2 + \left(y + \frac{1}{2}\right)^2 - \frac{1}{4}$ and we get a non-empty ball constraint.

**(c)** The Lagrangian is

$$L(\mathbf{x}, \lambda) = x^4 - 2y^2 - y + \lambda\left(x^2 + y^2 + y\right), \quad \lambda \geq 0.$$

- From separability w.r.t. the primal variables $x, y$ we can write

$$q(\lambda) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \lambda) = \min_{x\in\mathbb{R}} x^2\left(x^2 + \lambda\right) + \min_{y\in\mathbb{R}}\left\{(\lambda - 2)\, y^2 + (\lambda - 1)\, y\right\}.$$

- Since $\lambda \geq 0$ we have $x^2\left(x^2 + \lambda\right) \geq 0$, and $x = 0$ is the optimal solution with value 0.
- The problem with respect to $y$ is an unconstrained minimization problem. Therefore, if a global minimum exists, then it is attained in a stationary point.
  - We notice that if $\lambda \leq 2$ then this is a minimization of a concave function – thus no global minimum exists: we can take $y \to -\infty$ or $y \to \infty$ and see that $(\lambda - 2)\, y^2 + (\lambda - 1)\, y \to -\infty$.
  - If $\lambda > 2$ then this is a strictly convex function with a unique stationary point $y = \frac{1-\lambda}{2(\lambda-2)}$ with an objective value $-\frac{(1-\lambda)^2}{4(\lambda-2)}$.
  - Ovewrall

$$\min_{y\in\mathbb{R}}\left\{(\lambda - 2)\, y^2 + (\lambda - 1)\, y\right\} = \begin{cases} -\infty, & \lambda \leq 2, \\ -\frac{(1-\lambda)^2}{4(\lambda-2)}, & \lambda > 2. \end{cases}$$

- Therefore, the dual is

$$\begin{aligned} \max_{\lambda\in\mathbb{R}} \quad & -\frac{(1-\lambda)^2}{4(\lambda - 2)} \\ \text{s.t.} \quad & \lambda > 2. \end{aligned}$$

- To solve the dual, we notice that it is a continuously differentiable concave function over an open domain – therefore, stationarity is a sufficient condition for optimality. The two stationary points are $\lambda = 1$ and $\lambda = 3$, and therefore the solution is $\lambda = 3$ (the only feasible point), with objective value of $-1$.

**(d)** Since the primal problem is non-convex, there is no guarantee for strong duality. So we need to solve the primal problem in order to check for strong duality.

- The objective satisfies $x^4 - 2y^2 - y \geq -2y^2 - y$. Additionally

$$\left\{y \in \mathbb{R} \colon x^4 + y^2 + y \leq 0\right\} \subseteq \left\{y \in \mathbb{R} \colon y^2 + y \leq 0\right\}.$$

- So the optimal solution (if exists) of

$$\begin{aligned} (P') \quad \min_{x,y\in\mathbb{R}} \quad & -2y^2 - y \\ \text{s.t.} \quad & y^2 + y \leq 0, \end{aligned}$$

  is a lower bound on the optimal solution of the primal. Meaning, if $(P)$ attains the optimal value of $(P')$, then this is also an optimal value of $(P)$.
- We solve $(P')$:
  - Since $y^2 + y \leq 0$ if and only if $y \in [-1, 0]$, then the solution is attained in a stationary point or on the boundary.
  - Since the objective is concave, the minimizer is attained at the boundary.
  - $f(0) = 0 > f(-1) = -1$. Therefore, $(P')$ attains a minimum at $y = -1$, and from separability $(P)$ attains a minimum at $(0, -1)$ with value $-1$.
- Strong duality indeed holds.

**Problem 3 (HW9, Problem 5(a))**

Consider the problem

$$(P_\alpha) \qquad \min_{\mathbf{x}\in\mathbb{R}^n} \quad \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$$
$$\text{s.t.} \quad \mathbf{e}^T\mathbf{x}=\alpha,$$

where $\mathbf{A}\in\mathbb{R}^{m\times n}$, $\mathbf{b}\in\mathbb{R}^m$, $\mathbf{e}=(1,1,...,1)^T\in\mathbb{R}^n$ and $\alpha\in\mathbb{R}$ is a parameter. Prove that $(P_\alpha)$ has a unique solution if and only if $\operatorname{Ker}(\mathbf{A})\cap\operatorname{Ker}\left(\mathbf{e}^T\right)=\{\mathbf{0}_n\}$.

**Solution**

- The problem is a continuously differentiable, convex and generalized Slater's condition is satisfied. So {optimal solutions} = {KKT points}. It is enough to find a unique feasible KKT point.
- The Lagrangian is

$$L(\mathbf{x},\mu)=\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2+\mu\left(\mathbf{e}^T\mathbf{x}-\alpha\right), \quad \mu\in\mathbb{R}.$$

- The KKT conditions are

$$\begin{cases} 2\mathbf{A}^T\mathbf{A}\mathbf{x}-2\mathbf{A}^T\mathbf{b}+\mu\mathbf{e}=\mathbf{0}_n & (i)\\ \mathbf{e}^T\mathbf{x}=\alpha & (ii)\end{cases}$$

- Writing the conditions in matrix form

$$\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mu\end{pmatrix}=\begin{pmatrix}2\mathbf{A}^T\mathbf{b}\\ \alpha\end{pmatrix}.$$

- There is a unique feasible point iff $\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}$ is invertible.
- $\Longrightarrow$: Assume $\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}$ is invertible.
  - If $\mathbf{v}\in\operatorname{Ker}(\mathbf{A})\cap\operatorname{Ker}\left(\mathbf{e}^T\right)$ then $2\mathbf{A}^T\mathbf{A}\mathbf{v}=\mathbf{0}_n$ and $\mathbf{e}^T\mathbf{v}=0$.
  - Meaning $\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{v}\\ 0\end{pmatrix}=\mathbf{0}_{n+1}$ and we must have $\mathbf{v}=\mathbf{0}_n$.
- $\Longleftarrow$: Assume $\operatorname{Ker}(\mathbf{A})\cap\operatorname{Ker}\left(\mathbf{e}^T\right)=\mathbf{0}_n$. Assume on the contrary that $\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}$ is not invertible.
  1. It is not invertible iff there exists $(\mathbf{v},t)\neq\mathbf{0}_{n+1}$ such that $\begin{pmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{e}\\ \mathbf{e}^T & 0\end{pmatrix}\begin{pmatrix}\mathbf{v}\\ t\end{pmatrix}=\mathbf{0}_{n+1}$.
  2. This implies $2\mathbf{A}^T\mathbf{A}\mathbf{v}+t\mathbf{e}=\mathbf{0}_n$ and $\mathbf{e}^T\mathbf{v}=0$. So $\mathbf{v}\in\operatorname{Ker}\left(\mathbf{e}^T\right)$.
  3. From 2 we have $2\mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v}+2t\mathbf{e}^T\mathbf{v}=\mathbf{0}_n$ and therefore $\mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v}=0$ which means $\mathbf{v}\in\operatorname{Ker}(\mathbf{A})$.
  4. From the assumption $\mathbf{v}=\mathbf{0}_n$ and so $t\neq0$.
  5. From 1 we derive $t\mathbf{e}=\mathbf{0}_n$, so $t=0$ which contradicts 4.