Problem Set #6

Reiko Laski

Exercise 9.1
An unconstrained linear objective function is either constant or has no minimum.
Proof:
Consider the unconstrained linear objective function $f(\mathbf{x}) = \mathbf{b}^T\mathbf{x} + c$. By the FONC, we know that if a minimum exists, it will occur when $Df(\mathbf{x}) = \mathbf{0}$. If $f(\mathbf{x})$ is a constant function, then $Df(\mathbf{x}) = \mathbf{0}$ and we have a minimum. If $f(\mathbf{x})$ is not a constant function, then $Df(\mathbf{x}) = \mathbf{b}^T$ and there is no minimum.

Exercise 9.2
If $\mathbf{b} \in \mathbb{R}^m$ and $A \in M_{m\times n}(\mathbb{R})$, then the problem of finding an $\mathbf{x}^* \in \mathbb{R}^n$ to minimize $\|A\mathbf{x} - \mathbf{b}\|_2$ is equivalent to minimizing

$$\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{b}^T A\mathbf{x}.$$

Proof:

$$\begin{aligned}
\|A\mathbf{x} - \mathbf{b}\| &= (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}) \\
&= (\mathbf{x}^T A^T - \mathbf{b}^T)(A\mathbf{x} - \mathbf{b}) \\
&= \mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} - \mathbf{x}^T A^T\mathbf{b} + \mathbf{b}^T\mathbf{b} \\
&= \mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{b}^T A\mathbf{x} + \mathbf{b}^T\mathbf{b}
\end{aligned}$$

The FOC of this system is equivalent to that of $\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{b}^T A\mathbf{x}$,

$$\begin{aligned}
2A^T A\mathbf{x} - 2A^T\mathbf{b} &= \mathbf{0} \\
\implies A^T A\mathbf{x} &= A^T\mathbf{b}
\end{aligned}$$

Since $A^T A$ is positive definite, the solution to the normal equation is the unique minimizer of $\|A\mathbf{x} - \mathbf{b}\|_2$.

Exercise 9.3
Gradient descent, Newton, Quasi-Newton, Conjugate gradient
For each of the multivariate optimization methods, list:
(i)
(ii)
(iii)
(iv)
Proof:

Exercise 9.4
Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x}$, where $Q \in M_n(\mathbb{R})$ satisfies $Q > 0$ and $\mathbf{b} \in \mathbb{R}^n$. The Method of Steepest Descent (that is, gradient descent with optimal line search), converges in one step (that is, $\mathbf{x}_1 = Q^{-1}\mathbf{b}$), if and only if $\mathbf{x}_0$ is chosen such that $Df(\mathbf{x}_0)^T = Q\mathbf{x}_0 - \mathbf{b}$

is an eigenvector of Q $\left(\text{and } \alpha_0 = \frac{Df(\mathbf{x}_0)Df(\mathbf{x}_0)^T}{Df(\mathbf{x}_0)QDf(\mathbf{x}_0)^T}\right)$.
Proof:
First, suppose that $\mathbf{x}_0$ is chosen such that $Df(\mathbf{x}_0)^T = Q\mathbf{x}_0 - \mathbf{b}$ is an eigenvector of Q. Then we have that $Q(Q\mathbf{x}_0 - \mathbf{b}) = \lambda(Q\mathbf{x}_0 - \mathbf{b})$ for some $\lambda \in \mathbb{R}$. We can then evaluate $\mathbf{x}_1$ as

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{x}_0 - \alpha_0 Df(\mathbf{x}_0)^T \\
&= \mathbf{x}_0 - \frac{Df(\mathbf{x}_0)Df(\mathbf{x}_0)^T}{Df(\mathbf{x}_0)QDf(\mathbf{x}_0)^T} Df(\mathbf{x}_0)^T \\
&= \mathbf{x}_0 - \frac{(Q\mathbf{x}_0 - \mathbf{b})^T(Q\mathbf{x}_0 - \mathbf{b})}{(Q\mathbf{x}_0 - \mathbf{b})^T Q(Q\mathbf{x}_0 - \mathbf{b})}(Q\mathbf{x}_0 - \mathbf{b}) \\
&= \mathbf{x}_0 - \frac{(Q\mathbf{x}_0 - \mathbf{b})^T(Q\mathbf{x}_0 - \mathbf{b})}{(Q\mathbf{x}_0 - \mathbf{b})^T \lambda(Q\mathbf{x}_0 - \mathbf{b})}(Q\mathbf{x}_0 - \mathbf{b}) \\
&= \mathbf{x}_0 - \frac{1}{\lambda}(Q\mathbf{x}_0 - \mathbf{b}) \\
&= \mathbf{x}_0 - Q^{-1}(Q\mathbf{x}_0 - \mathbf{b}) \\
&= \mathbf{x}_0 - Q^{-1}Q\mathbf{x}_0 - Q^{-1}\mathbf{b} \\
&= Q^{-1}\mathbf{b}
\end{aligned}$$

Now suppose that $\mathbf{x}_1 = Q^{-1}\mathbf{b}$ (the Method of Steepest Descent converges in one step). MUST PROVE REVERSE

Exercise 9.5
Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1. Let $\{\mathbf{x}_k\}_{k=0}^{\infty}$ be defined by the Method of Steepest Descent. Then $\mathbf{x}_{k+1} - \mathbf{x}_k$ is orthogonal to $\mathbf{x}_{k+2} - \mathbf{x}_{k+1}$ for each k.
Proof:

Exercise 9.6
See Jupyter Notebook

Exercise 9.7
See Jupyter Notebook

Exercise 9.8
See Jupyter Notebook

Exercise 9.9
See Jupyter Notebook

Exercise 9.10
Consider the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x}$, where $Q \in M_n(\mathbb{R})$ is symmetric and positive definite and $\mathbf{b} \in \mathbb{R}^n$. For any initial guess $\mathbf{x}_0 \in \mathbb{R}^n$, one iteration of Newton's method lands at the unique minimizer of f.
Proof:

Since Q is positive definite, we know that there is a unique minimizer of $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x}$. By the FONC, $Q\mathbf{x}^* - \mathbf{b} = \mathbf{0} \implies \mathbf{x}^* = Q^{-1}\mathbf{b}$. Using Newton's method with an arbitrary $\mathbf{x}_0$, we have

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{x}_0 - D^2 f(\mathbf{x}_0)^{-1} Df(\mathbf{x}_0)^T \\
&= \mathbf{x}_0 - Q^{-1}(Q\mathbf{x}_0 - \mathbf{b}) \\
&= \mathbf{x}_0 - Q^{-1}Q\mathbf{x}_0 + Q^{-1}\mathbf{b} \\
&= Q^{-1}\mathbf{b} \\
&= \mathbf{x}^*
\end{aligned}$$

Exercise 9.12
If $A \in M_n(\mathbb{F})$ has eigenvalues $\lambda_1, ..., \lambda_n$ and $B = A + \mu I$, then the eigenvectors of A and B are the same, and the eigenvalues of B are $\mu + \lambda_1, \mu + \lambda_2, ..., \mu + \lambda_n$.
Proof:
Let $\mathbf{x}_i$ be the eigenvector of A corresponding to the eigenvalue λ_i. Then we have that

$$\begin{aligned}
B\mathbf{x}_i &= (A + \mu I)\mathbf{x}_i \\
&= A\mathbf{x}_i + \mu I\mathbf{x}_i \\
&= \lambda_i\mathbf{x}_i + \mu\mathbf{x}_i \\
&= (\lambda_i + \mu)\mathbf{x}_i
\end{aligned}$$

Exercise 9.15
Let A be a nonsingular $n \times n$ matrix, B an $n \times \ell$ matrix, C a nonsingular $\ell \times \ell$ matrix, and D an $\ell \times n$ matrix. We have

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Proof:
The following is Matt's code:

$$\begin{aligned}
&(A + BCD)(A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}) \\
&= AA^{-1} - AA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= I - B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= I - B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= I + BCDA^{-1} - (B(C^{-1} + DA^{-1}B)^{-1} + BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1})DA^{-1} \\
&= I + BCDA^{-1} - ((B + BCDA^{-1}B)(C^{-1} + DA^{-1}B)^{-}1))DA^{-1} \\
&= I + BCDA^{-1} - (BC(C^{-1} + DA^{-1}B)(C^{-1} + DA^{-1}B)^{-}1))DA^{-1} \\
&= I + BCDA^{-1} - BCDA^{-1} \\
&= I
\end{aligned}$$

Exercise 9.16
Let $Q \in M_n(\mathbb{R})$ satisfy $Q > 0$, and let f be the quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x} + c$. Given a starting point $\mathbf{x}_0$ and Q-conjugate directions $\mathbf{d}_0, \mathbf{d}_1, ..., \mathbf{d}_{n-1}$ in $\mathbb{R}^n$,

the optimal line search solution for $x_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ (that is, the α which minimizes $\phi_k(\alpha) = f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)$) is given by $\alpha_k = \frac{\mathbf{r}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}$, where $\mathbf{r}_k = \mathbf{b} - Q\mathbf{x}_k$. *Proof:*

Exercise 9.18
Proof:

Exercise 9.20
In the Conjugate Gradient Algorithm $\mathbf{r}_i^T \mathbf{r}_k = 0$ for all $i < k$.
Proof: