2011 CS5321 Numerical Optimization Midterm

May 16, 15:20-17:20

There are 10 questions with total 110 points. If you think any question is unclear or has ambiguity, write down your assumptions for it, and based on your assumption to answer it.

1. (10) Your cheat sheet
2. (10) Briefly answer the following questions.
   (a) What is symmetric positive definite (SPD)? why it is important in numerical optimization.
   (b) Name two algorithms that can only work for SPD Hessian matrix or reduced Hessian matrix.
3. (10) For a quadratic unconstrained optimization problem,

$$\min_{\vec{x}} f(\vec{x}) = \frac{1}{2}\vec{x}^T Q\vec{x} - \vec{b}^T\vec{x},$$

if Q is symmetric positive definite, at $\vec{x}_k$, given a search direction $\vec{p}_k$, the optimal step length is

$$\alpha_k = -\frac{\vec{p}_k^T \nabla f(\vec{x}_k)}{\vec{p}_k^T Q\vec{p}_k}.$$

Show the optimal step length you derived satisfies the Goldstein conditions:

$$f(\vec{x}_k) + (1-c)\alpha_k \nabla\vec{p}_k^T f(\vec{x}_k) \le f(\vec{x}_k + \alpha_k\vec{p}_k) \le f(\vec{x}_k) + c\alpha_k\vec{p}_k^T\nabla f(\vec{x}_k),$$

for $0 < c < 1/2$.

$$\begin{aligned} f(\vec{x}_k + \alpha_k\vec{p}_k) &= \frac{1}{2}(\vec{x}_k + \alpha_k\vec{p}_k)^T Q(\vec{x}_k + \alpha_k\vec{p}_k) - \vec{b}^T(\vec{x}_k + \alpha_k\vec{p}_k) \\ &= \left[\frac{1}{2}\vec{x}_k^T Q\vec{x}_k - \vec{b}^T\vec{x}_k\right] + \frac{1}{2}\alpha_k^2\vec{p}_k^T Q\vec{p}_k + \alpha_k\vec{p}_k^T(Q\vec{x}_k\vec{b}) \end{aligned}$$

The term $Q\vec{x}_k\vec{b} = \nabla f(\vec{x}_k)$, and $f(\vec{x}_k) = \frac{1}{2}\vec{x}_k^T Q\vec{x}_k - \vec{b}^T\vec{x}_k$. Using the definition of α_k and above relations, we have

$$f(\vec{x}_k + \alpha_k\vec{p}_k) = f(\vec{x}_k) + \frac{1}{2}\alpha_k\vec{p}_k^T\nabla f(\vec{x}_k)$$

which satisfies the Goldenstein conditions for $0 < c < 1/2$.

4. (10) Let B_k be the BFGS approximation to the Hessian H_k matrix. The formula of updating B_k is

$$B_{k+1} = B_k - \frac{B_k\vec{y}_k\vec{y}_k^T B_k}{\vec{y}_k^T B_k\vec{y}_k} + \frac{\vec{s}_k\vec{s}_k^T}{\vec{y}_k^T\vec{s}_k}, \tag{1}$$

where $\vec{s}_k = \vec{x}_{k+1} - \vec{x}_k$, $\vec{y}_k = \nabla f_{k+1} - \nabla f_k$. Show that if $\vec{s}_k^T\vec{y}_k > 0$ and B_k is SPD, B_{k+1} obtained by (1) is SPD. (Hint: (1) you need to prove B_{k+1} is symmetric first, (2) use the secant equation.)

$B_{k+1}^T = B_k - \dfrac{B_k\vec{y}_k\vec{y}_k^T B_k}{\vec{y}_k^T B_k\vec{y}_k} + \dfrac{\vec{s}_k\vec{s}_k^T}{\vec{y}_k^T\vec{s}_k} = B_{k+1}$, so it is symmetric.

$$\begin{aligned} \vec{x}^T B_{k+1}\vec{x} &= \vec{x}^T B_k\vec{x} - \frac{\vec{x}^T B_k\vec{y}_k\vec{y}_k^T B_k\vec{x}}{\vec{y}_k^T B_k\vec{y}_k} + \frac{\vec{x}^T\vec{s}_k\vec{s}_k^T\vec{x}}{\vec{y}_k^T\vec{s}_k} \\ &= \frac{\vec{x}^T B_k\vec{x}\vec{y}_k^T B_k\vec{y}_k - \vec{x}^T B_k\vec{y}_k\vec{y}_k^T B_k\vec{x}}{\vec{y}_k^T B_k\vec{y}_k} + \frac{\vec{x}^T\vec{s}_k\vec{s}_k^T\vec{x}}{\vec{y}_k^T\vec{s}_k} \end{aligned}$$

I could not find a simple way to prove the SPD part by the hints I gave, so the point of this question is given for grace.

An easier way to prove it is to show B_{k+1}^{-1} is SPD, which is left for exercise.

5. (10) Consider the linear least square problem:

$$\min_{\vec{x}\in\mathbb{R}^2} \|A\vec{x} - \vec{b}\|^2,$$

where

$$A = \begin{pmatrix} 4 & 8 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}, \vec{b} = \begin{pmatrix} 21/4 \\ 0 \\ 0 \end{pmatrix}$$

(a) Write its normal equation.

$$A^T A\vec{x} = A^T\vec{b} \Rightarrow \begin{pmatrix} 21 & 42 \\ 42 & 84 \end{pmatrix}\vec{x} = \begin{pmatrix} 21 \\ 42 \end{pmatrix}$$

(b) Express $\vec{b} = \vec{b}_1 + \vec{b}_2$ such that $\vec{b}_1$ is in the subspace spanned by A's column vectors, and $\vec{b}_2$ is orthogonal to A's column vectors.

$$\vec{b}_1 = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 5/4 \\ -2 \\ -1 \end{pmatrix}$$

6. (10) Consider the linear programming problem:

$$\begin{array}{ll} \min_{\vec{x}} & -400x_1 + 600x_2 - 100x_3 - 950x_4 \\ \text{s.t.} & 2x_1 + x_2 + x_3 + x_5 + 3x_6 = 8 \\ & x_1 + x_2 - x_3 + x_4 + 4x_6 = 5 \\ & x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{array}$$

Suppose the current $\vec{x} = (0, 0, 0, 0, 17/4, 5/4)$ and the simplex method wants to increase x_4.

(a) What is the search direction?
$\vec{p}^T = (0, 0, 0, 1, 3/4, -1/4)$

(b) What is the step length?
$\alpha = 5$

7. (10) Solve the problem

$$\begin{array}{ll} \min\limits_{\vec{x} \in \mathbb{R}^n} & \vec{c}^T \vec{x} \\ \text{s.t.} & \vec{x}^T \vec{e} = x_1 + x_2 + \cdots + x_n = 0 \\ & \vec{x}^T \vec{x} = 1 \end{array}$$

Partial credit will be given if only the case of $n = 2$ or $n = 3$ is considered.

$$\mathcal{L} = \vec{c}^T \vec{x} - \lambda_1 (\vec{x}^T \vec{e}) - \lambda_2 (\vec{x}^T \vec{x} - 1)$$

Use the first order condition,

$$\nabla_x \mathcal{L} = \vec{c} - \lambda_1 \vec{e} - \lambda_2 \vec{x} = 0.$$

So,

$$\vec{x} = \frac{1}{\lambda_2} (\vec{c} - \lambda_1 \vec{e}).$$

Also, $\vec{x}^T \vec{e} = 0$ implies $\vec{c}^T \vec{e} - \lambda_1 \vec{e}^T \vec{e} = 0$, which concludes $\lambda_1 = \vec{c}^T \vec{e}/n$ where n is the dimension.

Using $\vec{x}^T\vec{x} = 1$, we have

$$\frac{1}{\lambda_2^2}(\vec{c} - \lambda_1\vec{e})^T(\vec{c} - \lambda_1\vec{e}) = 1.$$

Thus, $\lambda_2 = \|\vec{c} - \lambda_1\vec{e}\|$.

$$\vec{x} = \frac{\vec{c} - \vec{c}^T\vec{e}/n\vec{e}}{\|\vec{c} - \vec{c}^T\vec{e}/n\vec{e}\|}$$

8. (10) Consider the quadratic programming problem

$$\begin{array}{ll} \min_{\vec{x}} & \frac{1}{2}\vec{x}^T G\vec{x} + \vec{c}^T\vec{x} \\ \text{s.t.} & A\vec{x} \geq \vec{b} \end{array}$$

where G is symmetric positive definite. What is its dual problem? Express it in the following form

$$\begin{array}{ll} \max_{\vec{\lambda}} & q(\vec{\lambda}) \\ \text{s.t.} & \vec{\lambda} \geq 0 \end{array} \tag{2}$$

Partial credit will be given if the objective function q contains variable $\vec{x}$. (Hint: use Wolfe's duality.)

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{c}^T\vec{x} - \vec{\lambda}^T(A\vec{x} - \vec{b})$$

Wolfe's duality says

$$\nabla_x\mathcal{L}(\vec{x}, \vec{\lambda}) = G\vec{x} + \vec{c} - A^T\vec{\lambda} = 0$$

$$\vec{x} = G^{-1}(A^T\vec{\lambda} - \vec{c}).$$

Plugging that into the Lagrangian function,

$$\begin{aligned} \mathcal{L}(\vec{x}, \vec{\lambda}) &= \frac{1}{2}(A^T\vec{\lambda} - \vec{c})^T G^{-1}GG^{-1}(A^T\vec{\lambda} - \vec{c}) + (\vec{c}^T - \vec{\lambda}^T A)G^{-1}(A^T\vec{\lambda} - \vec{c}) + \vec{\lambda}^T\vec{b} \\ &= \frac{-1}{2}(A^T\vec{\lambda} - \vec{c})^T G^{-1}(A^T\vec{\lambda} - \vec{c}) + \vec{\lambda}^T\vec{b} \end{aligned}$$

So the dual problem is

$$\begin{array}{ll} \max_{\vec{\lambda}} & \frac{-1}{2}(A^T\vec{\lambda} - \vec{c})^T G^{-1}(A^T\vec{\lambda} - \vec{c}) + \vec{\lambda}^T\vec{b} \\ \text{s.t.} & \vec{\lambda} \geq 0 \end{array}$$

9. (10) Suppose the sequential quadratic programming method is used, and the formula of local QP model is given as follows:

$$\begin{array}{ll} \min\limits_{\vec{p}} & \frac{1}{2}\vec{p}^T \nabla^2_{xx}\mathcal{L}_k \vec{p} + \nabla f_k^T \vec{p} \\ \text{s.t.} & \nabla c_k^T \vec{p} + c_k = 0 \end{array}$$

Consider the problem

$$\begin{array}{ll} \min\limits_{x_1,x_2} & 2(x_1^2+x_2^2-1) - x_1 \\ \text{s.t.} & x_1^2+x_2^2-1 = 0 \end{array} \tag{3}$$

What is the local quadratic programming model for (3) at $(\cos(\theta), \sin(\theta))^T$?

Just use the formula: First, we have

$$\mathcal{L} = 2(x_1^2+x_2^2-1) - x_1 - \lambda(x_1^2+x_2^2-1)$$

$$\nabla_x \mathcal{L} = \begin{pmatrix} 4x_1 - 1 - 2\lambda x_1 \\ 4x_2 - 2\lambda x_2 \end{pmatrix} = \begin{pmatrix} (4-2\lambda)\cos(\theta) - 1 \\ (4-2\lambda)\sin(\theta) \end{pmatrix},$$

$$\nabla_{xx}\mathcal{L} = \begin{pmatrix} 4-2\lambda & 0 \\ 0 & 4-2\lambda \end{pmatrix}, \nabla f(\vec{x}_k) = \begin{pmatrix} 4\cos(\theta) - 1 \\ 4\sin(\theta) \end{pmatrix}$$

$$\nabla_x c = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 2\cos(\theta) \\ 2\sin(\theta) \end{pmatrix}, c(\vec{x}_k) = 0.$$

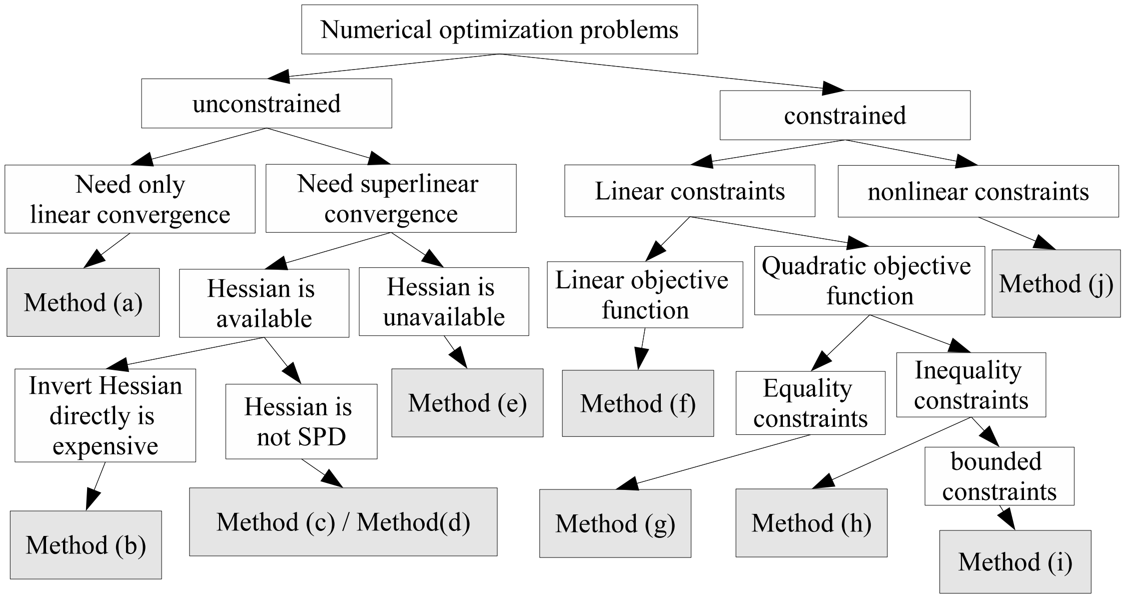
Therefore, the model problem is

$$\begin{array}{ll} \min\limits_{p_1,p_2,\lambda} & (2-\lambda)p_1^2 + (2-\lambda)p_2^2 + 4\cos(\theta)p_1 + 4\sin(\theta)p_2 - p_1 \\ \text{s.t.} & 2\cos(\theta)p_1 + 2\sin(\theta)p_2 = 0 \end{array}$$

or

$$\begin{array}{ll} \min\limits_{p_1,p_2,\lambda} & (2-\lambda)p_1^2 + (2-\lambda)p_2^2 - p_1 \\ \text{s.t.} & \cos(\theta)p_1 + \sin(\theta)p_2 = 0 \end{array}$$

10. (20) To your best knowledge, find at least one method for each gray block.


Numerical optimization problems
unconstrained
constrained
Need only linear convergence
Need superlinear convergence
Linear constraints
nonlinear constraints
Method (a)
Hessian is available
Hessian is unavailable
Linear objective function
Quadratic objective function
Method (j)
Invert Hessian directly is expensive
Hessian is not SPD
Method (e)
Method (f)
Equality constraints
Inequality constraints
bounded constraints
Method (b)
Method (c) / Method(d)
Method (g)
Method (h)
Method (i)