Exercises for Lecture Course on Numerical Optimization (NUMOPT)
Albert-Ludwigs-Universität Freiburg – Winter Term 2023/2024

Exercise 2: Duality and Fitting Problems

Prof. Dr. Moritz Diehl, Dimitris Kouzoupis, Florian Messerer, Yizhen Wang

1. **Lagrange duality and dual problems:**

   (a) Consider the following *logarithmic barrier* problem,

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^T x - \sum_{j=1}^{n} \log x_j \\ \text{s.t.} \quad & a^T x = b, \end{aligned}$$

where $a, c \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Remark 1: Problems using a logarithmic barrier as the one above will be at the core of interior point methods that we will analyze later in this course.

Remark 2: $\log x_j$ *is only defined for* $x_j \in \mathbb{R}_{++}$. *For simplicity, and without discussing this further here, we will assume that* $-\log x_j$ *takes the value* $+\infty$ *whenever* $x_j \in \mathbb{R}_-$.

   i. Derive the explicit form of the dual of this problem.
   ii. Does strong duality hold?

Solution: the Lagrangian of the problem reads

$$\mathcal{L}(x,\lambda) := c^\top x - \sum_{j=1}^{n} \log x_j - \lambda(a^\top x - b) = (c^\top - \lambda a^\top)x - \sum_{j=1}^{n} \log x_j + \lambda b,$$

with $\lambda \in \mathbb{R}$. The dual function is

$$q(\lambda) = \inf_{x \,\in\, \mathbb{R}^n} \mathcal{L}(x,\lambda),$$

If $c_j - \lambda a_j \leq 0$ for some component j, the Lagrangian is unbounded below, so we will impose the condition

$$c_j - \lambda a_j > 0, \quad \forall j = 0, \dots, n.$$

Taking the derivative of the Lagrangian with respect to x, we have

$$\nabla_x \mathcal{L}(x,\lambda) = c - X^{-1}\mathbf{1} - \lambda a,$$

with $X = \operatorname{diag}(x)$. The j-th component of gradient of the Lagrangian vanishes for

$$x_j^* = \frac{1}{c_j - \lambda a_j},$$

where, in order to derive such condition, we have to impose $c_j - \lambda a_j > 0$, which is anyway required for the Lagrangian function to be unbounded. As $\mathcal{L}(x,\lambda)$ is convex in x, such conditions attain the global minimum of the Lagrangian over $\mathbb{R}^n$. We therefore get ‘

$$q(\lambda) = \lambda b + \begin{cases} c^\top x^* - \sum_{j=1}^{n} \log x_j^* + \lambda a^\top x^* & \text{if } c_j - \lambda a_j > 0 \text{ for } i = 1,\dots,n \\ -\infty & \text{else} \end{cases}$$

Substituting the expression for x^*, we obtain the dual problem:

$$\begin{aligned} \sup_{\lambda\in\mathbb{R}} \quad & n + \lambda b + \sum_{j=1}^{n} \log(c_j - \lambda a_j) \\ \text{s.t.} \quad & c - \lambda a > 0. \end{aligned}$$

(b) Consider the following *mixed-integer quadratic program* (MIQP):

$$\begin{aligned} \min_{x \in \{0,1\}^n} \quad & x^T Q x + q^T x \\ \text{s.t.} \quad & Ax \geq b, \end{aligned}$$

where $Q \in \mathbb{R}^{n \times n}$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. where the optimization variables x_i are restricted to take values in $\{0, 1\}$. Solving mixed-integer problems is in general a challenging task, thus it is common practice to exploit continuous reformulations as the following:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & x^T Q x + q^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i(1 - x_i) = 0 \qquad i = 1, \ldots, n. \end{aligned}$$

i. Is this reformulation convex?
Solution: no, it has nonlinear equality constraints, hence is not convex.

ii. A lower bound to the optimal solution can be computed by solving the (convex) dual problem (not required here). Derive the explicit form of the dual of the continuous reformulation.
Solution: the Lagrangian of the problem reads

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= x^\top Q x + q^\top x - \mu^\top (Ax - b) - \lambda^\top X(1 - x) \\ &= x^\top (Q + \Lambda) x + (q^\top - \mu^\top A - \lambda^\top) x + \mu^\top b, \end{aligned}$$

where $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^m$, and $\Lambda = \operatorname{diag} \lambda$. This is unbounded below in x if $Q + \Lambda \prec 0$ or $Q + \Lambda \succeq 0$ and $q - \lambda - A^\top \mu \in \mathcal{N}(Q + \Lambda)$. (with $\mathcal{N}(Q + \Lambda)$ denoting the null space). Otherwise the minimum is attained for

$$x_*(\lambda, \mu) = -\frac{1}{2}(Q + \Lambda)^+ (q - \lambda - A^T \mu).$$

The dual function is therefore

$$q(\lambda, \mu) = \begin{cases} x_*^\top (Q + \Lambda) x_* + (q^\top - \mu^\top A - \lambda^\top) x_* + \mu^\top b & \text{if } \begin{array}{l} Q + \Lambda \succeq 0 \\ q - \lambda - A^\top \eta \in \mathcal{R}(Q + \Lambda) \end{array} \\ -\infty & \text{else,} \end{cases}$$

with $\mathcal{R}(\cdot)$ denoting the row-space. The dual problem is then given as

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^n, \mu \in \mathbb{R}^m} \quad & x_*^\top (Q + \Lambda) x_* + (q^\top - \mu^\top A - \lambda^\top) x_* + \mu^\top b \\ \text{s.t.} \quad & \mu \geq 0, \\ & Q + \Lambda \succeq 0, \\ & q - \lambda - A^\top \eta \in \mathcal{R}(Q + \Lambda), \end{aligned}$$

where the dependency of x_* depends on λ and μ is dropped for notational brevity.

2. **Regularized linear least squares:** Given a matrix $J \in \mathbb{R}^{m \times n}$, a symmetric positive definite matrix $Q \succ 0$, a vector of measurements $\eta \in \mathbb{R}^m$ and a point $\bar{x} \in \mathbb{R}^n$, compute the limit:

$$\lim_{\substack{\alpha \to 0 \\ \alpha > 0}} \arg\min_x \frac{1}{2}||\eta - Jx||_2^2 + \frac{\alpha}{2}(x - \bar{x})^\top Q(x - \bar{x}). \tag{1}$$

Hint: Use matrix square root and Lemma 6.1 from the lecture notes.

Q is symmetric positive definite and therefore has unique square root $Q^{\frac{1}{2}}$, such that $Q = Q^{\frac{1}{2}\top} Q^{\frac{1}{2}}$
$\frac{1}{2}\|\eta - Jx\|_2^2 + \frac{\alpha}{2}(x - \bar{x})^\top Q(x - \bar{x})$
$= \frac{1}{2}\|\eta - Jx\|_2^2 + \frac{\alpha}{2}(x - \bar{x})^\top Q^{\frac{1}{2}\top} Q^{\frac{1}{2}}(x - \bar{x}) = \frac{1}{2}\|\eta - Jx\|_2^2 + \frac{\alpha}{2}\|\underbrace{Q^{\frac{1}{2}}(x - \bar{x})}_{=:y}\|_2^2 = \dots$

Substitute $y := Q^{\frac{1}{2}}(x - \bar{x}) \Leftrightarrow x = Q^{-\frac{1}{2}}y + \bar{x}$
$\dots = \frac{1}{2}\|\underbrace{\eta - J\bar{x}}_{\widetilde{\eta}} - \underbrace{JQ^{-\frac{1}{2}}}_{\widetilde{J}} y\|_2^2 + \frac{\alpha}{2}\|y\|_2^2 = \frac{1}{2}\|\widetilde{\eta} - \widetilde{J}y\|_2^2 + \frac{\alpha}{2}\|y\|_2^2$

This is now the same form as problem (6.20) in the lecture notes (p. 46), and to obtain the limit in a clean way we can follow the steps outlined in the proof of Lemma 6.1.

$$\lim_{\substack{\alpha \to 0 \\ \alpha > 0}} \arg\min_y \frac{1}{2}\|\widetilde{\eta} - \widetilde{J}y\|_2^2 + \frac{\alpha}{2}\|y\|_2^2 = \lim_{\substack{\alpha \to 0 \\ \alpha > 0}} (\widetilde{J}^\top \widetilde{J} + \alpha I)^{-1} \widetilde{J}^\top \widetilde{\eta} \stackrel{\text{Lem. 6.1}}{=} J^\dagger \widetilde{\eta}$$

with $\widetilde{J}^\dagger$ the Moore Penrose Pseudo inverse. So in the limit we obtain $y^* = J^\dagger \widetilde{\eta}$, and substituting back, $x^* = Q^{-\frac{1}{2}}y^* + \bar{x}$.

3. **Linear L_2 fitting:** Assume we have modeled the dependency of some output $y \in \mathbb{R}$ on some input $x \in \mathbb{R}$ as the linear model $y = ax + b$ with parameters $a, b \in \mathbb{R}$. The value of these parameters is unknown, but we have a data set of N noisy measurements $(x_i, \tilde{y}_i)$, $i = 1, \dots, N$. These measurements are obtained as $\tilde{y}_i = ax_i + b + \eta_i$, where η_i is noise drawn from a normal distribution with zero mean and variance one, $\eta_i \sim \mathcal{N}(0, 1)$.

   One way of finding an estimate of the parameter values is to minimize a least-squares loss of the residuals $ax_i + b - \tilde{y}_i$, which can be formulated as the optimization problem

$$\min_{a,b \in \mathbb{R}} \sum_{i=1}^{N} \frac{1}{2}(ax_i + b - \tilde{y}_i)^2 = \min_{a,b} \frac{1}{2} \left\| J \begin{bmatrix} a \\ b \end{bmatrix} - \tilde{y} \right\|_2^2, \tag{2}$$

   where $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_N)$ and it will be part of the exercise to define J. As discussed in the lecture, the optimal solution of (2) can be calculated explicitly by solving the linear system

$$J^\top J \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = J^\top \tilde{y}, \tag{3}$$

   where $\hat{a}$, $\hat{b}$ are the resulting estimates of the parameter values.

   (a) Define J by writing it down on paper.
   $J = \begin{bmatrix} x & \mathbf{1} \end{bmatrix} \in \mathbb{R}^{N \times 2}$, where $x = (x_1, \dots, x_N)$, and $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$

   (b) Generate the problem data. Take $N = 30$ and generate $x = (x_1, \dots, x_N)$ as N equally spaced points in the interval $[0, 5]$ and, for $i = 1, \dots, N$, generate the measurements as $\tilde{y}_i = 3x_i + 4 + \eta_i$, where η_i is sampled from the normal distribution $\mathcal{N}(0, 1)$. Plot the results.
   Hint: look up the `linspace` and `randn` commands, e.g., via NumPy documentation (Python) / using `help` or `doc` command (MATLAB). If you want a reproducible 'random' sequence, you can use `rng`.

(c) Calculate the estimates $\hat{a}, \hat{b}$ in MATLAB using Equation (3) and plot the obtained line in the same graph as the measurements.

(d) Introduce 3 outliers in $\tilde{y}$ by replacing arbitrary measurements and plot the new fitted line in your plot.

You will need the measurements $\tilde{y}$ (both with and without outliers) and the matrix J for the next task.

4. **Linear L_1 fitting:** In this task we want to fit a line to the same set of measurements, but we use a different cost function:

$$\min_{a,b\in\mathbb{R}} \sum_{i=1}^{N} |(ax_i + b - \tilde{y}_i)|. \tag{4}$$

   (a) Problem (4) is not differentiable. Find an (equivalent) smooth reformulation.
   Hint 1: Introduce slack variables $s_1, \dots, s_N \in \mathbb{R}$ as additional decision variables.
   Hint 2: The resulting problem will be a Linear Program (LP).

$$\begin{aligned} \min_{a,b\in\mathbb{R},\, s\in\mathbb{R}^N} \quad & \sum_{i=1}^{N} s_i \\ \text{s.t.} \quad & -s_i \le ax_i + b - \tilde{y}_i \le s_i \quad i = 1,\dots,N \end{aligned}$$

   (b) The result of the previous task is a LP. In order to solve it with `linprog`, the native LP solver of MATLAB, we need to bring it to the form:

$$\min_{z\in\mathbb{R}^n} f^T z \tag{5a}$$

$$\text{s.t. } Az \le b \tag{5b}$$

$$Cz = d \tag{5c}$$

$$l_z \le z \le u_z, \tag{5d}$$

   Define matrices A, C and vectors f, b, d, l_z, u_z by writing them down on paper. You may not need all of these. In this case you can define them as 'empty'. Order your variables as $z = (a, b, s_1, \dots, s_N)$. Use matrix J from the previous exercise to define A.

$$f = \begin{bmatrix} 0 \\ 0 \\ \mathbf{1} \end{bmatrix}, \quad A = \begin{bmatrix} J & -I \\ -J & -I \end{bmatrix}, \quad b = \begin{bmatrix} \tilde{y} \\ -\tilde{y} \end{bmatrix}$$

   (c) Solve the problem with `linprog` (SciPy / MATLAB). Use the measurements $\tilde{y}$ from the previous exercise (both with and without outliers) and plot the results against those of the L2 fitting. Which norm performs better?

   The L1 norm is more robust against the outliers (as it does not penalize the model-measurement-mismatch quadratically). Which norm performs better depends on the context, but here it seems like we want our method to 'ignore' the outliers (the outliers seem nonsensical). That means L1 performs better.

   (d) Solve the problem resulting from task 4a with CasADi and compare the results.

   Should be identical.