Numerical Optimization: Final Exam Solution

ID: ________________ Name:________________

December 30, 2021

Problem 1: (16 points) Consider the equality constrained problem

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i = 1, ..., m, \end{aligned}$$

where $f, c_i, i = 1, ..., m$ are smooth functions. Show that in this case the IPM (interior-point method) search step is also a SQP (sequential quadratic programming) step.

Answer: There is no inequalities. Therefore, we write the KKT system of it

$$\begin{aligned} \nabla f(x) + \nabla c(x)^T \lambda = 0, \\ c(x) = 0. \end{aligned}$$

At x_k, derive the Newton equation:

$$\begin{bmatrix} \nabla^2_{xx} f(x_k) + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x_k) & \nabla c(x_k)^T \\ \nabla c(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_\lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) - \nabla c(x_k)^T \lambda_k \\ -c(x_k) \end{bmatrix}.$$

Consider the QP subproblem

$$\begin{aligned} \min_{d} \quad & \nabla f(x_k)^T d + \frac{1}{2} d^T \left[\nabla^2_{xx} f(x_k) + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x_k) \right] d \\ \text{s.t.} \quad & c_i(x_k) + \nabla c_i(x)^T d = 0, \quad i = 1, ..., m \end{aligned}$$

with Lagrangian

$$L(d, \mu) = \nabla f(x_k)^T d + \frac{1}{2} d^T \left[\nabla^2_{xx} f(x_k) + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x_k) \right] d + \sum_{i=1}^{m} \left[c_i(x_k) + \nabla c_i(x)^T d \right] \mu_i$$

The KKT conditions for the QP subproblem is

$$\begin{aligned} \nabla_d L(d, \mu) = \nabla f(x_k)^T + \left[\nabla^2_{xx} f(x_k) + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x_k) \right] d + \nabla c(x_k)^T \mu = 0, \\ c_i(x_k) + \nabla c_i(x)^T d = 0, \quad i = 1, ..., m \end{aligned}$$

which can be written as the following matrix form,

$$\begin{bmatrix} \nabla^2_{xx} f(x_k) + \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x_k) & \nabla c(x_k)^T \\ \nabla c(x_k) & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda_k + \delta \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) \\ -c(x_k) \end{bmatrix}.$$

Comparing with the Newton equation, one should notice they are equivalent despite the different naming of variables.

Problem 2: (8 points) Fill out the following form. Put a "✓" to indicate that this case could happen. Put a "×" to indicate that this case cannot happen. (see the example in the bottom right cell).

Primal \ Dual	Infeasible	Unbounded (from above)	Optimal solution exists
Infeasible	✓	✓	×
Unbounded (from below)	✓	×	×
Optimal solution exists	×	×	✓

Problem 3: (20 points = 4×5) For the standard form

$$\min_x \quad c^T x \quad \text{s.t. } Ax = b, x \geq 0.$$

(1) Write down the KKT system of this problem.

Answer: The Lagrangian fo the standard form is

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

where $\lambda \in \mathbb{R}^m, \mu \geq 0$. Then the KKT conditions for the problem is

$$\begin{aligned} \nabla_x L(x, \lambda, \mu) = c^T + \lambda^T A - \mu^T &= 0 \\ Ax - b &= 0 \\ x &\geq 0 \\ \mu &\geq 0 \\ \mu^T x &= 0 \end{aligned}$$

(2) Write down the conditions in the above KKT system, for which the iterates of the primal simplex method must satisfy.

Answer:

$$\begin{aligned} c^T + \lambda^T A - \mu^T &= 0 \\ Ax - b &= 0 \\ x &\geq 0 \\ \mu^T x &= 0 \end{aligned}$$

(3) Write down the conditions in the above KKT system, for which the iterates of the dual simplex method must satisfy.

Answer:

$$\begin{aligned} c^T + \lambda^T A - \mu^T &= 0 \\ Ax - b &= 0 \end{aligned}$$

$$\mu \geq 0$$
$$\mu^T x = 0$$

(4) Write down the conditions in the above KKT system, for which the iterates of the interior point method must satisfy.

Answer:

$$\begin{aligned} c^T + \lambda^T A - \mu^T &= 0 \\ Ax - b &= 0 \\ x &\geq 0 \\ \mu &\geq 0 \end{aligned}$$

(5) Show that the strong duality holds for any KKT point for the LP.

Answer: Suppose there exits x^*, λ^*, μ^* that satisfies the KKT conditions and use f and g to denote the primal and dual functions respectively, then

$$\begin{aligned} g(\lambda^*, \mu^*) &= \min_x L(x, \lambda^*, \mu^*) \\ &= c^T x^* + (\lambda^*)^T (Ax^* - b) - (\mu^*)^T x^* \\ &= c^T x^* \\ &= f(x^*). \end{aligned}$$

Therefore, the duality gap is zero, x^* and λ^*, μ^* are primal and dual optimal respectively.

Problem 4: (17points = 7+10) Suppose we are solving the unconstrained optimization problem

$$\min_x \quad f(x),$$

with f being smooth. We use local model

$$m_k(d) = f(x_k) + \nabla f(x_k)^T d + \tfrac{1}{2} d^T H_k d$$

to approximate $f(x)$ at x_k.

(1) Suppose you were using a multiple of identity matrix αI to approximate the inverse of the Hessian matrix, which may not satisfy the secant equation. Find the α as the least-squares solution of the secant equation.

Answer: The secant equation can be written as

$$s_{k-1} = H_k^{-1} y_{k-1},$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$. The approximate inverse Hessian is

$$H_k^{-1} \approx \alpha I.$$

We aim to find an α that matches the secant equation in the sense that it minimizes the sum of squared errors

$$\min_{\alpha>0} \quad \frac{1}{2}\|\alpha I y_{k-1} - s_{k-1}\|_2^2,$$

This gives

$$\alpha = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|_2^2}.$$

(2) Suppose we are using Quasi-Newton iteration

$$x^{k+1} \leftarrow x_k - {H_k}^{-1}\nabla f(x_k).$$

If f is strongly convex, then the curvature condition

$$s_k^T y_k > 0$$

holds at each iteration if $x_{k+1} - x_k \neq 0$, where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
Answer: Since $f(x)$ is strongly convex, there exists $\sigma > 0$ such that

$$\begin{aligned} f(x^{k+1}) &\geq f(x^k) + \nabla f(x^k)^T(x^{k+1} - x^k) + \tfrac{\sigma}{2}\|x^{k+1} - x^k\|_2^2, \\ f(x^k) &\geq f(x^{k+1}) + \nabla f(x^{k+1})^T(x^k - x^{k+1}) + \tfrac{\sigma}{2}\|x^{k+1} - x^k\|_2^2. \end{aligned}$$

Summing them up, we have

$$[\nabla f(x^k) - \nabla f(x^{k+1})]^T(x^{k+1} - x^k) + \sigma\|x^{k+1} - x^k\|_2^2 \leq 0.$$

This is equivalent to

$$y_k^T s_k = [\nabla f(x^{k+1}) - \nabla f(x^k)]^T(x^{k+1} - x^k) \geq \sigma\|x^{k+1} - x^k\|_2^2 > 0.$$

Problem 5: (25 points = 5+5+15) Suppose we are solving the unconstrained optimization problem

$$\min_x \quad f(x),$$

with f being smooth and (locally) L-Lipschitz differentialble on the level set

$$\mathcal{L} := \{x \mid f(x) \leq f(x_0)\}.$$

(1) Show that for a given descent direction d_k, moving with sufficiently small stepsize can cause decrease in the objective.
Answer: Let the stepsize be α_k, then $x_{k+1} = x_k + \alpha_k d_k$. Applying the first order Taylor expansion of $f(x_{k+1})$ at the point x_k,

$$\begin{aligned} f(x_{k+1}) &= f(x_k + \alpha_k d_k) \\ &= f(x_k) + \langle \nabla f(x_k), \alpha_k d_k\rangle + \frac{\alpha_k^2}{2}(d_k)^T\nabla^2 f(x_k + \theta\alpha_k d_k)d_k, \end{aligned}$$

where $\theta \in (0,1)$. For a sufficiently small α_k, the term

$$\frac{\alpha_k^2}{2}(d_k)^T \nabla^2 f(x_k + \theta \alpha_k d_k) d_k$$

can be ignored. Since $\langle \nabla f(x_k), \alpha_k d_k \rangle < 0$ and $\alpha_k > 0$, it holds that

$$f(x_{k+1}) = f(x_k) + \langle \nabla f(x_k), \alpha_k d_k \rangle < f(x_k),$$

which completes the proof.

(2) Write down the Armijo backtracking condition for a given descent direction (define the parameters you need)
Answer: Let d_k be a descent direction, α_k be the stepsize, and $c_1 \in (0,1)$ be a user-defined constant, then the Armijo condition can be written as

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k,$$

For Armijo backtracking, the stepsize α_k is chosen as the largest value in the set

$$\left\{ \gamma^0, \gamma^1, \gamma^2, \cdots \right\},$$

where $\gamma \in (0,1)$ is a given constant satisfying the Armijo contidion.

(3) Find the maximum number of (objective) function evaluations you need for this linear search.
Answer: For an arbitrary α, we achieve an upper bound of $f(x_k + \alpha d_k)$ by the Lipschitz continuity of ∇f

$$f(x_k + \alpha d_k) \leq f(x_k) + \langle \nabla f(x_k), \alpha d_k \rangle + \frac{\alpha^2}{2} L \|d_k\|_2^2.$$

By making α further satisfying the sufficient decrease condition, it holds

$$f(x_k) + \langle \nabla f(x_k), \alpha d_k \rangle + \frac{\alpha^2}{2} L \|d_k\|_2^2 \leq f(x_k) + c_1 \langle \nabla f(x_k), \alpha d_k \rangle.$$

Therefore, for any $\alpha \in \left[0, \frac{2(c_1 - 1)\langle \nabla f(x_k), d_k \rangle}{L\|d_k\|_2^2}\right]$, the Armijo line search condition is satisfied. And then, the backtracking procedure must end up with

$$\alpha_k \geq 2\gamma \frac{(c_1 - 1)\langle \nabla f(\boldsymbol{x}_k), d_k \rangle}{L\|d_k\|_2^2},$$

where γ is the decay constant as described above. Hence, the maximum number of decaying is

$$\left\lceil \log_{\frac{1}{\gamma}} \frac{1}{\min\{\alpha_k\}} \right\rceil = \left\lceil \log_\gamma 2\gamma \frac{(c_1 - 1)\langle \nabla f(\boldsymbol{x}_k), d_k \rangle}{L\|d_k\|_2^2} \right\rceil.$$

Taking the first evaluation of Armijo condition into account, and assuming two (objective) funciton evaluations required for each Armijo evaluation, then the maximum number of (objective) funciton evaluations is

$$2 + 2\left\lceil \log_\gamma 2\gamma \frac{(c_1 - 1)\langle \nabla f(\boldsymbol{x}_k), d_k \rangle}{L\|d_k\|_2^2} \right\rceil.$$

Problem 6: (14 points) Consider the integer linear programming

$$
\begin{aligned}
\min_x \quad & c^T x \\
\text{s.t.} \quad & Ax = b, \\
& Dx \leq e \\
& x \geq 0 \qquad x \text{ is integer}
\end{aligned}
$$

The Lagrangian relaxation problem is

$$
\begin{aligned}
\min_x \quad & c^T x + \lambda^T (Ax - b) \\
\text{s.t.} \quad & Dx \leq e \\
& x \geq 0 \qquad\qquad\qquad x \text{ is integer}
\end{aligned}
$$

Show that the linear programming relaxation solution has objective value smaller or equal to the Lagrangian relaxation solution.

Answer: Use z_{LP} and z_{LR} to denote the objective values of the linear programming relaxation and the Lagrangian relaxation respectively. Then we have the following relation

$$
\begin{aligned}
z_{\mathrm{LR}} &= \max_{\lambda} \left\{ \min_x c^T x + \lambda^T (Ax - b) \mid Dx \leq e, x \geq 0 \text{ and integer } \right\} \\
&\geq \max_{\lambda} \left\{ \min_x c^T x + \lambda^T (Ax - b) \mid Dx \leq e, x \geq 0 \right\} \\
&= \max_{\lambda} \left\{ \min_x (c + A^T \lambda)^T x - \lambda^T b \mid Dx \leq e, x \geq 0 \right\} \\
&= \max_{\lambda} \left\{ \max_{\mu} \mu^T e - \lambda^T b \mid \mu^T D \leq (c + A^T \lambda)^T, \mu \leq 0 \right\} \\
&= \max_{\lambda, \mu} \left\{ \mu^T e - \lambda^T b \mid \mu^T D \leq (c + A^T \lambda)^T, \mu \leq 0 \right\} \\
&= \max_{\lambda, \mu} \left\{ \mu^T e - \lambda^T b \mid \mu^T D - \lambda^T A \leq c^T, \mu \leq 0 \right\} \\
&= \min_y \left\{ c^T y \mid Ay = b, Dy \leq e, y \geq 0 \right\} \\
&= z_{\mathrm{LP}}
\end{aligned}
$$

Hence, the objective value of LP relaxation is smaller than or equal to the Lagrangian relaxation.