EE 127 / EE 227AT 10/25/18
L. El Ghaoui

HW #5 Solutions

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

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

where $c > 0$ and $\alpha_i > 0$, $i = 1, \ldots, n$. 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 $\alpha \doteq \sum_{i=1}^n \alpha_i$.

Solution 1 *Note: This solution makes choices on modifying the formulation of the problem. These are not necessary and we accepted solutions that do not rely on that. Notably, it is not necessary to turn* max *into* min *or to express the equality* $\mathbf{1}^T x = c$ *as the inequality* $\mathbf{1}^T x \leq c$.

Let us consider the equivalent problem

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

Since the objective is strictly decreasing over $x \geq 0$ and $\mathbf{1}^\top x$ is nondecreasing over $x \geq 0$, we can replace the equality constraint by an inequality one, thus we consider the problem

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

The partial Lagrangian for this problem is

$$\begin{aligned} \mathcal{L}(x, \mu) &= \sum_{i=1}^n \alpha_i \ln 1/x_i + \mu(\mathbf{1}^\top x - c) \\ &= \sum_{i=1}^n (\alpha_i \ln 1/x_i + \mu x_i) - \mu c, \end{aligned}$$

and, for $\mu \geq 0$,

$$\begin{aligned} g(\mu) &= \min_{x \geq 0} \mathcal{L}(x, \mu) = -\mu c + \sum_{i=1}^{n} \min_{x_i \geq 0} \left(\alpha_i \ln 1/x_i + \mu x_i\right) \\ &= -\mu c + \sum_{i=1}^{n} \left(\alpha_i \ln(\mu/\alpha_i) + \alpha_i\right) \\ &= -\mu c + \ln \mu \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \alpha_i (1 - \ln \alpha_i). \end{aligned}$$

The minimum with respect to $x_i \geq 0$ in the first expression is attained at the unique point $x_i = \alpha_i/\mu \geq 0$, which we obtain by verifying that the expression is convex with respect to x and setting the gradient to 0.

The dual is thus $d^* = \max_{\mu \geq 0} g(\mu)$. Since the primal problem is strictly feasible, strong duality holds. The optimal dual solution is obtained as

$$\mu^* = \frac{\sum_{i=1}^{n} \alpha_i}{c} = \frac{\alpha}{c},$$

from which we obtain the optimal primal solution as

$$x_i^* = \frac{\alpha_i}{\mu^*} = \frac{c\alpha_i}{\alpha}, \quad i = 1, \ldots, n.$$

The expression for the optimal objective value follows by substituting this optimal solution back into the objective:

$$\begin{aligned} p^* &= \sum_{i=1}^{n} \alpha_i \ln \left(\frac{c\alpha_i}{\alpha}\right) \\ &= \sum_{i=1}^{n} \alpha_i \ln \left(\frac{c}{\alpha}\right) + \alpha_i \ln \alpha_i \\ &= \alpha \ln \left(\frac{c}{\alpha}\right) + \alpha_i \ln \alpha_i. \end{aligned}$$

Exercise 2 (KKT conditions) Consider the optimization problem

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

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: $x \in \mathcal{D}, f_i(x) \leq 0, i = 1, \ldots, m$
- Dual feasibility: $\lambda \geq 0$
- Complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \ldots, m$
- 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.

   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.

Solution 2 (KKT conditions — Suvrit Sra, 2013)

1. Write the interval constraints on x_i as $x_i^2 \leq 1$, and define $D = \operatorname{diag}(d_1, \ldots, d_n)$, $\Lambda = \operatorname{diag}(\lambda)$, $\lambda = (\lambda_1, \ldots, \lambda_n)$, $r = (r_1, \ldots, r_n)$. The problem is a QCQP of the form

$$\begin{aligned} \min_x \quad & \tfrac{1}{2} x^\top D x + r^\top x \\ \text{s.t.} \quad & a^\top x = 1 \\ & x_i^2 \leq 1 \qquad i = 1, \ldots, n. \end{aligned}$$

Strong duality holds for this problem, due to satisfaction of the Slater's conditions. The Lagrangian of the problem is

$$\mathcal{L}(x, \lambda, \mu) = \frac{1}{2} x^\top (D + \Lambda) x + (r + \mu a)^\top x - \left(\mu + \frac{1}{2} \sum_{i=1}^n \lambda_i \right)$$

The KKT necessary and sufficient conditions for optimality of x^* and of λ^*, μ^* require primal feasibility, dual feasiblity (i.e., $\lambda^* \geq 0$), complementary slackness $\lambda_i^*((x_i^*)^2 - 1) = 0$, and Lagrangian stationarity:

$$\nabla_x \mathcal{L}(x, \lambda^*, \mu^*) = (D + \Lambda^*) x + (r + \mu^* a) = 0.$$

That is, since D is strictly positive,

$$x^* = -(D + \Lambda^*)^{-1}(r + \mu^* a),$$

i.e., since D and Λ are diagonal,

$$x_i^* = -\frac{r_i + \mu^* a_i}{d_i + \lambda_i^*}, \quad i = 1, \ldots, n. \tag{1}$$

Notice that the problem is essentially separable in the variables x_i, except for the coupling constraint $a^\top x = 1$.

2. The dual function is

$$\begin{aligned} g(\lambda,\mu) &= \inf_x \mathcal{L}(x,\lambda,\mu) = \mathcal{L}(x^*,\lambda,\mu) \\ &= \frac{1}{2}(r+\mu a)((D+\Lambda)^{-1})^\top (D+\Lambda)(D+\Lambda)^{-1}(r+\mu a) \\ &\quad -(r+\mu a)^\top (D+\Lambda)^{-1}(r+\mu a) - \left(\mu + \frac{1}{2}\sum_{i=1}^n \lambda_i\right) \\ &= -\frac{1}{2}(r+\mu a)^\top (D+\Lambda)^{-1}(r+\mu a) - \left(\mu + \frac{1}{2}\sum_{i=1}^n \lambda_i\right) \\ &= -\mu - \frac{1}{2}\sum_{i=1}^n \left[\frac{(r_i+\mu a_i)^2}{d_i+\lambda_i} + \lambda_i\right], \end{aligned}$$

3. The dual problem amounts to maximizing $g(\lambda,\mu)$ w.r.t. μ and $\lambda \geq 0$. Let us find the maximum of $g(\lambda,\mu)$ w.r.t. λ, for fixed μ. The problem is separable, and it amounts to finding

$$\max_{\lambda_i \geq 0} \; -\frac{(r_i+\mu a_i)^2}{d_i+\lambda_i} - \lambda_i.$$

The unconstrained optimal point is at $\lambda_i = |r_i + \mu a_i| - d_i$. When this quantity is nonnegative, it is also the optimal solution of the constrained problem, otherwise the optimal solution is at the boundary, i.e., $\lambda_i = 0$. We thus have that

$$\lambda_i^*(\mu) = \begin{cases} |r_i+\mu a_i| - d_i & \text{if } |r_i+\mu a_i| - d_i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

4. Substituting (2) into (1), we can express $x_i^*(\mu)$ as a function of the scalar dual variable μ only:

$$x_i^*(\mu) = \begin{cases} -\text{sgn}(r_i+\mu a_i) & \text{if } |r_i+\mu a_i| - d_i \geq 0 \\ -(r_i+\mu a_i)/d_i & \text{otherwise.} \end{cases} \qquad i = 1,\ldots,n. \tag{3}$$

5. This algorithm returns a value of μ that is within ϵ from the true optimum μ^* in a number of iterations upper bounded by $\lceil \log_2(\ell/\epsilon) - 1 \rceil$, where ℓ is the length of the initial localization interval.

6. The numerical simulations confirm precisely this iteration bound. See the solution code in `kkt_conditions_sols.ipynb`. For this solution with $n = 500$, the bisection algorithm was approximately 30x faster than CVX. The time elapsed for solving via CVX was 0.1589 seconds while the bisection algorithm solved the problem in 0.00502 seconds. The euclidean distance between the solutions obtained with CVX and bisection was 0.00763. This result demonstrates a significant speedup in solving for optimal μ with minimal change to the optimal value.

Exercise 3 (Trust region subproblem) Consider the problem

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

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

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

   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

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

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?

Solution 3

1. The problem, as stated, is not convex, due to the non-convex constraint (only the objective is convex, since $Q \in \mathbb{S}^n$ and $c^T x$ is linear).

2. Since Q is PSD, let $Q = U \Lambda U^\top$ be the eigenvalue decomposition of Q, with U being unitary and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ the diagonal matrix containing the eigenvalues in decreasing order. With the change of variable $y = Ux$, and with $d = Uc$, we have $c^\top x = d^\top y$, and $x^\top Q x = y^\top \Lambda y$. The objective then reads $\min_y \; y^\top \Lambda y + 2 d^\top y$. Since U is unitary, the constraint on x becomes $\|y\|_2 = 1$. This proves the result.

3. Note that at optimum, y_i is of the opposite sign of d_i (since we are minimizing, we want the second term in the sum to be negative, since the first is always positive). That is, the problem can be written as

$$\min_{\xi} \sum_{i=1}^{n} \left(\lambda_i \xi_i^2 - 2|d_i|\xi_i\right) \;\; : \;\; \sum_{i=1}^{n} \xi_i^2 = 1, \;\; \xi \geq 0,$$

with $\xi_i = -\mathbf{sign}(d_i) y_i$, $i = 1, \ldots, n$.

Then, the new formulation results from the change of variable $z_i = \xi_i^2$, $i = 1, \ldots, n$. The optimal y is then obtained as $y_i = -\mathbf{sign}(d_i)\sqrt{z_i}$, $i = 1, \ldots, n$.

4. Dualizing the equality constraint leads to the dual function

$$\begin{aligned} g(\nu) &\doteq \min_{z \geq 0} \sum_{i=1}^{n} (\lambda_i z_i - 2|d_i|\sqrt{z_i}) + \nu(1 - \sum_{i=1}^{n} z_i) \\ &= \nu + \min_{z \geq 0} \sum_{i=1}^{n} ((\lambda_i - \nu) z_i - 2|d_i|\sqrt{z_i}) \\ &= \begin{cases} \nu - \displaystyle\sum_{i=1}^{n} \frac{d_i^2}{\lambda_i - \nu} & \text{if } \nu \leq \lambda_i, i = 1, \ldots, n, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Strong duality holds, thanks to Slater's condition. Note that in the above there is a single optimal z for any ν:

$$z_i^*(\nu) \doteq \frac{d_i^2}{(\lambda_i - \nu)^2}, \;\; i = 1, \ldots, n.$$

The dual problem writes

$$p^* = \max_{\nu} \; \nu - \sum_{i=1}^{n} \frac{d_i^2}{\lambda_i - \nu} \;\; : \;\; \nu \leq \nu_{\max} \doteq \min_{i} \; \lambda_i.$$

5. To implement bisection we need an initial interval for ν. With $\eta = \lambda_n - \nu \geq 0$ and $\delta_i = \lambda_i - \lambda_n \geq 0$, we obtain $p^* = \lambda_n - \phi^*$, where

$$\phi^* \doteq \min_{\eta \geq 0} \; \eta + \sum_{i=1}^{n} \frac{d_i^2}{\delta_i + \eta}.$$

Since $\delta_i \geq 0$, $i = 1, \ldots, n$, we have

$$\phi^* \leq \min_{\eta \geq 0} \; \eta + \frac{\|d\|_2^2}{\eta} = 2\|d\|_2,$$

Further observing that $\eta^* \leq \phi^*$, we obtain $0 \leq \eta^* \leq 2\|d\|_2$, which provides the desired interval.

6. We have a single optimal point for any dual variable ν:

$$z_i^*(\nu) = \frac{d_i^2}{(\lambda_i - \nu)^2}, \quad i = 1, \ldots, n.$$

Hence the above is primal optimal whenever ν is dual optimal. We then set $x^* = U^\top y^*$, with $y_i^* = -\mathbf{sign}(d_i)\sqrt{z_i^*}$, $i = 1, \ldots, n$.