Math 408 Homework Set 5 Solutions

(1) Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite, and let $d_1, d_2, \ldots, d_n \in \mathbb{R}^n$ be a basis of Q-conjugate vectors. Show that

$$Q^{-1} = \sum_{i=1}^n \frac{d_i d_i^T}{d_i^T Q d_i} \ .$$

Solution: Let $M := \sum_{i=1}^n \frac{d_i d_i^T}{d_i^T Q d_i}$. Then, for every $x \in \mathbb{R}^n$, we have $x = \sum_{j=1}^n \lambda_j d_j$ for a unique set of coefficients $\lambda_1, \ldots, \lambda_n$ since $d_1, d_2, \ldots, d_n$ for a basis of $\mathbb{R}^n$. But then

$$Mx = \sum_{i=1}^n \frac{d_i}{d_i^T Q d_i}(d_i Q \sum_{j=1}^n \lambda_j d_j) = \sum_{i=1}^n \frac{d_i}{d_i^T Q d_i}(\sum_{j=1}^n \lambda_j d_i Q d_j) = \sum_{i=1}^n \frac{d_i}{d_i^T Q d_i}\lambda_i d_i Q d_i = \sum_{j=1}^n \lambda_j d_j = x,$$

which implies that $M = Q^{-1}$.

(2) Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, and suppose that $\mathrm{Nul}(A) = \{0\}$. Consider the function

$$f(x) := \frac{1}{2}\|Ax - b\|_2^2 \ .$$

Show how to apply the CG algorithm to minimize f on $\mathbb{R}^n$ under the assumption that A is not available but Ax and $A^T y$ can be obtained for arbitrary vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, respectively.

Solution:
Initialization: $x_0 \in \mathbb{R}^n$, $r_0 = Ax_0 - b$, $d_0 = -g_0 = -A^T r_0$.
For $k = 0, 1, 2, \ldots$

$$\begin{array}{lll}
\alpha_k & := -r_k^T A d_k / \|Ad_k\|_2^2 & \\
x_{k+1} & := x_k + \alpha_k d_k & \\
r_{k+1} & := Ax_{k+1} - b & \\
g_{k+1} & := A^T r_{k+1} & \text{(STOP if } g_{k+1} = 0) \\
\beta_k & := (Ag_{k+1})^T (Ad_k) / \|Ad_k\|_2^2 & \\
d_{k+1} & := -g_{k+1} + \beta_k d_k & \\
k & := k + 1. &
\end{array}$$

(3) Apply the conjugate gradient algorithm initialized at $x^0 = (0, 0, 0)^T$ to solve the problem $\min_{x\in\mathbb{R}^3} \frac{1}{2} x^T H x + g^T x$, where

$$H = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad g = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Solution: Later.

(4) Let $f(x) := \frac{1}{2} x^T H x + g^T x$, where $H \in \mathcal{S}^n_{++}$ and $g \in \mathbb{R}^n$. Let $\{x^0, \ldots, x^n\}$ be the iterates of the conjugate gradient algorithm applied to f initiated at $x^0 = 0$.

(a) Show that $f(x^k) - f(x^{k+1}) = \frac{1}{2}\frac{\langle g^k\,,\,d^k\rangle^2}{\langle Hd^k\,,\,d^k\rangle}$, which establishes that $f(x^k) > f(x^{k+1})$ whenever $d^k \neq 0$.

Solution: Since $t_k = -\frac{\langle g^k\,,\,d^k\rangle}{(d^k)^T H d^k}$, we have

$$f(x^{k+1}) = f(x^k) + t_k \left\langle g^k\,,\,d^k\right\rangle + \frac{t_k^2}{2}(d^k)^T H d^k = f(x^k) - \frac{\left\langle g^k\,,\,d^k\right\rangle^2}{(d^k)^T H d^k}.$$

(b) Show that the sequence $\{\beta_k\}$ generated by the CG-algorithm satisfies $\beta_k = \frac{\langle g^{k+1}\,,\,g^{k+1}-g^k\rangle}{\|g^k\|^2}$.

Solution:

$$\begin{aligned}\beta_k &= \frac{\left\langle g^{k+1}\,,\,Hd^k\right\rangle}{\left\langle d^{kT}\,,\,Hd^k\right\rangle} = \frac{\left\langle g^{k+1}\,,\,t_k^{-1}H(x^{k+1}-x^k)\right\rangle}{\langle d^k\,,\,Hd^k\rangle} \\ &= \frac{\left\langle g^{k+1}\,,\,(Hx^{k+1}+g)-(Hx^k+g)\right\rangle}{\langle -g^k\,,\,d^k\rangle} = \frac{\left\langle g^{k+1}\,,\,g^{k+1}-g^k\right\rangle}{\langle -g^k\,,\,-g^k+\beta_{k-1}d^{k-1}\rangle} \\ &= \frac{\left\langle g^{k+1}\,,\,g^{k+1}-g^k\right\rangle}{\langle g^k\,,\,g^k\rangle}.\end{aligned}$$

(5) Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, and consider the function

$$f(x) := \frac{1}{2}\|Ax-b\|_2^2\ .$$

Show that $\nabla f(x) = A^T(Ax-b)$ and $\nabla^2 f(x) = A^T A$.

Solution:

$$\begin{aligned}f(x+\Delta x) &= \frac{1}{2}\|A(x+\Delta x)-b\|_2^2 \\ &= \frac{1}{2}\|(Ax-b)+A\Delta x)-b\|_2^2 \\ &= \frac{1}{2}\|Ax-b\|_2^2 + \left\langle A^T(Ax-b)\,,\,\Delta x\right\rangle + \frac{1}{2}\|A\Delta\|^2 \\ &= f(x) + \left\langle A^T(Ax-b)\,,\,\Delta x\right\rangle + \frac{1}{2}(\Delta x)^T(A^T A)(\Delta x).\end{aligned}$$

(6) Let $f(x) := \frac{1}{2}x^T Hx + g^T x$, where $H \in \mathbb{R}^{n\times n}$ and $g \in \mathbb{R}^n$.
(a) Show that $\nabla f(x) = \frac{1}{2}(H+H^T)x + g$ so that if $H \in \mathcal{S}^n$, then $\nabla f(x) = Hx + g$.
(b) Show that $\nabla^2 f(x) = \frac{1}{2}(H+H^T)$ so that if $H \in \mathcal{S}^n$, then $\nabla^2 f(x) = H$.

Solution:

$$\begin{aligned}f(x+\Delta x) &= \frac{1}{2}(x+\Delta x)^T H(x+\Delta x) + g^T(x+\Delta x) \\ &= f(x) + \frac{1}{2}((\Delta x)^T Hx + x^T H\Delta x) + g^T\Delta x + \frac{1}{2}(\Delta x)^T H\Delta x \\ &= f(x) + \left\langle \frac{1}{2}(H+H^T)x+g)\,,\,\Delta x\right\rangle + (\Delta x)^T(\frac{1}{2}(H+H^T))\Delta x.\end{aligned}$$

(7) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function (has as many continuous derivatives as you want), and define $f(x) := \frac{1}{2}\|F(x)\|_2^2$.

(a) Show that $\nabla f(x) = \nabla F(x)^T F(x)$.

(b) Show that $\nabla^2 f(x) = \nabla F(x)^T \nabla F(x) + \sum_{j=1}^m F_j(x)\nabla^2 F_j(x)$, where $F_j : \mathbb{R}^n \to \mathbb{R}$ are the component functions of F, $j = 1, \ldots, m$.

Solution: Let F_i be the ith component function of F so that $f(x) = \frac{1}{2}\sum_{j=1}^m (F_i(x))^2$. For each $j = 1, \ldots, m$, we have

$$F_i(x + \Delta x) = F(x) + \nabla F_i(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 F_i(x)\Delta x + o(\|\Delta x\|^2).$$

Hence,

$$\frac{1}{2}(F_i(x+\Delta x))^2 = \frac{1}{2}(F_i(x))^2 + \langle F_i(x)\nabla F_i(x)\, ,\, \Delta x\rangle + \frac{1}{2}\Delta x^T(\nabla F_i(x)\nabla F_i(x)^T + F_i(x)\nabla^2 F_i(x))\Delta x + o(\|\Delta x\|^2).$$

Consequently,

$$\begin{aligned}
f(x+\Delta x) &= \frac{1}{2}\sum_{j=1}^m (F_i(x+\Delta x))^2 \\
&= \sum_{j=1}^m \frac{1}{2}(F_i(x))^2 + \langle F_i(x)\nabla F_i(x)\, ,\, \Delta x\rangle + \frac{1}{2}\Delta x^T(\nabla F_i(x)\nabla F_i(x)^T + F_i(x)\nabla^2 F_i(x))\Delta x + o(\|\Delta x\|^2) \\
&= f(x) + \left\langle \nabla F(x)^T F(x)\, ,\, \Delta x \right\rangle + \frac{1}{2}\Delta x^T\left(\nabla F(x)^T\nabla F(x) + \sum_{j=1}^m F_i(x)\nabla^2 F_i(x)\right)\Delta x + o(\|\Delta x\|^2),
\end{aligned}$$

since

$$\nabla F(x) = \begin{bmatrix} \nabla F_1(x)^T \\ \nabla F_2(x)^T \\ \vdots \\ \nabla F_m(x)^T \end{bmatrix} \quad \text{so that} \quad \nabla F(x)^T \nabla F(x) = \sum_{j=1}^m \nabla F_i(x)\nabla F_i(x)^T \ .$$

(8) Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $V \in \mathbb{R}^{n\times k}$, and $\hat{x} \in \mathbb{R}^n$, where $k < n$. Consider the function given by $h(z) := f(\hat{x} + Vz)$.

(a) Show that $\nabla h(z) = V^T \nabla f(\hat{x} + Vz)$.

Solution:

$$\begin{aligned}
h(z+\Delta z) &= f(\hat{x} + V(z\Delta z)) = f((\hat{x} + Vz) + V\Delta z) \\
&= f(\hat{x}) + \langle \nabla f(\hat{x} + Vz)\, ,\, V\Delta z\rangle + o(\|V\Delta z\|) \\
&= f(\hat{x}) + \left\langle V^T\nabla f(\hat{x} + Vz)\, ,\, \Delta z\right\rangle + o(\|\Delta z\|).
\end{aligned}$$

(b) Show that if $\bar{z}$ is a local solution to $\min_z h(z)$, then $\nabla f(\hat{x} + V\bar{z}) \perp \mathrm{Ran}\,(V)$.

Solution: If $\bar{z}$ is a local solution to $\min_z h(z)$, then $0 = \nabla h(\bar{z}) = V^T\nabla f(\hat{x} + V\bar{z})$, or equivalently, $\nabla f(\hat{x} + V\bar{z}) \in \mathrm{Nul}\left(V^T\right) = \mathrm{Ran}\,(V)^\perp$.

(c) Let S be a subspace of $\mathbb{R}^n$. Show that if $\bar{x} \in \hat{x} + S$ solves $\min\{f(x)\ |x \in \hat{x} + S\}$, then $\nabla f(\bar{x}) \perp S$.

Solution: Let the columns of the matrix $V \in \mathbb{R}^{n\times k}$ form a basis for the subspace S where $k = \dim S$. Then the problem $\min\{f(x)\ |x \in \hat{x} + S\}$ is equivalent to the problem $\min_z f(\hat{x} + Vz)$. By part (b), we have $\nabla f(\bar{x}) \perp \mathrm{Ran}\,(V) = S$.