Interior Point Methods (II)

Lecture 15, Convex Optimization

National Taiwan University

June 17, 2021

Table of contents

Feasibility and phase I methods

- The barrier method requires a strictly feasible starting point $x^{(0)}$.
- When such a point is not known, the barrier method is preceded by a preliminary stage, called **phase I**, in which a strictly feasible point is computed (or the constraints are found to be infeasible).
- The strictly feasible point found during phase I is then used as the starting point for the barrier method, which is called the **phase II stage**.

Basic phase I method (1/4)

- We consider a set of inequalities and equalities in the variables $x \in \mathbf{R}^n$,
$$f_i(x) \leq 0, i = 1, \ldots, m, Ax = b,$$
where $f_i : \mathbf{R}^n \rightarrow \mathbf{R}$ are convex, with continuous second derivatives.
- We assume that we are given a point $x^{(0)} \in \textbf{dom}\ f_1 \bigcap \ldots \bigcap \textbf{dom}\ f_m$, with $Ax^{(0)} = b$.
- Our goal is to find a strictly feasible solution of these inequalities and equalities, or determine that none exists.

Basic phase I method (2/4)

- To do this we form the following optimization problem:

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

  in the variables $x \in \mathbf{R}^n, s \in \mathbf{R}$.
- The variable s can be interpreted as a bound on the maximum infeasibility of the inequalities; the goal is to drive the maximum infeasibility below zero.
- This problem is always strictly feasible, since we can choose $x^{(0)}$ as starting point for x, and for s, we can choose any number larger than $\max_{i=1,\ldots,m} f_i(x^{(0)})$.

Basic phase I method (3/4)

- We can therefore apply the barrier method to solve the problem in the previous page, called the **phase I optimization problem** associated with the inequality and equality system.
- We can distinguish three cases depending on the sign of the optimal value $\bar{p}^*$ of the phase I problem.
  1. If $\bar{p}^* < 0$, then $f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad Ax = b$ has a strictly feasible solution. (i.e., the original problem is strictly feasible.) We do not need to solve the phase I optimization problem with high accuracy; we can terminate when $s < 0$.
  2. If $\bar{p}^* > 0$, then the original problem is infeasible. We also do not need to solve the phase I optimization problem to high accuracy; we can terminate when a dual feasible point is found with positive dual objective (which proves that $\bar{p}^* > 0$).
  3. If $\bar{p}^* = 0$ and the minimum is attained at x^* and $s^* = 0$, then the set of inequalities is feasible, but not strictly feasible.

Basic phase I method (4/4)

- If $\bar{p}^* = 0$ and the minimum is not attained, then the inequalities are infeasible.
- In practice it is impossible to determine exactly that $\bar{p}^* = 0$.
- Instead, an optimization algorithm applied to the basic phase I problem will terminate with the conclusion that $|\bar{p}^*| < \epsilon$ for some small, positive ϵ.
- This allows us to conclude that the inequalities $f_i(x) \leq -\epsilon$ are infeasible, while the inequalities $f_i(x) \leq \epsilon$ are feasible.

Termination near the phase II central path (1/3)

- A simple variation on the basic phase I method, using the barrier method, has the property that (when the equalities and inequalities are strictly feasible) the central path for the phase I problem intersects the central path for the original optimization problem.
- We assume a point $x^{(0)} \in \mathcal{D} = \textbf{dom } f_0 \cap \textbf{dom } f_1 \cap \ldots \cap \textbf{dom } f_m$, with $Ax^{(0)} = b$ is given.

Termination near the phase II central path (2/3)

- We form the phase I optimization problem

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, i = 1, \ldots, m \\ & f_0(x) \leq M \\ & Ax = b, \end{array}$$

  where M is a constant chosen to be larger than $\max\left\{f_0(x^{(0)}), p^*\right\}$.

- We assume now that the original problem is strictly feasible, so the optimal value $\bar{p}^*$ of the phase I problem in the previous bullet point is negative.

Termination near the phase II central path (3/3)

- The central path of the phase I problem is characterized by

$$\sum_{i=1}^{m} \frac{1}{s - f_i(x)} = \bar{t}, \quad \frac{1}{M - f_0(x)} \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{s - f_i(x)} \nabla f_i(x) + A^T \nu = 0,$$

where $\bar{t}$ is the parameter.

- If (x, s) is on the central path and $s = 0$, then x and ν satisfy

$$t \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T \nu = 0$$

for $t = 1/(M - f_0(x))$.

- This means that x is on the central path for the original optimization problem.

Lagrange duality for Problems with Generalized Inequality Constraints

- In this section we examine how Lagrange duality extends to a problem with generalized inequality constraints

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p, \end{array}$$

  where $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones.
- We assume the domain of the problem, $D = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$, is nonempty.
- We do not assume convexity of the problem.

Monotonicity and Convexity with for a generalized inequality

Monotonicity with respect to a generalized inequality

Suppose $K \subseteq \mathbf{R}^m$ is a proper cone with associated generalized inequality $\preceq_K$. A function $f : \mathbf{R}^m \to \mathbf{R}$ is called ***K*-nondecreasing** if

$$x \preceq_K y \implies f(x) \leq f(y),$$

and ***K*-increasing** if

$$x \preceq_K y, \; x \neq y \implies f(x) < f(y).$$

Convexity with respect to a generalized inequality

Suppose $K \subseteq \mathbf{R}^m$ is a proper cone with associated generalized inequality $\preceq_K$. We say $f : \mathbf{R}^n \to \mathbf{R}^m$ is ***K*-convex** if for all $x, y \in \mathbf{R}^n$, and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y).$$

The Lagrange Dual Function

- With each generalized inequality $f_i(x) \preceq_{K_i} 0$, we associate a Lagrange multiplier vector $\lambda_i \in \mathbf{R}^{k_i}$ and define the associated Lagrangian as

$$L(x, \lambda, \nu) = f_0(x) + \lambda_1^T f_1(x) + \ldots + \lambda_m^T f_m(x) + \nu_1 h_1(x) + \ldots + \nu_p h_p(x),$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $\nu = (\nu_1, \ldots, \nu_p)$.

- The dual function is defined exactly as in a problem with scalar inequalities:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).$$

- Since the Lagrangian is affine in the dual variables (λ, ν), and the dual function is a pointwise infimum of the Lagrangian, the dual function is concave.

Nonnegativity Requirement of The Dual Problem

- As in a problem with scalar inequalities, the dual function gives lower bounds on p^*, the optimal value of the primal problem.
- Here the nonnegativity requirement on the dual variables is replaced by the condition

$$\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m,$$

where K_i^* denotes the dual cone of K_i.

  - Recall that for a problem with scalar inequalities, we require $\lambda_i \geq 0$.

Weak Duality and the Dual Problem

- If $\lambda_i \succeq_{K_i^*} 0$ and $f_i(\tilde{x}) \preceq_{K_i} 0$, then $\lambda_i^T f_i(\tilde{x}) \leq 0$.
- Therefore, for any primal feasible point $\tilde{x}$ and any $\lambda_i \succeq_{K_i^*} 0$, we have

$$f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}).$$

- Taking the infimum over $\tilde{x}$ yields $g(\lambda, \nu) \leq p^*$.
- The Lagrange dual optimization problem is

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, i = 1, \dots, m. \end{array}$$

- We always have weak duality, i.e., $d^* \leq p^*$, where d^* denotes the optimal value of the dual problem, whether or not the primal problem is convex.

Slater's condition and strong duality

- A sufficient condition for the strong duality ($d^* = p^*$) to hold is when
  1. the primal problem is convex, and
  2. the primal problem satisfies a generalized version of Slater's condition for the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Ax = b, \end{array}$$

  where f_0 is convex and f_i is K_i-convex. That is, there exists an $x \in \textbf{relint } \mathcal{D}$ with $Ax = b$ and $f_i(x) \prec_{K_i} 0, i = 1, \ldots, m$.

Example – Lagrange dual of semidefinite program (1/2)

- We consider a semidefinite program (SDP) in inequality form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \ldots + x_n F_n + G \preceq 0, \end{array}$$

  where $F_1, \ldots, F_n, G \in \mathbf{S}^k$.
  - Here f_1 is affine, and K_1 is $\mathbf{S}_+^k$, the positive semidefinite cone.
- We associate with the constraint a dual variable or multiplier $Z \in \mathbf{S}^k$, so the Lagrangian is

$$\begin{aligned} L(x, Z) &= c^T x + \mathbf{tr}((x_1 F_1 + \ldots + x_n F_n + G)Z) \\ &= x_1(c_1 + \mathbf{tr}(F_1 Z)) + \ldots + x_n(c_n + \mathbf{tr}(F_n Z)) + \mathbf{tr}(GZ), \end{aligned}$$

  which is affine in x.

Example – Lagrange dual of semidefinite program (2/2)

- The dual function is given by

$$g(Z) = \inf_x L(x, Z) = \begin{cases} \mathbf{tr}(GZ), & \mathbf{tr}(F_i Z) + c_i = 0, \ i = 1, \dots, n \\ -\infty, & \text{otherwise.} \end{cases}$$

- The dual problem can therefore be expressed as[1]

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, i = 1, \dots, n \\ & Z \succeq 0, \end{array}$$

  which is an SDP in standard form.
- Strong duality obtains if the semidefinite program (i.e., the primal problem) is strictly feasible: there exists an $x \in \mathbf{R}^n$ with

$$x_1 F_1 + \dots + x_n F_n + G \prec 0.$$

[1]Recall the fact that $\mathbf{S}_+^k$ is self-dual, i.e., $(\mathbf{S}_+^k)^* = \mathbf{S}_+^k$.

Example – Cone program in standard form (1/3)

- We consider the cone program (in standard form)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq_K 0, \end{array}$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $K \subseteq \mathbf{R}^n$ is a proper cone.

Example – Cone program in standard form (2/3)

- We associate with the equality constraint a multiplier $\nu \in \mathbf{R}^m$, and with the nonnegativity constraint a multiplier $\lambda \in \mathbf{R}^n$. The Lagrangian is

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b),$$

so the dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu, & A^T \nu - \lambda + c = 0 \\ -\infty, & \text{otherwise.} \end{cases}$$

Example – Cone program in standard form (3/3)

- The dual problem can be expressed as

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\nu + c = \lambda \\ & \lambda \succeq_{K^*} 0. \end{array}$$

- By eliminating λ and defining $y = -\nu$, this problem can be simplified to

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \preceq_{K^*} c, \end{array}$$

  which is a cone program in inequality form, involving the dual generalized inequality.
- Strong duality obtains if the Slater condition holds, i.e., there is an $x \succ_K 0$ with $Ax = b$.

Semidefinite program in standard form (1/2)

- In the previous example, if we repalce $K \subseteq \mathbf{R}^n$ with the semidefinite cone $\mathbf{S}_+^n$, we obtain an SDP problem in standard form.
- Specifically, the problem

$$\begin{array}{ll} \underset{x \in \mathbf{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq_K 0 \end{array}$$

would be transformed into

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n}{\text{minimize}} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\ & X \succeq 0. \end{array}$$

Semidefinite program in standard form (2/2)

- With some derivations, the dual problem can be shown to be

$$\begin{array}{ll} \underset{y \in \mathbf{R}^m}{\text{maximize}} & b^T y \\ \text{subject to} & y_1 A_1 + y_2 A_2 + \cdots y_m A_m \preceq C, \end{array}$$

  which is an SDP in inequality form.
- Compare the dual problem of cone program in standard form:

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \preceq_{K^*} c. \end{array}$$

- Recall again that the dual cone of $\mathbf{S}_+^n$ is $\mathbf{S}_+^n$ itself.

Optimality conditions

- In the following, we extend the optimality conditions (i.e., KKT conditions) earlier derived for problems with scalar inequalities

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p, \end{array}$$

  to problems with generalized inequalities:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p. \end{array}$$

Complementary slackness (1/2)

- Assume that the primal and dual optimal values are equal, and attained at the optimal points x^*, λ^*, ν^*.
- Then, we have

$$
\begin{aligned}
f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_{x \in \mathcal{D}} \underbrace{\left(f_0(x) + \sum_{i=1}^{m} {\lambda_i^*}^T f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)}_{L(x, \lambda^*, \nu^*)} \\
&\leq f_0(x^*) + \sum_{i=1}^{m} {\lambda_i^*}^T f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\leq f_0(x^*),
\end{aligned}
$$

and therefore we conclude that x^* minimizes $L(x, \lambda^*, \nu^*)$, and also that the two sums in the second line are zero.

Complementary slackness (2/2)

- From the previous page, we have

$$\sum_{i=1}^{m} {\lambda_i^*}^T f_i(x^*) = 0.$$

- Since each term in this sum is nonpositive, we conclude that

$${\lambda_i^*}^T f_i(x^*) = 0, i = 1, \ldots, m,$$

which is the generalized complementary slackness condition.

- Therefore,

$$\lambda_i^* \succ_{K_i^*} 0 \Rightarrow f_i(x^*) = 0, \quad \text{and } f_i(x^*) \prec_{K_i} 0 \Rightarrow \lambda_i^* = 0.$$

- However, in contrast to problems with scalar inequalities, it is possible to satisfy the generalized complementary slackness with $\lambda_i^* \neq 0$ and $f_i(x^*) \neq 0$.

KKT conditions (1/3)

- Now we add the assumption that the functions f_i, h_i are differentiable, and generalize the KKT conditions to problems with generalized inequalities.
- Since x^* minimizes $L(x, \lambda^*, \nu^*)$, its gradient with respect to x vanishes at x^*:

$$\nabla f_0(x^*) + \sum_{i=1}^{m} Df_i(x^*)^T \lambda_i^* + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0,$$

where $Df_i(x^*) \in \mathbf{R}^{k_i \times n}$ is the derivative[2] of f_i evaluated at x^*.

[2]See p.10 of Lecture 3, or §A.4.1 in the textbook, for the definition of the derivative operator D.

KKT conditions (2/3)

- Thus, if strong duality holds, any primal optimal x^* and any dual optimal (λ^*, ν^*) must satisfy the optimality conditions (or KKT conditions)

$$
\begin{aligned}
f_i(x^*) &\preceq_{K_i} 0, i = 1, ..., m \\
h_i(x^*) &= 0, i = 1, ..., p \\
\lambda_i^* &\succeq_{K_i^*} 0, i = 1, ..., m \\
{\lambda_i^*}^T f_i(x^*) &= 0, i = 1, ..., m \\
\nabla f_0(x^*) + \sum_{i=1}^{m} Df_i(x^*)^T \lambda_i^* + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0.
\end{aligned}
$$

KKT conditions (3/3)

- If the primal problem is convex, the converse also holds, i.e., the conditions

$$
\begin{aligned}
f_i(x^*) &\preceq_{K_i} 0, i = 1, \ldots, m \\
h_i(x^*) &= 0, i = 1, \ldots, p \\
\lambda_i^* &\succeq_{K_i^*} 0, i = 1, \ldots, m \\
{\lambda_i^*}^T f_i(x^*) &= 0, i = 1, \ldots, m \\
\nabla f_0(x^*) + \sum_{i=1}^{m} Df_i(x^*)^T \lambda_i^* + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0
\end{aligned}
$$

are sufficient conditions for optimality of x^*, (λ^*, ν^*).

Problems with Generalized Inequalities (1/2)

- In this section, we show how the barrier method can be extended to problems with generalized inequalities. We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Ax = b, \end{array}$$

  where $f_0 : \mathbf{R}^n \to \mathbf{R}$ is convex, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}, i = 1, \ldots, k$, are K_i-convex, and $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones.
- We assume that the functions f_i are twice continuously differentiable, that $A \in \mathbf{R}^{p \times n}$ with rank $A = p$, and that the problem is solvable.

Problems with Generalized Inequalities (2/2)

- The KKT conditions for the problem are

$$
\begin{aligned}
Ax^* &= b \\
f_i(x^*) &\preceq_{K_i} 0, i = 1, \dots, m \\
\lambda_i^* &\succeq_{K_i^*} 0, i = 1, \dots, m \\
\nabla f_0(x^*) + \sum_{i=1}^{m} Df_i(x^*)^T \lambda_i^* + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0, \\
{\lambda_i^*}^T f_i(x^*) &= 0, i = 1, \dots, m
\end{aligned}
$$

  where $Df_i(x^*) \in \mathbf{R}^{k_i \times n}$ is the derivative of f_i at x^*.

- We will assume that the problem is strictly feasible, so the KKT conditions are necessary and sufficient conditions for optimality of x^*.

Generalized logarithm for a proper cone (1/2)

- We first define the analog of the logarithm, $\log x$, for a proper cone $K \subseteq \mathbf{R}^q$.
- We say that $\psi : \mathbf{R}^q \to \mathbf{R}$ is a **generalized logarithm** for K if
  - ψ is concave, closed, twice continuously differentiable, $\textbf{dom}\ \psi = \textbf{int}\ K$, and $\nabla^2\psi(y) \prec 0$ for $y \in \textbf{int}\ K$.
  - There is a constant $\theta > 0$ such that for all $y \succ_K 0$, and all $s > 0$,
  $$\psi(sy) = \psi(y) + \theta \log s.$$
- In other words, ψ behaves like a logarithm along any ray in the cone K.

Generalized logarithm for a proper cone (2/2)

- We call the constant θ the degree of ψ (since $\exp\psi$ is a homogeneous function of degree θ).
- If ψ is a generalized logarithm for K, then so is $\psi + a$, where $a \in \mathbf{R}$.
- The ordinary logarithm ($\log x$) is a generalized logarithm for $\mathbf{R}_+$.

Properties of a generalized logarithm for a proper cone

- We will use the following two properties, which are satisfied by any generalized logarithm:
  1. If $y \succ_K 0$, then
  $$\nabla\psi(y) \succ_{K^*} 0,$$
  which implies ψ is K-increasing.
  2. $y^T \nabla\psi(y) = \theta$.
- The second property follows immediately from differentiating $\psi(sy) = \psi(y) + \theta \log s$ with respect to s.

Example – Nonnegative orthant

- The function

$$\psi(x) = \sum_{i=1}^{n} \log x_i$$

is a generalized logarithm for $K = \mathbf{R}^n_+$, with degree n.

- For $x \succ 0$,

$$\nabla\psi(x) = (1/x_1, \ldots, 1/x_n),$$

so $\nabla\psi(x) \succ 0$, and $x^T \nabla\psi(x) = n$.

Example – Positive semidefinite cone

- The function $\psi(X) = \log\det X$ is a generalized logarithm for the cone $\mathbf{S}_+^p$.
- The degree is p, since

$$\log\det(sX) = \log\det X + p\log s$$

  for $s > 0$.
- The gradient of ψ at a point $X \in \mathbf{S}_{++}^p$ is equal to

$$\nabla\psi(X) = X^{-1}.$$

- Thus, we have $\nabla\psi(X) = X^{-1} \succ 0$, and the inner product of X and $\nabla\psi(X)$ is equal to $\mathbf{tr}(XX^{-1}) = p$.

Logarithmic barrier functions for generalized inequalities

- Returning to the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, \dots, m \\ & Ax = b, \end{array}$$

  let $\psi_1, \dots, \psi_m$ be generalized logarithms for the cones $K_1, \dots, K_m$, respectively, with degrees $\theta_1, \dots, \theta_m$.

- We define the logarithmic barrier function for the problem as

$$\phi(x) = -\sum_{i=1}^{m} \psi_i(-f_i(x)),$$

  dom $\phi = \{x \mid f_i(x) \prec_{K_i} 0, i = 1, \dots, m\}$.

- Convexity of ϕ follows from the fact that the functions ψ_i are K_i-increasing, and the functions f_i are K_i-convex.

The central path

- We define the central point $x^*(t)$, for $t \geq 0$, as the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) - \sum_{i=1}^{m} \psi_i(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

  (assuming the minimizer exists, and is unique).

- Central points are characterized by the optimality condition

$$\begin{aligned} & t\nabla f_0(x) + \nabla\phi(x) + A^T\nu \\ = \quad & t\nabla f_0(x) + \sum_{i=1}^{m} Df_i(x)^T \nabla\psi_i(-f_i(x)) + A^T\nu = 0, \end{aligned}$$

  for some $\nu \in \mathbf{R}^p$, where $Df_i(x)$ is the derivative of f_i at x.

Dual points on central path (1/4)

- For $i = 1, ..., m$, define

$$\lambda_i^*(t) = \frac{1}{t} \nabla \psi_i(-f_i(x^*(t))),$$

and let $\nu^*(t) = \nu/t$, where ν is the optimal dual variable of the approximated problem in the previous page.

- We will show that $\lambda_1^*(t), ..., \lambda_m^*(t)$, together with $\nu^*(t)$, are dual feasible for the original problem.

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, i = 1, ..., m \\ & Ax = b. \end{array}$$

Dual points on central path (2/4)

- First, by the monotonicity property of generalized logarithms $\nabla\psi(y) \succ_{K^*} 0$, we have

$$\lambda_i^*(t) \succ_{K_i^*} 0.$$

- Second, following from the fact that

$$t\nabla f_0(x) + \sum_{i=1}^{m} Df_i(x)^T \nabla\psi_i(-f_i(x)) + A^T\nu = 0,$$

we argue that the Lagrangian

$$L(x, \lambda^*(t), \nu^*(t)) = f_0(x) + \sum_{i=1}^{m} \lambda_i^*(t)^T f_i(x) + \nu^*(t)^T(Ax - b)$$

is minimized over x by $x = x^*(t)$.

Dual points on central path (3/4)

- The dual function g evaluated at $(\lambda^*(t), \nu^*(t))$ is therefore equal to

$$
\begin{aligned}
& g(\lambda^*(t), \nu^*(t)) \\
= \quad & f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t)^T f_i(x^*(t)) + \nu^*(t)^T (Ax^*(t) - b) \\
= \quad & f_0(x^*(t)) + (1/t) \sum_{i=1}^{m} \nabla\psi_i(-f_i(x^*(t)))^T f_i(x^*(t)) \\
= \quad & f_0(x^*(t)) - (1/t) \sum_{i=1}^{m} \theta_i,
\end{aligned}
$$

where θ_i is the degree of ψ_i.

Dual points on central path (4/4)

- In the last line, we use the fact that $y^T \nabla \psi_i(y) = \theta_i$ for $y \succ_{K_i} 0$, and therefore

$$\lambda_i^*(t)^T f_i(x^*(t)) = -\theta_i/t, i = 1, ..., m.$$

- Thus, if we define $\theta = \sum_{i=1}^m \theta_i$, then the primal feasible point $x^*(t)$ and the dual feasible point $(\lambda^*(t), \nu^*(t))$ have duality gap θ/t.
- This is just like the scalar case, except that θ, the sum of the degrees of the generalized logarithms for the cones, appears in place of m, the number of inequalities.

Semidefinite programming in inequality form (1/3)

- We consider the SDP with variable $x \in \mathbf{R}^n$,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = x_1 F_1 + \ldots + x_n F_n + G \preceq 0, \end{array}$$

  where $G, F_1, \ldots, F_n \in \mathbf{S}^p$.

- The dual problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, i = 1, \ldots, n \\ & Z \succeq 0. \end{array}$$

Semidefinite programming in inequality form (2/3)

- Using the generalized logarithm $\log\det X$ for the positive semidefinite cone $\mathbf{S}^p_+$, we have the barrier function (for the primal problem)

$$\phi(x) = \log\det(-F(x)^{-1})$$

with $\mathbf{dom}\ \phi = \{x \mid F(x) \prec 0\}$.

- For strictly feasible x, the gradient of ϕ is equal to

$$\frac{\partial \phi(x)}{\partial x_i} = \mathbf{tr}(-F(x)^{-1}F_i), i = 1, \ldots, n,$$

which gives us the optimality conditions that characterize central points:

$$tc_i + \mathbf{tr}(-F(x^*(t))^{-1}F_i) = 0, i = 1, \ldots, n.$$

Semidefinite programming in inequality form (3/3)

- Hence the matrix

$$Z^*(t) = \frac{1}{t}(-F(x^*(t)))^{-1}$$

is strictly dual feasible, and the duality gap associated with $x^*(t)$ and $Z^*(t)$ is p/t.

Summary (1/2)

- In this semester, we studied theory and algorithms on mathematical optimization problems, particularly convex optimization problems.
  - Chapter 2: convex sets, affine sets, cones, convexity-preserving operations, separating hyperplane theorem, proper cones, dual cones, etc.
  - Chapter 3: convex functions, Jensen's inequality, convexity-preservation operations, quasiconvex function, conjugate functions.
  - Chapter 4: convex optimization problems, linear programs (LP), quadratic programs, quadratic constrained quadratic programs (QCQP), second-order cone problems (SOCP), quasi-convex problems, convex problems with genearalized inequalities, convex conic problems, semidefinite programs (SDP).

Summary (2/2)

- Chapter 5: Lagrangian, dual functions, dual problems, weak duality, strong duality, KKT conditions, dual problems of SDPs.
- Chapters 9 - 11: Descent methods, Newton's methods, Newton's method with equality constriants, interior-point methods.