Tutorial 10 Xiaojian Xu 04/26/2019

ESE415 OPTIMIZATION

Dr. Ulugbek Kamilov

Guideline





FINAL REVIEW

PROBLEMS IN LASTER YEAR'S FINAL EXAM

Key points

Underdetermined System

How to solve such problems using the derived

KKT conditions

- KKT conditions.
- How to use KKT to solve problems.

Proximal operations

- expression.
- How to get a closed form solution for some commonly used g(x).
- How to solve simple composite proximal problems.

Penalty method

- Expression;
- How to solve a constrain problem using penalty method.

Dual problem

- How to formulate a dual problem.
- How to solve a dual problem.
- •

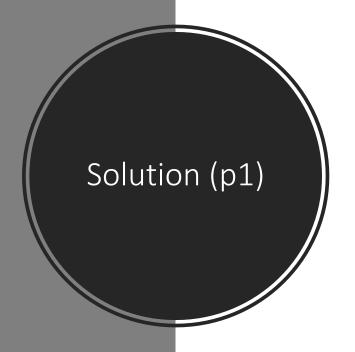
P1

Problem 1

Consider the following system of linear equations:

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 - x_2 + x_3 = 1 \end{cases}$$

- (a) Find the solution that is closest to the origin in \mathbb{R}^3 .
- (b) Find the solution that is closest to the vector $e_1 = (1, 0, 0)$.
- (c) Find the solution that is closest to the vector $e_2 = (0, 1, 0)$.



(a) We first express the given system of equations in a matrix-vector form:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$
, with $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. (1)

The Gram matrix and its inverse are given by

$$G = AA^{T} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \quad \Rightarrow \quad G^{-1} = \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}. \tag{2}$$

The minimum norm solution is then given by

$$\boldsymbol{x}^* = \boldsymbol{A}^T (\boldsymbol{A} \boldsymbol{A}^T)^{-1} \boldsymbol{b} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix}.$$
 (3)

- (b) The vector e_1 is already the solution of the system.
- (c) We need to solve the following optimization problem

$$\min \|x - e_2\| \quad \text{subject to} \quad Ax = b. \tag{4}$$

By introducing the change of the variables $z=x-e_2$ we get the following minimum norm problem

minimize
$$||z||$$
 subject to $Az = c$, with $c = b - Ae_2 = (0, 2)$. (5)

Therefore we have that

$$z^* = A^T (AA^T)^{-1} c = (1/2, -1, 1/2).$$
(6)

Finally we obtain

$$x^* = z^* + e_2 = (1/2, 0, 1/2).$$
 (7)

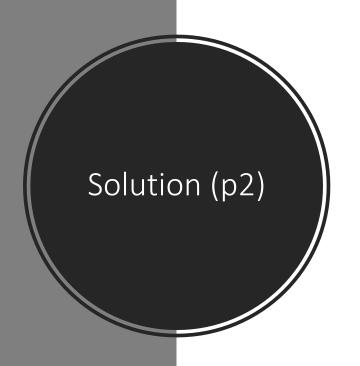
Ρ2

Problem 2

Consider the following optimization problem

minimize
$$f(x)=\frac{1}{2}\|x\|_{\ell_2}^2$$
 subject to $g(x)=x_1+\cdots+x_n+1\leq 0$ with $x\in\mathbb{R}^n$.

- (a) Use the KKT conditions to solve this problem analytically.
- (b) Find the sequence of solutions $\{x^t\}$ obtained with the quadratic penalty method.
- (c) Comment on the convergence rate of the sequence $\delta_t \triangleq \|x^t x^*\|_{\ell_2}$.



- (a) This problem is a superconsistent convex problem with differentiable functions. We can thus apply KKT conditions:
 - primal feasibility: $x_1 + \cdots + x_n \leq -1$;
 - *dual feasibility*: $\lambda \geq 0$;
 - complementary slackness: $\lambda(x_1 + \cdots + x_n + 1) = 0$;
 - stationarity:

$$x_i + \lambda = 0, \quad \forall i \in [1, \dots, n].$$

By solving for (x, λ) , we obtain that $\lambda > 0$ and

$$x_i^* = -\lambda^* = -\frac{1}{n}, \quad \forall i \in [1, \dots, n].$$

(b) We form the quadratic penalty function

$$q_t(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_{\ell_2}^2 + \frac{t}{2} (x_1 + \dots + x_n + 1)_+^2 \quad \Rightarrow \quad \nabla q_t(\mathbf{x}) = \mathbf{x} + t(x_1 + \dots + x_n + 1)_+ = \mathbf{0}.$$

Note that for all $x \in \mathcal{G}$, we get x = 0 as the critical point of q_t . On the other hand for $x \notin \mathcal{G}$, we get

$$x_i^t = -\frac{t}{1+nt}.$$

Since

$$q_t(\mathbf{x}^t) = \frac{t}{2(1+nt)} \le q_t(\mathbf{0}) = \frac{t}{2},$$

we conclude that x^t is the minimizer of q_t .

(c) We compute

$$\delta_t = \sqrt{\sum_{i=1}^n \left(\frac{-nt+1+nt}{n(1+nt)}\right)^2} = \frac{1}{\sqrt{n}(1+nt)},$$

which converges as O(1/t). This is a sublinear rate of convergence.

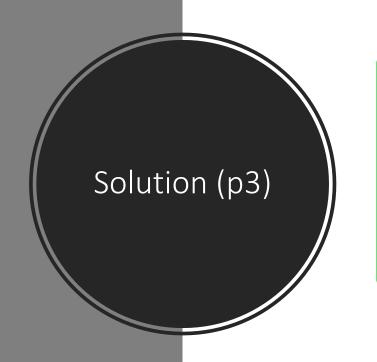
Р3

Problem 3

Show that the inequality

$$\sup_{\boldsymbol{\mu} \in \mathbb{R}^p} \left\{ \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ L(\boldsymbol{x}, \boldsymbol{\mu}) \right\} \right\} \leq \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \sup_{\boldsymbol{\mu} \in \mathbb{R}^p} \left\{ L(\boldsymbol{x}, \boldsymbol{\mu}) \right\} \right\},$$

always holds for any function $L: \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, and vectors $x \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^p$.



Note that for any $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^p$, we have that

$$d(\boldsymbol{\mu}^*) \triangleq \inf_{\boldsymbol{x} \in \mathbb{R}^n} \{ L(\boldsymbol{x}, \boldsymbol{\mu}^*) \} \leq L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \leq \sup_{\boldsymbol{\mu} \in \mathbb{R}^p} \{ L(\boldsymbol{x}^*, \boldsymbol{\mu}) \} \triangleq q(\boldsymbol{x}^*).$$

Hence we must also have that

$$\sup_{\boldsymbol{\mu} \in \mathbb{R}^p} \left\{ d(\boldsymbol{\mu}) \right\} \le \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ q(\boldsymbol{x}) \right\},$$

which proves the result.

P4

Problem 4

The proximal operator of a function g is defined as

$$\operatorname{prox}_g(\boldsymbol{y}) \triangleq \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell_2}^2 + g(\boldsymbol{x}) \right\}.$$

Many practically interesting functions admit closed form expressions for the proximal.

(a) Find the expression for the following quadratic proximal

$$g_1(x) = \frac{1}{2}x^\mathsf{T} Q x - c^\mathsf{T} x + d$$
, with $Q \succ 0, c \in \mathbb{R}^n$, and $d \in \mathbb{R}$.

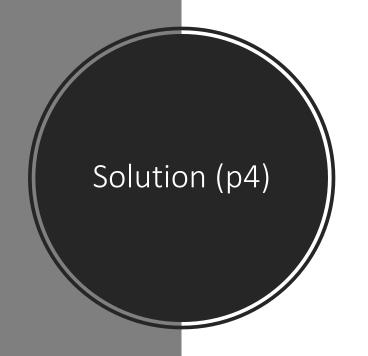
(b) Find the expression for the logarithmic proximal

$$g_2(\boldsymbol{x}) = -\lambda \sum_{i=1}^n \log(x_i), \quad \boldsymbol{x} > \boldsymbol{0},$$

where $\log(\cdot)$ denotes the natural logarithm.

(c) Find the expression for the following composite proximal

$$g_3(x) = \tau^2 \varphi(x/\tau + z)$$
, with $\tau > 0, z \in \mathbb{R}^n$, and $\varphi \in \Gamma^0(\mathbb{R}^n)$.



(a) We seek the minimizer of the following objective

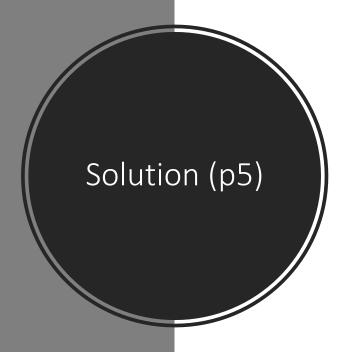
$$h_1(x) = \frac{1}{2} ||x - y||^2 + \frac{1}{2} x^{\mathsf{T}} Q x - c^{\mathsf{T}} x + d.$$

We compute the gradient

$$abla h_1(oldsymbol{x}) = oldsymbol{x} - oldsymbol{y} + oldsymbol{Q} oldsymbol{x} - oldsymbol{c} = oldsymbol{0} \qquad \Rightarrow \quad oldsymbol{x}^* = (oldsymbol{Q} + oldsymbol{I})^{-1}(oldsymbol{y} + oldsymbol{c}),$$

where \boldsymbol{I} is the identity matrix. Hence, we have

$$\operatorname{prox}_{g_1}(\boldsymbol{y}) = (\boldsymbol{Q} + \boldsymbol{I})^{-1}(\boldsymbol{y} + \boldsymbol{c})$$



Solution 4 (cont.)

(b) We seek the minimizer of

$$h_2(x) = \frac{1}{2} ||x - y||^2 - \lambda \sum_{i=1}^n \log(x_i).$$

Note that g_2 is separable and we can optimize along each variable x_i . We hence consider a scalar objective

$$h(x) = \frac{1}{2}(x - y)^2 - \lambda \log(x).$$

We seek the minimizer

$$h'(x) = x - y - \frac{\lambda}{x} = 0 \quad \Rightarrow \quad x^2 - yx - \lambda = 0 \quad \Rightarrow \quad x^* = \frac{y + \sqrt{y^2 + 4\lambda}}{2},$$

where we only kept the positive solution. Hence, we have

$$[\operatorname{prox}_{g_2}(\boldsymbol{y})]_i = \frac{y_i + \sqrt{y_i^2 + 4\lambda}}{2}$$
 where $i \in [1, \dots, n]$.

(c) We consider the following minimization problem

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \tau^2 \varphi \left(\boldsymbol{x} / \tau + \boldsymbol{z} \right) \right\} \\ &= \min_{\boldsymbol{u} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\tau(\boldsymbol{u} - \boldsymbol{z}) - \boldsymbol{y}\|^2 + \tau^2 \varphi \left(\boldsymbol{u} \right) \right\} \\ &= \min_{\boldsymbol{u} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\boldsymbol{u} - (\boldsymbol{y} / \tau + \boldsymbol{z})\|^2 + \varphi(\boldsymbol{u}) \right\}, \end{aligned}$$

where we used the variable change $u = x/\tau + z$, which is equivalent to $x = \tau(u - z)$. The solution of the last problem is given by

$$oldsymbol{u}^* = \mathrm{prox}_{arphi}(oldsymbol{y}/ au + oldsymbol{z}) \quad \Rightarrow \quad \mathrm{prox}_{g_3}(oldsymbol{y}) = au(\mathrm{prox}_{arphi}(oldsymbol{y}/ au + oldsymbol{z}) - oldsymbol{z}).$$

P5

Bonus Problem

Consider the following optimization problem

minimize
$$f(x)=x_1^2-x_2^2-x_1$$
 subject to $g(x)=x_1^3+x_2^2-1\leq 0$ and $h(x)=4x_1+x_2^2-1=0.$

- (a) Formulate the corresponding dual optimization problem with the dual objective function $d(\lambda, \mu)$, where $\lambda \geq 0$ and $\mu \in \mathbb{R}$ are the dual variables corresponding to the inequality and equality constraints, respectively.
- (b) Solve the dual optimization problem.
- (c) Find the duality gap by supposing that the optimal solution is $x^* = (-2, \pm 3)$.

Solution (p6)

Bonus Solution

(a) The dual function is given by

$$\begin{split} d(\lambda,\mu) &= \inf_{\boldsymbol{x} \in \mathbb{R}^2} \left\{ L(\boldsymbol{x},\lambda,\mu) = (x_1^2 - x_2^2 - x_1) + \lambda (x_1^3 + x_2^2 - 1) + \mu (4x_1 + x_2^2 - 1) \right\} \\ &= \inf_{\boldsymbol{x} \in \mathbb{R}^2} \left\{ \lambda x_1^3 + x_1^2 + (4\mu - 1)x_1 + (\lambda + \mu - 1)x_2^2 - \lambda - \mu \right\} \\ &= \begin{cases} \inf_{\boldsymbol{x} \in \mathbb{R}^2} \left\{ x_1^2 + (4\mu - 1)x_1 + (\mu - 1)x_2^2 - \mu \right\} & \text{when } \lambda = 0, \mu \geq 1 \\ -\infty & \text{when } \lambda \neq 0, \mu < 1. \end{cases} \end{split}$$

Therefore, for $\lambda = 0$ and $\mu \ge 1$, the minimizer is given by

$$x_1^* = \frac{1}{2}(1-4\mu) \quad \text{and} \quad x_2^* = 0 \quad \Rightarrow \quad d(\lambda,\mu) = -\frac{1}{4}(4\mu-1)^2 - \mu = -4\left(\mu - \frac{1}{8}\right)^2 - \frac{3}{16}.$$

(b) We solve the following problem

$$d^* = \max_{\lambda = 0, \, \mu \geq 1} \left\{ d(\lambda, \mu) = -4 \left(\mu - \frac{1}{8} \right)^2 - \frac{3}{16} \right\} = -\frac{13}{4} \quad \text{with} \quad \lambda^* = 0 \text{ and } \mu^* = 1.$$

(c) The duality gap is given by

$$\delta = f(x^*) - d(\lambda^*, \mu^*) = -3 + \frac{13}{4} = \frac{1}{4}.$$

Thanks

for all your support through the whole semester and good luck to your finals!