Tutorial and Additional Problems partly with solutions

1. Consider optimizing a function $f : \Re^n \rightarrow \Re$ that is bounded in $\Re^n$ and that is continuously diffirentiable in an open set $\mathcal{R}$ containing the sub-level set $\mathcal{S} = \left\{\mathbf{x} | f(\mathbf{x}) \leq f(\mathbf{x}^0)\right\}$. Also assume that that the gradient ∇f is Lipschitz continuous on $\mathcal{R}$.

   Consider applying a quasi-newton descent algorithm for optimizing f using $\mathbf{x}^0$ as the initial point, with the update rule

   $$\mathbf{x}^{(k+1)} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$$

   such that α_k satisfy the Wolfe conditions for each k.

   While studying the convergence analysis of the class of steepest descent methods, we stated that under the Wolfe line-search conditions, for Newton-like methods that are characterized by a sequence of positive definite matrices B_k, convergence is assured if $||B_k|| ||B_k^{-1}|| \leq M$ for all k. This condition is called the "uniformly bounded condition number" criterion. Prove this statement of convergence from scratch. You may reproduce steps from your class notes, as and when required.

   You can also assume the following, wherever required:

   (a) $||A\mathbf{x}|| \leq ||A|| ||\mathbf{x}||$.

   (b) For a symmetric and positive definite matrix A, $||A^{\frac{1}{2}}|| = ||A||^{\frac{1}{2}}$.

2. **Convergence of Cutting Plane algorithm:** We present a generalized cutting plane algorithm for the optimization problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{D} \end{aligned} \tag{1}$$

for some closed and convex set $\mathcal{D}$ and concave f. Let $\mathbf{g}(\mathbf{x})$ be a subgradient[1] at some point $\mathbf{x}$ for the function f. A version of the general cutting plane algorithm consists of solving the following problem in the k^{th} iteration to get $\mathbf{x}^k$.

$$\begin{array}{rll} \mathbf{x}^k = & \text{maximize} & f^k(\mathbf{x}) \\ & \text{subject to} & \mathbf{x} \in \mathcal{D} \end{array} \tag{2}$$

where the function f is replaced by a polyhedral approximation f^k constructed using the points $\mathbf{x}^i$ generated so far, along with their subgradients $\mathbf{g}(\mathbf{x}^i) \equiv \mathbf{g}^i$. More specifically,

$$f^k(\mathbf{x}) = \min\left\{f(\mathbf{x}^0) + (\mathbf{x} - \mathbf{x}^0)^T\mathbf{g}^0), \ldots, f(\mathbf{x}^{(k-1)}) + (\mathbf{x} - \mathbf{x}^{(k-1)})^T\mathbf{g}^{(k-1)}\right\}$$

Assume that the maximum of f^k is attained for all k. Prove that the cutting plane algorithm, with the updates presented as above, converges finitely for the dual of a linear program, with atleast one strategy for choosing the subgradient (in fact, it converges for any choice of the subgradients). Also state the choice of the subgradients.

You can assume that the dual function for a linear progran is of the form

$$\min_{i \in \mathcal{I}} \left\{\mathbf{a}_i^T\mathbf{x} + b_i\right\}$$

where $\mathcal{I}$ is a finite index set and $a_i \in \Re^n$ and b_i are given vectors respectively.

3. Solve the minimization problem

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{0} \\ & ||\mathbf{x}||_2 \leq 1 \end{array} \tag{3}$$

4. Suppose that $\widetilde{f}(\mathbf{z}) = f(\mathbf{x})$, where $\mathbf{x} = S\mathbf{z} + \mathbf{s}$ for some $S \in \Re^{n\times m}$ and $\mathbf{s} \in \Re^n$. Show that

$$\nabla\widetilde{f}(\mathbf{z}) = S^T\nabla f(\mathbf{x})$$

and

$$\nabla^2\widetilde{f}(\mathbf{z}) = S^T\nabla^2 f(\mathbf{x})S$$

[1] $\mathbf{g}$ is a subgradient at $\mathbf{x}$ for a concave function f if and only if $-\mathbf{g}$ is a subgradient at $\mathbf{x}$ for the convex function $-f$.

5. Consider the general form of constrained convex optimization problem:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & A\mathbf{x} = b \end{array} \tag{4}$$

where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are convex functions.

Now consider the modified problem

$$\begin{array}{ll} \text{minimize} & B_\mu(\mathbf{x}) = f(\mathbf{x}) + \mu \sum_{i=1}^{m} h(-g_i(\mathbf{x})) \\ \text{subject to} & A\mathbf{x} = b \end{array} \tag{5}$$

with domain $\{\mathbf{x} | g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m\}$. Assume that $h(v) : \Re^+ \to \Re$ is an increasing and a differentiable convex function of v and $\mu > 0$ is a parameter.

Let $\widehat{\mathbf{x}}(\mu)$ be the solution to (5) and consider it as an approximation to the solution of (4). Show how to construct a dual (for (4)) feasible λ from $\widehat{\mathbf{x}}(\mu)$. Find the associated duality gap. Derive the general form of the function h under which the duality gap obtained thus, depends only on μ and m and no other data from the problem.

Show that the form must be:

$$h(z) = -a \log(-z) + b$$

with some constraints on a and b. What are the constraints?

6. Consider the problem

$$\text{minimize} \quad \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} \tag{6}$$

where A is a symmetric positive definite matrix. Let $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(n-1)}\}$ be a set of nonzero vectors that are mutually conjugate with respect to A.

The algorithm is iterative (like the conjugate gradient method outlined in notes). The k^{th} iteration consists of the following step:

- $\mathbf{x}^{(k+1)} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$ where α_k is the one dimensional minimizer of $\phi(\alpha) = f(\mathbf{x}^k + \alpha\mathbf{d}^k)$ and is given as $\alpha_k = -\frac{\nabla^T f(\mathbf{x}^k)\mathbf{d}^k}{(\mathbf{d}^k)^T A\mathbf{d}^k}$.

Let $\mathbf{x}^0 \in \Re^n$ be the initial point. We will prove that the sequence $\{\mathbf{x}^k\}$ generated by the repeated application of the conjugate gradient step above, for increasing values of k, converges to the solution $\mathbf{x}^*$ of the problem (6) in at most n steps (for step (b) onwards, provide brief justification):

(a) Prove that the directions $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(n-1)}\}$ are linearly independent.

(b) Since the directions $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(n-1)}\}$ are linearly independent, we can write the following for some choice of scalars $\gamma_0, \gamma_1, \ldots, \gamma_{n-1}$.

$$\mathbf{x}^* - \mathbf{x}^0 = \ldots\ldots\ldots$$

(c) By premultiplying both sides of this inequality by $\left(\mathbf{d}^k\right)^T A$ and using properties determined so far, we obtain the following expression for γ_k:

$$\gamma_k = \ldots\ldots\ldots$$

(d) $\mathbf{x}^k$ can be expressed in terms of $\mathbf{x}^0, \mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(k-1)}$, *etc.* as:

$$\mathbf{x}^k = \ldots\ldots\ldots$$

(e) By premultiplying the expression by $\left(\mathbf{d}^k\right)^T A$ and using the properties determined so far, we have:

$$\left(\mathbf{d}^k\right)^T A(\mathbf{x}^k - \mathbf{x}^0) = \ldots\ldots\ldots$$

(f) And therefore

$$\left(\mathbf{d}^k\right)^T A(\mathbf{x}^* - \mathbf{x}^0) = \ldots\ldots\ldots$$

(g) Thus $\gamma_k = \ldots\ldots\ldots$, which establishes the result.

7. Consider the half space defined by $\mathcal{H} = \left\{\mathbf{x} \in \Re^n | \mathbf{a}^T\mathbf{x} + \alpha \geq 0\right\}$ where $a \in \Re^n$ and $\alpha \in \Re^n$ are given. Formulate and solve the optimization problem for finding the point $\mathbf{x}$ in $\mathcal{H}$ that has the smallest Euclidean norm.

8. Let the feasible region $\mathcal{D}$ be given as

$$\mathcal{D}: \quad \begin{array}{ll} g_i(\mathbf{x}) \leq \mathbf{0} & \text{for } i = 1 \ldots m \\ h_j(\mathbf{x}) = \mathbf{0} & \text{for } j = 1 \ldots k \end{array}$$

At some feasible point $\mathbf{x}$, let $\mathcal{I}(\mathbf{x})$ be the active index set for the inequality constraints at $\mathbf{x}$, and define the sets $\mathcal{F}(\mathbf{x})$ and $F(\mathbf{x})$ as

$$\mathcal{F}(\mathbf{x}) = \left\{ \mathbf{s} : \begin{array}{ll} g_i(\mathbf{x}+\mathbf{s}) \leq \mathbf{0} & \text{for } i \in \mathcal{I}(\mathbf{x}) \\ h_j(\mathbf{x}+\mathbf{s}) = \mathbf{0} & \text{for } j = 1 \ldots k \end{array} \right\}$$

and

$$F(\mathbf{x}) = \left\{ \mathbf{s} : \begin{array}{ll} \mathbf{s}^T \nabla g_i(\mathbf{x}) \leq \mathbf{0} & \text{for } i \in \mathcal{I}(\mathbf{x}) \\ \mathbf{s}^T \nabla h_j(\mathbf{x}) = \mathbf{0} & \text{for } j = 1 \ldots k \end{array} \right\}$$

(a) Show that if the constraints that are active at $\mathbf{x}$ are all linear, $F(\mathbf{x}) = \mathcal{F}(\mathbf{x})$. This condition is called the *constraint qualificiation* of $\mathbf{x}$.

(b) Suppose the only constraints are given by

- $g_1(x_1, x_2) = x_2 - x_1^3$
- $g_2(x_1, x_2) = -x_2$.

Then, does the constraint qualification assumption hold at $\mathbf{x} = \mathbf{0}$? Prove your statement.

9. Let $\mathcal{I}(\mathbf{x}^*) = \{i_1, i_2, \ldots, i_m\}$ be the active index set at $\mathbf{x}^*$ for the constraints g_i's in the primal problem discussed in class (I guess equation (4.85) in the notes, but please confirm). Show that the set

$$\mathcal{S} = \left\{ \mathbf{s} \mid \mathbf{s}^T \nabla f(\mathbf{x}^*) < 0, \mathbf{s}^T \nabla h_j(\mathbf{x}^*) = 0 \ for\ j = 1 \ldots k,\ and\ \mathbf{s}^T \nabla g_i(\mathbf{x}^*) \leq 0 \ for\ i \in \mathcal{I}(\mathbf{x}^*) \right\}$$

is empty if and only if there exist multipliers λ_j^* for $1 \leq j \leq k$ and $\mu_j^* \geq 0$, such that

$$\nabla f(\mathbf{x}^*) = \sum_{j=1}^{k} \lambda_j^* \nabla h_j(\mathbf{x}^*) - \sum_{i \in \mathcal{I}(\mathbf{x}^*)} \mu_i^* \nabla g_i(\mathbf{x}^*)$$

This lemma is known as the Extension of Farkas lemma.

10. Consider the objective function

$$f(\mathbf{x}) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 100(x_1 - x_4)^4$$

(a) Assume throughout that, for the algorithm in Figure 1, $\rho = 0.1$, $\sigma = 0.1$, $\tau = 0.1$ and $\xi = 0.75$. Is the line seach in Figure 1, (i) exact (ii) approximate (iii) inexact using Wolfe conditions (iv) inexact using Goldstein conditions or (v) none of these? Reason it out.

(b) Solve the problem using the steepest-descent method with stopping criterion $||\alpha_k \mathbf{d}^k|| < \epsilon$ where $\epsilon = 10^{-6}$, using the line search in Figure 1. Report using both initial points $[-2 \quad -1 \quad 1 \quad 2]^T$ and $[200 \quad -200 \quad 100 \quad -100]^T$.

(c) Solve the problem using the modified Newton method:

$$\widehat{H}^k = \frac{\nabla^2 f(\mathbf{x}^k) + \beta I}{1 + \beta}$$

where

$$\beta = \begin{cases} 0 & \text{if} \quad \lambda_{min}\left(\nabla^2 f(\mathbf{x}^k)\right) > 0 \\ 0.25 - \lambda_{min}\left(\nabla^2 f(\mathbf{x}^k)\right) & \text{if} \quad \lambda_{min}\left(\nabla^2 f(\mathbf{x}^k)\right) \leq 0 \end{cases} \tag{7}$$

with the same termination tolerance and initial points as in (b). You should use the line search in Figure 1 where needed.

(d) Solve the problem using the Gauss-Newton method with the same termination tolerance and initial points as in (b). You should use the line search in Figure 1 where needed.

(e) Based on the results of (b)-(d), compare the computational efficiency and solution accuracy of the three methods.

11. By applying quadratic primal active-set algorithm, solve the following QP problem:

$$\begin{aligned} \text{minimize} \quad & f(x) = x_1^2 - x_1x_2 + x_2^2 - 3x_1 \\ \text{subject to} \quad & -x_1 - x_2 \geq -2 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{8}$$

You can use $\mathbf{x}_0 = [0 \ \ 0]^T$.

12. Find and classify (as local or global maximum or minimum or as a saddle point) the stationary points for the following function.

$$f(x) = x_1^2x_2^2 - 4x_1^2x_2 + 4x_1^2 + 2x_1x_2^2 + x_2^2 - 8x_1x_2 + 8x_1 - 4x_2$$

ANS: $[x_1, 2]$ and $[-1, x_2]$ for arbitrary x_1 and x_2 are global minimisers.

13. Find and classify (as local or global maximum or minimum or as a saddle point) the stationary points for the following function.

(a)
$$f(x) = x_1^2 - x_2^2 + x_3^2 - 2x_1x_3 - x_2x_3 + 4x_1 + 12$$

(b)
$$f(x) = x_1^2x_2^2 - 4x_1^2x_2 + 4x_1^2 + 2x_1x_2^2 + x_2^2 - 8x_1x_2 + 8x_1 - 4x_2$$

14. Find $\beta, \theta \in \Re$ for which the function $f(x, y) = \beta(x^2 + y^2) + \theta xy + x + y$

(a) has no stationary points

(b) has exactly one stationary point and it is a global strict minimum

(c) has infinite stationary points, and all of them are global minimizers

15. Prove that if $f(\mathbf{x})$ is a convex function on a convex set $\mathcal{C}$, then the set

$$\mathcal{S} = \{\mathbf{x}\,|\,\mathbf{x} \in \mathcal{C}, f(\mathbf{x}) \leq K\}$$

is convex for every real number K.

16. What is the distance between two parallel hyperplanes $\left\{\mathbf{x} \in \Re^n \,\middle|\, \mathbf{a}^T\mathbf{x} = b\right\}$ and $\left\{\mathbf{x} \in \Re^n \,\middle|\, \mathbf{c}^T\mathbf{x} = d\right\}$?

17. Let $f : \mathcal{D} \to \Re$ where $\mathcal{D} \subseteq \Re^n$. Let $f(\mathbf{x})$ have continuous partial derivatives and continuous mixed partial derivatives in an open region $\mathcal{R}$ containing a point $\mathbf{x}^*$ where $\nabla f(\mathbf{x}^*) = 0$. State and prove necessary condition(s) for $\mathbf{x}^*$ to be a point of local minimum/maximum. You can prove any one of the conditions.

18. If $f : \mathcal{D} \to \Re$ where $\mathcal{D} \subseteq \Re^n$. Let $\mathbf{x}^*$ be a local minimizer of f. Let $f(\mathbf{x})$ have continuous partial derivatives in an open ball $\mathcal{R}$ containing a point $\mathbf{x}^*$. Then for every feasible direction $\mathbf{d}$ at $\mathbf{x}^*$

$$\nabla^T f(\mathbf{x}^*)\mathbf{d} \geq 0$$

19. Prove/disprove mathematically: The complement of a non-convex set is convex.

20. Let $\mathcal{S} \subseteq \Re^n$ be defined as,

$$\mathcal{S} = \left\{\mathbf{x} \in \Re^n \,\middle|\, \mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \leq 0\right\};$$

with A being an $n \times n$ symmetric matrix, $\mathbf{b} \in \Re^n$, and $c \in \Re$. Prove/disprove: $\mathcal{S}$ is convex if and only if $A \succeq 0$.

21. Give an explicit description of the positive semidefinite cone $\mathcal{S}_+^n$, in terms of the matrix coefficients and ordinary inequalities, for $n = 3$. To describe a general element of $\mathcal{S}_+^n$, for $n = 3$, use the following notation for any cone $S \in \mathcal{S}_+^3$

$$S = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_4 & x_5 \\ x_3 & x_5 & x_6 \end{bmatrix}$$

22. Let $f_1(x)$ and $f_2(x)$ be two convex functions such that $f_1(-8) = 7.2$, $f_1(12) = 51.2$, $f_2(-8) = 73.2$, and $f_2(20) = 40$ and define the function $f(x) = max\ \{f_1(x), f_2(x)\}$.

    (a) Is $f(x)$ convex?

(b) Identify the smallest interval in which the minimizer of $f(x)$ is guaranteed to exist. The explanation will carry most of the weightage.
ANS: $[-8, 12]$.

23. A convex quadratic function $\phi(x) : \Re \to \Re$ assumes the values ϕ_1, ϕ_2, and ϕ_3 at $x_0 - \alpha, x_0$, and $x_0 + \alpha$, respectively for some $\alpha > 0$ and $x_0 \in \Re$. Find an expression for the minimum of the function in terms of ϕ_1, ϕ_2 and ϕ_3.

ANS:

$$f_{min} = f_2 - \frac{(f_1 - f_3)^2}{8(f_1 - 2f_2 + f_3)}$$

24. Suppose that $\widetilde{f}(\mathbf{z}) = f(\mathbf{x})$, where $\mathbf{x} = S\mathbf{z} + \mathbf{s}$ for some $S \in \Re^{n\times m}$ and $\mathbf{s} \in \Re^n$. Show that

$$\nabla \widetilde{f}(\mathbf{z}) = S^T \nabla f(\mathbf{x})$$

and

$$\nabla^2 \widetilde{f}(\mathbf{z}) = S^T \nabla^2 f(\mathbf{x}) S$$

25. Suppose three industries are interrelated so that parts of their outputs are used as inputs by themselves, according to the 3×3 consumption matrix

$$C = [c_{ij}] = \begin{bmatrix} 0.2 & 0.3 & 0.1 \\ 0.4 & 0.1 & 0.2 \\ 0.4 & 0.6 & 0.7 \end{bmatrix}$$

where, c_{ij} is the fraction of the output of industry j consumed by industry i. Let p_i be the price charged by industry i for its total output. Find the vector of charges $\mathbf{p} = [p_i]$ so that for each industry, total expenditure equals total income.

$$C = [c_{ij}] = \begin{bmatrix} 0.1 & 0.25 & 0 \\ 0.3 & 0 & 0.15 \\ 0.1 & 0.25 & \alpha \end{bmatrix}$$

where, c_{ij} is the fraction of the output of industry j consumed by industry i. Let p_i be the price charged by industry i for its total output. For each industry, its total expenditure equals half its total income.

(a) Find the value of α.

(b) Find the vector of charges $\mathbf{p} = [p_i]$.

26. Prove that if $\mathcal{C}$ is a convex set of joint probabilities for (x, y), then the associated set of conditional probabilities of x given y is also convex.

27. Prove that the hyperbolic cone $\mathcal{H}$, specified below is convex:

$$\mathcal{H} = \left\{ \mathbf{x} \,\middle|\, \mathbf{x}^T A \mathbf{x} \leq (\mathbf{b}^T \mathbf{x})^2, \mathbf{b}^T \mathbf{x} \geq 0 \right\}$$

where $A \in \mathcal{S}_+^n$ (a positive semi-definite cone) and $\mathbf{b} \in \Re^n$.

ANS: You can make use of covexity preserving operations discussed in the class. The hyperbolic cone is the inverse image of the second-order cone

$$\left\{ (\mathbf{z}, t) \,\middle|\, \mathbf{z}^T \mathbf{z} \leq t^2, t \geq 0 \right\}$$

under the affine function $f(\mathbf{x}) = (A^{1/2}\mathbf{x}, \mathbf{b}^T \mathbf{x})$.

Step 1
Input $\mathbf{x}^k$, $\mathbf{d}^k$.
Initialize algorithm parameters ρ, σ, τ, and ξ.
Set $\alpha_L = 0$ and $\alpha_U = 10^{99}$.
Step 2
Compute $f_L = f(\mathbf{x}^k + \alpha_L \mathbf{d}^k)$.
Compute $f_L' = \nabla^T f(\mathbf{x}^k + \alpha_L \mathbf{d}^k)\mathbf{d}^k$.
Step 3
Estimate α_0 by exact line search on the quadratic approximation for $g(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k)$, as was discussed in class.
Step 4
Compute $f_0 = f(\mathbf{x}^k + \alpha_0 \mathbf{d}^k)$.
Step 5 (Interpolation)
if $f_0 > f_L + \rho(\alpha_0 - \alpha_L)f_L'$ **then**
  If $\alpha_0 < \alpha_U$, then set $\alpha_U = \alpha_0$.
  $\widehat{\alpha}_0 = \alpha_L + \frac{(\alpha_0 - \alpha_L)^2 f_L'}{2[f_L - f_0 + (\alpha_0 - \alpha_L) f_L']}$.
  If $\widehat{\alpha}_0 < \alpha_L + \tau(\alpha_U - \alpha_L)$ then set $\widehat{\alpha}_0 = \alpha_L + \tau(\alpha_U - \alpha_L)$.
  If $\widehat{\alpha}_0 > \alpha_U - \tau(\alpha_U - \alpha_L)$ then set $\widehat{\alpha}_0 = \alpha_U - \tau(\alpha_U - \alpha_L)$.
  Set $\alpha_0 = \widehat{\alpha}_0$ and go to **Step 4**.
end if
Step 6
Compute $f_0' = \nabla^T f(\mathbf{x}^k + \alpha_0 \mathbf{d}^k)\mathbf{d}^k$.
Step 7 (Extrapolation)
if $f_0' < \sigma f_L'$ **then**
  Compute $\Delta\alpha_0 = \frac{(\alpha_0 - \alpha_L) f_0'}{(f_L' - f_0')}$.
  If $\Delta\alpha_0 < \tau(\alpha_0 - \alpha_L)$, then set $\Delta\alpha_0 = \tau(\alpha_0 - \alpha_L)$.
  If $\Delta\alpha_0 > \xi(\alpha_0 - \alpha_L)$, then set $\Delta\alpha_0 = \xi(\alpha_0 - \alpha_L)$.
  Compute $\widehat{\alpha}_0 = \alpha_0 + \Delta\alpha_0$.
  Set $\alpha_L = \alpha_0$, $\alpha_0 = \widehat{\alpha}_0$, $f_L = f_0$, $f_L' = f_0'$, and go to **Step 4**.
end if
Step 8
Output α_0 and $f_0 = f(\mathbf{x}^k + \alpha_0 \mathbf{d}^k)$, and stop.

Figure 1: Line search.