

HW #5

Exercise 1 (Minimizing a sum of logarithms) Consider the following problem:

$$\begin{aligned} p^* = \max_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^n \alpha_i \ln x_i \\ \text{s.t.} \quad & x \geq 0, \quad \mathbf{1}^\top x = c, \end{aligned}$$

where $c > 0$ and $\alpha_i > 0$, $i = 1, \dots, n$. Problems of this form arise, for instance, in maximum-likelihood estimation of the transition probabilities of a discrete-time Markov chain. Determine in closed-form a minimizer, and show that the optimal objective value of this problem is

$$p^* = \alpha \ln(c/\alpha) + \sum_{i=1}^n \alpha_i \ln \alpha_i,$$

where $\alpha \doteq \sum_{i=1}^n \alpha_i$.

Exercise 2 (KKT conditions) Consider the optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^n \left(\frac{1}{2} d_i x_i^2 + r_i x_i \right) \\ \text{s.t.} \quad & a^\top x = 1, \quad x_i \in [-1, 1], \quad i = 1, \dots, n, \end{aligned}$$

where $a \neq 0$ and $d > 0$.

In this exercise, we will use the KKT conditions and/or the Lagrangian to come up with a fast algorithm to solve this optimization problem. Recall that the KKT conditions, as defined in Lecture 20, are:

- Primal feasibility: $x \in \mathcal{D}, f_i(x) \leq 0, i = 1, \dots, m$
- Dual feasibility: $\lambda \geq 0$
- Complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$
- Lagrangian stationarity: $x \in \arg \min \mathcal{L}(\cdot, \lambda)$

1. Verify that strong duality holds for this problem, and write down the KKT optimality conditions.

Note: For simplicity, please assume that a is such that the constraint set is strictly feasible.

Hint: For algebraic convenience throughout the problem, we recommend you express the constraint $x_i \in [-1, 1]$ as $x_i^2 \leq 1$.

2. The dual problem amounts to maximizing the dual function $g(\lambda, \mu)$ w.r.t. μ and $\lambda \geq 0$. Write this dual function.
3. Find the maximum of $g(\lambda, \mu)$ w.r.t. λ , for fixed μ . In particular, write down a closed-form expression for the corresponding optimal point $\lambda_i^*(\mu)$.
4. Using your previous results, express the optimal primal point $x_i^*(\mu)$ as a function of the scalar dual variable μ only.

The optimal μ is the value which makes the constraint $\sum_i a_i x_i^*(\mu) = 1$ satisfied. We can find this value by bisection over μ . Since the gradient of $g(\lambda^*, \mu)$ with respect to μ is simply given by $a^\top x_i^*(\mu) - 1$, we increase μ when the gradient is positive, and decrease it when it is negative. In practice, we initialize two values $\mu_l < 0, \mu_r > 0$ for which we know a priori that $\mu^* \in [\mu_l, \mu_r]$, and execute the following bisection algorithm.

- (a) Set $\mu = (\mu_r + \mu_l)/2$

- (b) Evaluate $h = a^\top x^*(\mu) - 1$
 - (c) If $h > 0$, let $\mu_l = \mu$, else let $\mu_r = \mu$
 - (d) If $|\mu_r - \mu_l| \leq \epsilon$ or $h = 0$, exit and return μ , else goto (a).
5. Let l be the length of the initial localization interval. Provide a simple upper bound on the number of iterations needed to reach a value of μ within ϵ from the true optimum μ^* .
 6. In the notebook `kkt_conditions.ipynb`, implement the bisection algorithm and compare it with the solution obtained via CVX (or another solver of your choice). Recommended values for μ_l, μ_r, ϵ are provided for you in the notebook.

In the following exercise, you will study through another example how to find a proper initialization for μ_l and μ_r .

Exercise 3 (Trust region subproblem) Consider the problem

$$p^* = \min_x x^\top Q x + 2c^\top x \quad : \quad \|x\|_2 = 1.$$

where $Q \in \mathbb{S}^n$ is symmetric positive semi-definite and $c \in \mathbb{R}^n$.

1. Is the problem, as stated, convex?
2. Show that the problem can be reduced to

$$p^* = \min_y \sum_{i=1}^n (\lambda_i y_i^2 + 2d_i y_i) \quad : \quad \sum_{i=1}^n y_i^2 = 1,$$

for appropriate vectors $\lambda, d \in \mathbb{R}^n$, which you will determine as functions of the problem data.

3. Show that the problem can be further reduced to the convex problem

$$p^* = \min_z \sum_{i=1}^n (\lambda_i z_i - 2|d_i|\sqrt{z_i}) \quad : \quad \sum_{i=1}^n z_i = 1, \quad z \geq 0.$$

4. Express the dual of the above problem as one with a single variable.
5. We can solve the one-dimensional dual problem by bisection on the dual variable (see previous exercise for an example of bisection). This requires finding an initial interval in which any optimal dual variable lies. To obtain fast convergence, we want the length of this interval to be as small as possible. Find an initial interval with length no more than $2\|d\|_2$.

Hint: Introduce variables η, δ_i and ϕ^* such that $\eta = \lambda_n - \nu \geq 0, \delta_i = \lambda_i - \lambda_n \geq 0, \phi^* = \lambda_n - p^*$ (or depending on your problem formulation, $\eta = \lambda_n + \nu \geq 0, \delta_i = \lambda_i - \lambda_n, \phi^* = p^* + \lambda_n$). Then, try to bound η^* with the problem data.

6. From now on, we assume we computed the optimal ν (one could implement the algorithm from the previous exercise). How do you recover an optimal primal variable x ?