Sergey Voronin
Department of Mathematics
Tufts University

December 7, 2016

HPC — Homework 1

(A) A matrix $A \in \mathbb{R}^{n\times n}$ is said to be *symmetric and positive definite* if $A^T = A$ and $x^T A x > 0$ for any nonzero vector $x \in \mathbb{R}^n$.

If A is SPD, then all eigenvalues of A are positive. Suppose v is an eigenvector.

$$v^T A v = v^T \lambda v = v^T v \lambda = \|v\|^2 \lambda > 0 \implies \lambda > 0$$

(B) First note that since $Ax = b$,

$$J(x) = \frac{1}{2} x^T A x - b^T x = \frac{1}{2} x^T A x - x^T A x = -\frac{1}{2} x^T A x. \tag{1}$$

For any vector $y \in \mathbb{R}^n$,

$$\begin{aligned}
J(y) &= \frac{1}{2} y^T A y - b^T y \\
&= \frac{1}{2} y^T A y - x^T A y \\
&= \left(\frac{1}{2} y^T A y - x^T A y + \frac{1}{2} x^T A x\right) - \frac{1}{2} x^T A x \\
\text{(by (1))} \qquad &= \left(\frac{1}{2} y^T A y - x^T A y + \frac{1}{2} x^T A x\right) + J(x) \\
\text{(by symmetry of } A\text{)} \qquad &= \frac{1}{2}(x-y)^T A (x-y) + J(x) \\
&\geq J(x),
\end{aligned}$$

where the last inequality is due to the positive definiteness of A, and equality holds if and only if $y = x$. In other words, $J(y) > J(x)$ for all $y \neq x$. Hence x is the unique global minimizer of J.

(C) Note that for any fixed vectors x and d, $\alpha^* := \frac{(b-Ax)^T d}{d^T A d}$ is the unique critical point of the function $f(\alpha) := J(x + \alpha d)$. First, let's see what exactly $f(\alpha)$ is:

$$\begin{aligned} f(\alpha) =& \frac{1}{2}(x+\alpha d)^T A(x+\alpha d) - b^T(x+\alpha d) \\ =& \frac{1}{2}(x^T Ax + 2\alpha x^T Ad + \alpha^2 d^T Ad) - b^T x - \alpha b^T d \\ =& \left(\frac{1}{2}x^T Ax - b^T x\right) + \alpha\left(x^T Ad - b^T d\right) + \frac{1}{2}\alpha^2 d^T Ad. \end{aligned}$$

From this, we see that

$$0 = f'(\alpha) = x^T Ad - b^T d + \alpha d^T Ad$$

if and only if

$$\alpha = \frac{b^T d - x^T Ad}{d^T Ad} = \frac{(b - Ax)^T d}{d^T Ad}.$$

This shows that $\alpha^* = \frac{(b-Ax)^T d}{d^T Ad}$ is the unique critical point of f.

Then we show that α^* is a minimizer via the second derivative test: indeed,

$$f''(\alpha) = d^T Ad > 0$$

for *any* α, since A is SPD and $d \neq 0$.

In conclusion, by definition of r^n, we have that $\arg\min_\alpha J(x^n + \alpha d^n) = \frac{r^n \cdot d^n}{d^n \cdot Ad^n}$.

(D) Recall (from Calculus 3) that the *gradient* of a function $g : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$, if exists, is the unique vector v that satisfies

$$g(x + d) = g(x) + v^T d + o(\|d\|), \quad \forall\, d \in \mathbb{R}^n.$$

To compute $\nabla J(x)$, fix any vector $d \in \mathbb{R}^n$, and compute the first order approximation of $J(x + d)$ as follows:

$$\begin{aligned} J(x+d) =& \frac{1}{2}(x+d)^T A(x+d) - b^T(x+d) \\ =& \frac{1}{2}\left(x^T Ax + 2x^T Ad + d^T Ad\right) - b^T x - b^T d \\ =& \frac{1}{2}x^T Ax - b^T x + (A^T x - b)^T d + \frac{1}{2}d^T Ad \\ =& J(x) + (Ax - b)^T d + o(\|d\|). \end{aligned}$$

Here we have that $d^T Ad = o(\|d\|)$ because

$$0 \leq \frac{d^T Ad}{\|d\|} \leq \|d\| \left\| A\frac{d}{\|d\|} \right\| \leq \|d\| \cdot (\textsf{maximum eigenvalue of } A) \to 0 \qquad \text{as } \|d\| \searrow 0.$$

Hence $\nabla J(x) = Ax - b = -r(x)$.

Submitted by Sergey Voronin on December 7, 2016.