Additional exercises

Contents

1 Conditional-gradient / Frank-Wolfe method

Let C be a compact convex set in $\mathbb{R}^n$, and let $f(x)$ be a smooth (nonlinear) convex function defined on C. Consider the following algorithm to minimize f on C:

- Initialize: $x_0 \in C$
- For $k = 0, 1, \ldots$
- Compute $y_k = \operatorname{argmin}_{x \in C} \nabla f(x_k)^T (x - x_k)$
- Set $x_{k+1} = (1 - \gamma_k) x_k + \gamma_k y_k$, where $\gamma_k = 2/(k+1)$
- End For

Let D be the Euclidean diameter of C, defined by $D = \max_{x,y \in C} \|x - y\|_2$, and L be the Lipschitz constant for ∇f with respect to the Euclidean norm. Let $x^* = \operatorname{argmin}_{x \in C} f(x)$ and $f^* = f(x^*)$.

(a) Prove that for any integer k we have

$$f(x_{k+1}) - f^* \leq (1 - \gamma_k)(f(x_k) - f^*) + \frac{\gamma_k^2 L D^2}{2}.$$

(b) Deduce, by induction, that $f(x_k) - f^* \leq \frac{2LD^2}{k+1}$ for all $k \geq 0$.

2 Lagrangian and KKT conditions

Consider a linearly constrained optimization problem

$$\inf_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad Ax = b. \tag{1}$$

We assume that f is smooth and convex, and that the solution to (1) is attained (i.e., the inf is really a min). Let L be the Lagrangian associated to (1). Show that a sufficient condition for an $\bar{x} \in \mathbb{R}^n$, satisfying $A\bar{x} = b$, to be optimal is that there exist $\bar{z}$ such that $\nabla_x L(\bar{x}, \bar{z}) = 0$.

Assuming that Slater's condition holds, show that the condition is also necessary.

3 Duality for linear programming

Consider the linear program

$$\min_{x \in \mathbb{R}^n} \quad c^T x \quad \text{s.t.} \quad Ax = b, \ x \geq 0 \tag{2}$$

where $c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

(a) By writing problem above as $\min_{x \in \mathbb{R}^n} \{c^T x + I(x) : Ax = b\}$ where $I(x)$ is the indicator function of $\mathbb{R}^n_+$, derive an explicit formulation of the Lagrange dual of (2) and give sufficient conditions for the dual problem to have the same optimal value as (2).

(b) Assuming that strong duality holds, and that primal and dual optimal values are finite and attained, show that necessary and sufficient conditions for x to be an optimal point of (2) is that there exists $z \in \mathbb{R}^m, s \in \mathbb{R}^n$ such that the following conditions hold:

$$\begin{cases} Ax = b, \ x \geq 0 \\ c + A^T z = s, \ s \geq 0 \\ x_i s_i = 0, \ \forall i = 1, \ldots, n. \end{cases} \tag{3}$$

(These are the KKT conditions of optimality.)

4 Network flow optimization

Consider a directed graph with n nodes and m arcs. The network flow problem seeks to find the optimal way to flow a certain amount $G > 0$ of goods (water, electricity, shipment, ...) from a source node $s \in \{1, \ldots, n\}$ to a destination node $t \in \{1, \ldots, n\}$, given the following constraints/costs:

- There is flow conservation at each node $i \in \{1, \ldots, n\}$ (i.e., ingoing flow must equal outgoing flow at each node)
- The cost of transferring flow x_j on arc j is $\phi_j(x_j)$.

Note that flow x_j on an arc j can potentially be negative, which means that flow traverses in reverse direction.

(a) Show that network flow problem can be written as

$$\min_{x_1,\ldots,x_m} \quad \sum_{j=1}^{m} \phi_j(x_j) \quad \text{s.t.} \quad Ax = b \tag{4}$$

where $A \in \mathbb{R}^{n\times m}$ is a matrix that you should specify, and $b \in \mathbb{R}^n$ is a vector that you should specify. [*Hint: your matrix A should have exactly two nonzero entries per column*].

(b) Write down the Lagrangian dual of (4). Give sufficient conditions for strong duality to hold.

(c) Assuming the ϕ_j are strongly convex, write down the gradient method for the dual problem derived in part (b).

(d) Assuming strong duality holds, show that $x \in \mathbb{R}^m$ is optimal for (4) iff there exists $z \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that

$$Ax = b, \quad A^T z = y, \quad \phi_j'(x_j) = y_j. \tag{5}$$

[*You can use the result of Exercise 2 if you want.*] Consider the case where the network is an electric circuit, x_j is electric current, and $\phi_j(x_j) = R_j x_j^2/2$ which means that arc j is a resistor with resistance R_j. Give a physical interpretation of the variables z_i $(i = 1, \ldots, n)$ and of the equations (5).

5 Convergence of Newton's method

(a) Write down the Newton iterations (with unit step size) to minimize the function $f(x) = |x|^{5/2}$ where $x \in \mathbb{R}$, starting from an initial condition $x_0 \in \mathbb{R}$. What convergence rate do you get? How do you reconcile this with the theorem proved in lecture?

(b) Consider Newton's method applied to the function whose derivative $f'(x)$ is given by the graph in Figure 1. Identify the initial points for which Newton's method (with unit step size) converges.

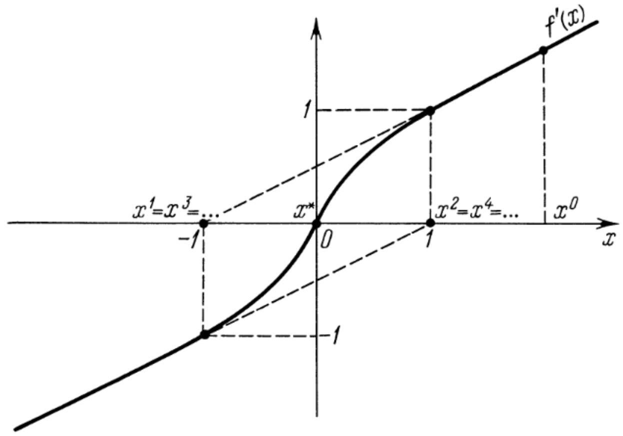

Figure 1: Figure from Polyak's "Introduction to Optimization", page 30

6 Newton's method for linearly constrained problems

Consider the problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Px + q^T x \text{ subject to } Ax = b \tag{6}$$

where P is an $n \times n$ real symmetric positive definite matrix, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

(a) Give a closed form expression for the solution of (6).

(b) Deduce a Newton's method to solve problems of the form

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad Ax = b \tag{7}$$

where f is a (strongly) convex function.

7 Approximate path-following method

The path-following method we saw in lecture assumes that we compute $x^*(t)$ exactly along points of the central path. In practice we cannot do this. Consider the following alternative path-following method:

- *Input:* $t_0 > 0$, x_0 such that $\lambda_{t_0}(x) \le 1/9$, $\epsilon > 0$, θ barrier parameter of s.c. function F.
- *Initialize:* set $t = t_0, x = x_0$, $\alpha = \frac{1/4+\sqrt{\theta}}{1/9+\sqrt{\theta}}$.
- While $\theta/t > \epsilon$
- Let $t^+ = \alpha t$.

(*) Let $x^+ = x - (\nabla^2 F(x))^{-1}(t^+ c + \nabla F_{t^+}(x))$.

- Update $x = x^+$, $t = t^+$.
- End While

Show that for each iteration in the main "While" loop we have: before execution of step (*), $\lambda_{t^+}(x) \le 1/4$ and after execution of step (*) we have $\lambda_{t^+}(x^+) \le 1/9$.

8 Path-following method and KKT conditions

Consider the linear program in inequality form

$$\min_{x \in \mathbb{R}^n} \; c^T x \;\text{ s.t. }\; Ax \le b \tag{8}$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and consider the central path, for $t > 0$

$$x^*(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} \; t\, c^T x \; - \; \sum_{i=1}^m \log(b_i - a_i^T x)$$

associated to the logarithmic barrier function $F(x) = -\sum_{i=1}^n \log(b_i - a_i^T x)$ of $Q = \{x : Ax \le b\}$.

The dual program of (8) is

$$\max \;\; -b^T z \;\; : \;\; c + A^T z = 0, z \ge 0. \tag{9}$$

For any $t > 0$, show that the vector $z(t) = \frac{1}{t} \left[\frac{1}{b_i - a_i^T x_i^*(t)} \right]_{i=1,\ldots,m}$ is feasible for (9). Compute $c^T x^*(t) + b^T z(t)$. What does this imply?

9 Second-order cone programming

Consider the regularized least-squares problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|Dx\|_1$$

where $A \in \mathbb{R}^{m \times n}, D \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^m$ and $\lambda \ge 0$. Show that this problem can be expressed as a second-order cone program. [*Hint: show that, for two real numbers s, t we have $s^2 \le t$ iff* $\|(2s, t-1)\|_2 \le t + 1$].

10 LP, SOCP, SDP

Show that linear programming is a special case of semidefinite programming. Show that second-order cone programming is also a special case of semidefinite programming. [*Hint: consider the constraint* $\begin{bmatrix} t & x^T \\ x & tI_n \end{bmatrix} \succeq 0$.]

11 Euclidean distance matrices

A family of $\binom{n}{2}$ numbers $(d_{ij})_{1 \le i < j \le n}$ is *Euclidean-realizable* if there exist $k \in \mathbb{N}$ and points $x_1, \ldots, x_n \in \mathbb{R}^k$ such that $d_{ij} = \|x_i - x_j\|_2$ for all $1 \le i < j \le n$. Show that one can decide whether $(d_{ij})_{ij}$ is Euclidean-realizable by solving a *semidefinite programming feasibility problem.*

12 Sum-of-squares and semidefinite programming

A polynomial $p(t)$ is called a *sum of squares* if we can write $p(t) = \sum_j q_j(t)^2$ for some other polynomials $q_j(t)$.

(a) Show that a degree $2d$ polynomial $p(t) = \sum_{k=0}^{2d} p_k t^k$ is a sum of squares iff there exists a symmetric positive semidefinite matrix M of size $(d+1)\times(d+1)$ such that

$$\sum_{\substack{0\le i,j\le d\\ i+j=k}} M_{ij} = p_k \qquad \forall k = 0,\ldots,2d.$$

[The rows/columns of matrix M are indexed by $0,1,\ldots,d$.]

(b) Consider the problem of finding the monic[1] polynomial of degree $2d$ with smallest sup-norm on $[-1,1]$, i.e.,

$$\min \quad \|p\|_\infty \text{ s.t. } \; p \text{ is monic of degree } 2d$$

where $\|p\|_\infty = \max_{t\in[-1,1]} |p(t)|$. Show that this problem can be formulated as a semidefinite program. [*Hint: you can use the fact that a polynomial of degree* $2d$ *is nonnegative on* $[-1,1]$ *iff it is of the form* $s_1(t) + (1-t^2)s_2(t)$ *where* s_1, s_2 *are sums of squares.*]

13 The definition of self-concordant functions (*)

Note: this exercise is more technical, you are advised to work on the exercises above first.

Let f be a C^3 (three-times continuously differentiable) convex function such that $\mathrm{dom}(f)$ is open, $\mathrm{epi}(f)$ is closed, and $\nabla^2 f(x)$ is positive definite for all $x \in \mathrm{dom}(f)$. For $x \in \mathrm{dom}(f)$ we let $\|h\|_x = \sqrt{h^T \nabla^2 f(x) h}$.

The goal of this exercise is to show that the definition of self-concordance we saw in lecture is equivalent to another commonly used definition.

(a) Show that if f is self-concordant (according to the definition we saw in lecture), then we have

$$\left|\frac{d}{dt}\left(h^T \nabla^2 f(x+tv) h\right)\right| \le 2\|h\|_{x+tv}^2 \|v\|_{x+tv} \tag{10}$$

for all $x \in \mathrm{dom}(f)$, $h, v \in \mathbb{R}^n$ and $t \in \mathbb{R}$ such that $x + tv \in \mathrm{dom}(f)$.

Note: by letting $v = h$ *and* $t \to 0$ *Equation* (10) *reads*

$$|\nabla^3 f(x)[h,h,h]| \le 2\|h\|_x^3 \tag{11}$$

where $\nabla^3 f(x)$ *is the third derivative of* f*, seen as a trilinear form. The equation above is often taken as the definition of self-concordance. It can be shown, using the theory of trilinear forms, that* (11) *is actually equivalent to* (10) *but we will omit this here.*

(b) We now prove that if f satisfies (10), then it is self-concordant (according to the definition given in lecture). We thus assume that f satisfies (10).

(i) Show that $\left|\frac{d}{dt}\left(\frac{1}{\|v\|_{x+tv}}\right)\right| \le 1$ for all $x \in \mathrm{dom}(f)$, $v \in \mathbb{R}^n$ such that $x + tv \in \mathrm{dom}(f)$

(ii) Deduce that $\|v\|_{x+tv} \le \frac{\|v\|_x}{1 - t\|v\|_x}$.

(iii) Conclude.

[1] A monic polynomial is one whose leading coefficient is equal to 1.