Problem Set 6

By: Bethany Bailey

9.1

Let f be an unconstrained linear objective function $f(\vec{x}) = \vec{b}^T\vec{x} + c$. By the first order condition, we know that if a minimum exists, it will occur when $Df(\vec{x}) = \vec{0}$. When $f(\vec{x})$ is a constant function, $Df(\vec{x}) = \vec{0}$ and we have a minimum. If $f(\vec{x})$ is not a constant function, then $Df(\vec{x}) = \vec{b}^T$ and there is no minimum. Thus, an unconstrained linear objective function is either constant or has no minimum.

9.2

As we know from previous proofs, A^TA is symmetric and positive definite. Thus, if $b \in \mathbb{R}^{>}$, to minimize $||Ax - b||_2$ we just need to find an x* that satisfies the first order condition and then it will be a global minimizer of $x^TA^TAx - 2b^TAx$. The first order condition is $2A^TAx - 2A^Tb = 0$. Thus, $A^TAx = A^Tb$, and we can see that solivng the normal equation is also the same as minimizing $||Ax - b||_2$.

9.3

(i)

Gradient: pick initial point and move in steepest direction

Newton's: quadratic approximation and iterative solution

BFGS: approximate Hessian

Conjugate Gradient: move along Q-conjugate directions

(ii)

Gradient: differentiable objective function

Newton's: non-large dimension quadratic

BFGS: functions that are not smooth

Conjugate Gradient: large dimension, quadratic with sparse matrices

(iii)

Gradient: iterations are less expensive

Newton's: quick convergence

BFGS: iterations are less expensive than Newton's

Conjugate Gradient: can solve large problems inexpensively

(iv)

Gradient: slow convergence

Newton's: iterations are expensive

BFGS: storage is expensive

Conjugate Gradient: need sparse matrices

9.4

Let D be the derivative of $f(x)^T$, and λ the eigenvalue of Q with $Df(x_0)^T = Qx_0 - b$. Then,

$$\begin{aligned}
x_1 =& x_0 - \frac{DD^T}{DQD^T} D^T \\
=& x_0 - \frac{DD^T}{D\lambda D^T} D^T \\
=& x_0 - \frac{1}{\lambda} D^T \\
=& x_0 - Q^{-1} D^T \\
=& x_0 - Q^{-1}(Qx_0 - b) \\
=& Q^{-1} b
\end{aligned}$$

9.5

9.6

See Jupyter Notebook.

9.7

See Jupyter Notebook.

9.8

See Jupyter Notebook.

9.9

See Jupyter Notebook.

9.10

See the following Newton iteration, which shows that the algorithm converges on the unique minimizer in one iteration.

$$
\begin{aligned}
x1 =& x0 - (D^2(f(x_0)))^{-1} Df(x_0)^T \\
=& x0 - Q^{-1}(Qx_0 - b) \\
=& x0 - x_0 + Q^{-1}b \\
=& Q^{-1}b
\end{aligned}
$$

9.12

Let $\vec{x}_i$ be the eigenvector of A corresponding to the eigenvalue λ_i. Then:

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

Therefore, the eigenvectors of A and B are the same, and the eigenvalues of B are as shown in the question.

9.15

$$
\begin{aligned}
&(A + BCD)\left[A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}\right] \\
&= \left\{I - B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}\right\} + \left\{BCDA^{-1} - BCDA^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}\right\} \\
&= \left\{I + BCDA^{-1}\right\} - \left\{B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1} + BCDA^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}\right\} \\
&= I + BCDA^{-1} - \left(B + BCDA^{-1}B\right)\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1} \\
&= I + BCDA^{-1} - BC\left(C^{-1} + DA^{-1}B\right)\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1} \\
&= I + BCDA^{-1} - BCDA^{-1} \\
&= I.
\end{aligned}
$$

9.16

Substituting $x = \frac{v - As}{\|s\|^2}, y = s$ into the Sherman-Morrison-Woodbury formula, we get:

$$
\begin{aligned}
A^{-1} - \frac{A^{-1}xy^T A^{-1}}{1 + y^T A^{-1} x} =& A^{-1} - \frac{A^{-1}(v - As)/\|s\|^2 s^T A^{-1}}{1 + s^T A^{-1}(v - As)/\|s\|^2} \\
=& A^{-1} - \frac{(A^{-1}v - s)s^T A^{-1}/\|s\|^2}{1 + s^T A^{-1} v/\|s\|^2 - 1} \\
=& A^{-1} + \frac{(s - A^{-1}v)s^T A^{-1}}{s^T A^{-1} v}
\end{aligned}
$$