Numerical Analysis

11/12/2012

1. Let $\mathbf{A}$ be a n-by-n symmetric matrix. Let $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x}$ in R^n. Vectors $\mathbf{p}_i$ and $\mathbf{p}_j$ are conjugate if $\langle \mathbf{p}_i, \mathbf{p}_j \rangle_A \equiv \mathbf{p}_i^T \mathbf{A} \mathbf{p}_j = 0$. Suppose $\{\mathbf{p}_j\}_{j=1}^n$ is a sequence of n mutually conjugate directions.

   (a) (10 points) Show that $\{\mathbf{p}_j\}_{j=1}^n$ forms a basis of R^n.

   (b) (10 points) We can expand the solution $\mathbf{x}^*$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$ in this basis, such as

   $$\mathbf{x}^* = \sum_{i=1}^{n} \alpha_i \mathbf{p}_i.$$

   Show that $\alpha_i = \frac{\langle \mathbf{p}_i, \mathbf{b} \rangle}{\langle \mathbf{p}_i, \mathbf{p}_i \rangle_{\mathbf{A}}}$.

   (c) (10 points) Let $\{\mathbf{x}_k\}$ denote a sequence of solutions derived by the conjugate gradient method. Let $\mathbf{r}_k$ be the residual at the ith iteration. Then $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$. Show the relation between $\mathbf{r}_k$ and the gradient of a quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x}$.

   (d) (20 points) Let $\mathbf{x}_0$ be an initial random guess, $\mathbf{r}_0 = \mathbf{A}\mathbf{x}_0 - \mathbf{b}$ and $\mathbf{p}_0 = -\mathbf{r}_0$. By the Gram-Schmidt orthonormalization,

   $$\mathbf{p}_{k+1} = -\mathbf{r}_k + \sum_{i \le k} \frac{\langle \mathbf{p}_i, \mathbf{r}_k \rangle_{\mathbf{A}}}{\langle \mathbf{p}_i, \mathbf{p}_i \rangle_{\mathbf{A}}} \mathbf{p}_i.$$

   Show all $\mathbf{p}_i$ paiwisely conjugate.

   (e) Let $\{\mathbf{p}_j\}_{j=0}^k$ be conjugate vectors obtained in (d). Show

      i. (10 points) $\mathbf{W}_k = span\{\mathbf{r}_0, ..., \mathbf{r}_{k-1}\} = span\{\mathbf{p}_0, ..., \mathbf{p}_{k-1}\}$

      ii. (10 points) $\mathbf{p}_k^T \mathbf{r}_j = -\mathbf{r}_k^T \mathbf{r}_k$ for all $0 \le j < k$

iii. (10 points) $\mathbf{p}_k = -\mathbf{r}_k + \beta_{k-1}\mathbf{p}_{k-1}$, where $\beta_{k-1} = \frac{\mathbf{r}_k^T\mathbf{r}_k}{\mathbf{r}_{k-1}^T\mathbf{r}_{k-1}}$.

2. (15 points) Draw a flow chart to illustrate applying the conjugate gradient method for minimizing

$$Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x},$$

where $\mathbf{A}$ is positive definite.

3. Let

$$f(\mathbf{x};\boldsymbol{\theta}) = \sum_{m=1}^{M} r_m \exp\left(-\frac{\|\mathbf{x}-\mathbf{a}_m\|^2}{s_m^2}\right) + r_{M+1},$$

where $\mathbf{x} \in R^d$ and $\boldsymbol{\theta} = (\mathbf{a}_1^T\ \mathbf{a}_2^T \dots \mathbf{a}_M^T\ s_1\ s_2\ \dots\ s_M\ r_1\ \dots\ r_M\ r_{M+1}\ )^T$ denotes collection of hyper parameters in f. The length of $\boldsymbol{\theta}$ is $dM + 2M + 1$.

(a) (10 points) Let $\mathbf{A}$ be an $M \times d$ matrix, $\mathbf{a}_m^T$ denote the mth row of matrix $\mathbf{A}$, $\mathbf{s} = [s_1, \dots, s_M]$, $\mathbf{r} = [r_1, \dots, r_M, r_0]$. Write a matlab function to extract $\mathbf{A}, \mathbf{s}$ and $\mathbf{r}$ from given $\boldsymbol{\theta}$.

(b) (5 points) Write a matlab function to form $\boldsymbol{\theta}$ for given $\mathbf{A}, \mathbf{s}$ and $\mathbf{r}$.

(c) (5 points) Write a matlab function to evaluate the output of f for given $\mathbf{x}$ and $\boldsymbol{\theta}$.

(d) (10 points) Let $\mathbf{X}$ be an $N \times d$ matrix and $\mathbf{Y}$ be an $N \times 1$ vector. Write a matlab function to calculate the following mean square error for given X, Y and $\boldsymbol{\theta}$,

$$E(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(\mathbf{x}_i;\boldsymbol{\theta}))^2, \tag{1}$$

where $\mathbf{x}_i^T$ denotes the ith row of $\mathbf{X}$ and y_i denotes the ith element of $\mathbf{Y}$.

(e) Calculation of the gradient of $E(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

i. (15 points) Derive $\frac{df(\mathbf{x};\boldsymbol{\theta})}{d\mathbf{a}_m}$, $\frac{df(\mathbf{x};\boldsymbol{\theta})}{ds_m}$ and $\frac{df(\mathbf{x};\boldsymbol{\theta})}{dr_m}$ respectively.

ii. (10 points) Let ga be an $N \times Md$ matrix and $ga(i,(m-1)d+1:md)$ denote a vector obtained by substituting $\mathbf{x}_i$ and current $\boldsymbol{\theta}$ to $\frac{df(\mathbf{x};\boldsymbol{\theta})}{d\mathbf{a}_m}$. Draw a flow chart to illustrate how to determine matrix ga.

iii. (10 points) Let gs be an $N \times M$ matrix and $gs(i,m)$ denote a vector obtained by substituting $\mathbf{x}_i$ and current $\boldsymbol{\theta}$ to $\frac{df(\mathbf{x};\boldsymbol{\theta})}{ds_m}$. Draw a flow chart to illustrate how to determine matrix gs.

iv. (10 points) Let gr be an $N \times (M+1)$ matrix and $gr(i,m)$ denote a vector obtained by substituting $\mathbf{x}_i$ and current $\boldsymbol{\theta}$ to $\frac{df(\mathbf{x};\boldsymbol{\theta})}{dr_m}$. Draw a flow chart to illustrate how to determine matrix gr.

v. Let $g\boldsymbol{\theta} = [ga\ gs\ gr]$ be an $N \times L$ matrix, where $L = Md + 2M + 1$. The ith row of $g\boldsymbol{\theta}$ represents the result of substituting $\mathbf{x}_i$ and current $\boldsymbol{\theta}$ to $\frac{df(\mathbf{x};\boldsymbol{\theta})}{d\boldsymbol{\theta}}$. Substituting current $\boldsymbol{\theta}$ to $\frac{dE(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ is a vector of length L which represents the gradient of $E(\boldsymbol{\theta})$ at current $\boldsymbol{\theta}$.

A. (10 points) State how to determine the gradient of $E(\boldsymbol{\theta})$ at current $\boldsymbol{\theta}$ for given $g\boldsymbol{\theta}$.

B. (10 points) Write a matlab function for determining $\frac{dE(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ at current $\boldsymbol{\theta}$.

4. $\{\boldsymbol{\theta}_k\}_k$ denotes a sequence of $\boldsymbol{\theta}$ obtained by an iterative approach for minimizing $E(\boldsymbol{\theta})$ in equation (1) with respect to $\boldsymbol{\theta}$. Typically it is represented by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k.$$

The gradient method sets $\Delta\boldsymbol{\theta}_k$ proportional to the negative gradient $E(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_k$, mathematically expressed by

$$\Delta\boldsymbol{\theta}_k \propto -\nabla(\boldsymbol{\theta}_k),$$

where

$$\nabla(\boldsymbol{\theta}_k) = \frac{dE(\boldsymbol{\theta})}{d\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k}.$$

5. (10 points) Draw a flow chart to illustrate applying the gradient method for minimizing $E(\boldsymbol{\theta})$ in equation (1) with respect to $\boldsymbol{\theta}$.

6. (10 points) The gradient method can be improved by the technique of line minimization. The step size η is optimized such that

$$\eta_{otp} = \min_{\{\eta\}} E(\boldsymbol{\theta}_k - \eta\nabla(\boldsymbol{\theta}_k)).$$

   (a) (10 points) Write a matlab function to evaluate $E(\boldsymbol{\theta}_k - \eta\nabla(\boldsymbol{\theta}_k))$ for given $\boldsymbol{\theta}_k, \eta$ and $\nabla(\boldsymbol{\theta}_k)$.

   (b) (10 points) Let $E(\eta)$ the output of function E for fixed $\boldsymbol{\theta}_k$ and $\nabla(\boldsymbol{\theta}_k)$. The matlab function mnbrek.m is a function that is able to adjust given a, b and c to form a bracket that satifies the following condition

$$\begin{aligned} a &< b < c, \\ E(b) &< E(a) \text{ and } E(b) < E(c). \end{aligned}$$

   Draw a flow chart to illustrate how to adjust given a, b and c to form a bracket.

   (c) Golden section search for given bracket a, b and c.

      i. (10 points) Let $a < b < c$. The golden search sets x to $b + W * (c - b)$ if $|b - c| > |b - a|$ and $b - W * (b - a)$. State two criteria for deriving $W = \frac{3-\sqrt{5}}{2}$.

ii. (10 points) Draw a flow chart to illustrate golden section search of minimizing $E(\eta)$ for given a, b and c.

(d) Parabolic interpolation for given bracket a, b and c:

i. (10 points) Express a quadratic polynomial that interpolates $(a, E(a)), (b, E(b))$ and $(c, E(c))$ and find x_{opt} that minimizes $E(\eta)$, where η is within the given bracket.

ii. (10 points) Draw a flow chart to illustrate minimizing $E(\eta)$ by iteratively finding the minimum of quadratic interpolating polynomials.

7. The nonlinear conjugate gradient method combines the conjugate gradient method with the line minimization method.

(a) (10 points) Draw a flow chart to illustrate minimizing $E(\boldsymbol{\theta})$ by the nonlinear conjugate gradient method.

(b) (100 points) Write matlab programs to implement the nonlinear conjugate gradient method for minimizing $E(\boldsymbol{\theta})$.

8. (30 points) Let $\mathbf{z} = \{z[t]\}_t$ and $\mathbf{x} = \{x[t]\}_t$. Series $\mathbf{x}$ is regadred as linear convolution of series z through a kernel $\mathbf{a} = (a_0, ..., a_{L-1})$, if

$$x[t] = \sum_{i=0}^{L-1} a_i z[t-i] + n_t \text{ for } L \le t \le N,$$

where $\{n_t\}_t$ denotes noises.

(a) (10 points) Design an objective function $E(\mathbf{a})$ whose minimization with respect to $\mathbf{a}$ leads to an optimal $\mathbf{a}$ for given $\mathbf{z}$ and $\mathbf{x}$.

(b) (10 points) Derive $\mathbf{a}_{opt}$ such that

$$\mathbf{a}_{opt} = \min_{\{\mathbf{a}\}} E(\mathbf{a}).$$

(c) (10 points) Write a matlab function to calculate $\mathbf{a}_{opt}$ for given $\mathbf{x}$ and $\mathbf{z}$.