Notes for Wednesday, March 25

Nocedal & Wright p. 103 claim

if $x_{k+1} = x_k + \alpha p_k$

the $\alpha \in \mathbb{R}$ that minizes $\varphi(x_k + \alpha p_k)$ over $\alpha \in \mathbb{R}$ is

$$\alpha_k = \frac{-r_k^T p_k}{p_k^T A p_k}$$

To see this, recall

$$\varphi(x) = \tfrac{1}{2} x^T A x - b^T x \qquad (A \text{ SPD})$$

$$\therefore \quad \varphi(x_k + \alpha p_k) = \tfrac{1}{2}(x_k + \alpha p_k)^T A (x_k + \alpha p_k) - b^T(x_k + \alpha p_k)$$

$$= \tfrac{1}{2} x_k^T A x_k + \alpha x_k^T A p_k + \tfrac{1}{2}\alpha^2 p_k^T A p_k$$

$$- b^T x_k - \alpha b^T p_k$$

$$\therefore \quad \frac{d}{d\alpha} \varphi(x_k + \alpha p_k) = x_k^T A p_k + \alpha p_k^T A p_k - b^T p_k$$

$$= (Ax_k - b)^T b_k + \alpha p_k^T A p_k$$

$$= r_k^T b + \alpha p_k^T A p_k$$

$$\therefore \quad \frac{d}{d\alpha} \varphi(x_k + \alpha p_k) = 0$$

$$\Longleftrightarrow \quad \alpha = \frac{-r_k^T b}{p_k^T A p_k}$$

Proof of Theorem 5.1

also $p_i \neq 0 \; \forall i$

N + W say the conjugate directions $\{p_0, p_1, \dots, p_{n-1}\}$

satisfying $\quad p_i^T A p_j = 0 \quad$ for $i \neq j$

are linearly indendent.

<u>Why?</u>

To show that $\{p_0, p_1, \dots, p_{n-1}\}$ are linearly ind. in R^n

need to show that if $\displaystyle\sum_{j=0}^{n-1} \alpha_j p_j = 0$

then all $\alpha_i = 0$.

So suppose $\displaystyle\sum_{j=0}^{n-1} \alpha_j p_j = 0$.

Then
$$
\begin{aligned}
0 &= p_i^T A \left(\sum_{j=0}^{n-1} \alpha_j p_j\right) \\
&= \sum_{j=0}^{n-1} \alpha_j \, p_i^T A p_j \\
&= \alpha_i \, p_i^T A p_i \qquad (\text{because } p_i^T A p_j = 0 \text{ for } i \neq j)
\end{aligned}
$$

Since $p_i^T A p_i > 0$ (because A is SPD and $p_i \neq 0$)

must have $\alpha_i = 0$. This argument holds for all $i = 0, 1, \dots, n-1$.

So $\alpha_i = 0$ for all $i = 0, 1, 2, \dots, n-1$.

Bottom of p. 105

N & W claim

$$S^T A S \text{ is diagonal}$$

where $S = [p_0, p_1, \cdots, p_{n-1}]$

and $p_i^T A p_j = 0$ for $i \neq j$.

Why?

$$S^T A S = \begin{bmatrix} p_0^T \\ p_1^T \\ \vdots \\ p_{n-1}^T \end{bmatrix} A \, [p_0, p_1, \cdots p_{n-1}]$$

$$= \begin{bmatrix} p_0^T \\ p_1^T \\ \vdots \\ p_{n-1}^T \end{bmatrix} [A p_0, A p_1, \cdots A p_{n-1}]$$

So the (i,j) element of $S^T A S$ is

$$p_i^T (A p_j) = p_i^T A p_j = 0 \quad \text{if } i \neq j.$$

If A is diagonal, $A = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ & & \ddots & & \\ 0 & & \cdots & 0 & d_n \end{pmatrix}$

then the co-ordinate directions $e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i$

are conjugate.

$$e_i^T A e_j = e_i^T (A e_j) = \begin{array}{c} (0 \cdots 0\,1\,0 \cdots 0) \\ \uparrow \\ i \end{array} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ d_j \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j$$

$$= 0$$

(5.10) $\qquad r_{k+1} = r_k + \alpha_k A p_k$

<u>why</u>?

Recall $\quad r_k = A x_k - b \quad$ and $\quad x_{k+1} = x_k + \alpha_k p_k$

$$\therefore \quad r_{k+1} = A x_{k+1} - b$$

$$= A(x_k + \alpha_k p_k) - b$$

$$= A x_k + \alpha_k A p_k - b$$

$$= A x_k - b + \alpha_k A p_k$$

$$= r_k + \alpha_k A p_k$$

N + W p. 106

$$h(\sigma) = \phi(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1})$$

is a strictly convex quadratic (in the $\sigma_0, \cdots, \sigma_{k-1}$)

First note that $h(\sigma)$ is a quadratic in $\sigma_0, \cdots, \sigma_{k-1}$.

$$\begin{aligned}
h(\sigma) &= \phi(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&= \tfrac{1}{2}(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1})^T A (x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&\quad - b^T (x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&= \tfrac{1}{2} x_0^T A x_0 + x_0^T A (\sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&\quad + \tfrac{1}{2}(\sigma_0 p_0 + \cdots \sigma_{k-1} p_{k-1})^T A (\sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&\quad - b^T x_0 - b^T (\sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}) \\
&= \tfrac{1}{2} x_0^T A x_0 + \sigma_0 x_0^T A p_0 + \cdots + \sigma_{k-1} x_0^T A p_{k-1} \\
&\quad + \tfrac{1}{2} \sigma_0^2 p_0^T A p_0 + \cdots + \tfrac{1}{2} \sigma_{k-1}^2 p_{k-1}^T A p_{k-1} \\
&\quad - b^T x_0 - \sigma_0 b^T p_0 - \sigma_{k-1} b^T p_{k-1}
\end{aligned}$$

(Note all the terms $\frac{1}{2} \sigma_i \sigma_j p_i^T A p_j = 0$ for $i \neq j$ because $\{p_i\}$ conjugate

$\therefore$ $h(\sigma)$ is a quadratic in the σ_i

To see the $h(\sigma)$ is strictly convex,

consider its Hessian matrix $H(\sigma) = \left(\frac{\partial^2 h(\sigma)}{\partial \sigma_i \partial \sigma_j} \right)$

$$\frac{\partial h(\sigma)}{\partial \sigma_j} = x_0^T A p_j + \sigma_j p_j^T A p_j - b^T p_j$$

$$\therefore \quad \frac{\partial^2 h(\sigma)}{\partial x_i \partial x_j} = \begin{cases} p_i^T A p_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note also $p_i^T A p_i > 0$ since A SPD and $p_i \neq 0$.

$\therefore$ $H(\sigma)$ is a diagonal matrix with positive diagonal elements

$\Rightarrow$ $H(\sigma)$ is symmetric positive-definite

$\Rightarrow$ $h(\sigma)$ is strictly convex.

Proof of Theorem 5.2 N&W p. 106

$h(\sigma)$ strictly convex so has a unique minimizer σ^*

that satisfies

$$\frac{\partial h(\sigma^*)}{\partial \sigma_i} = 0 \qquad i = 0, 1, \cdots, k-1 \qquad (1)$$

$$h(\sigma) = \varphi(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1})$$

So $$\frac{\partial h(\sigma)}{\partial \sigma_i} = \nabla \varphi(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1})^T p_i$$

Therefore (1) is equivalent to

$$\nabla \varphi(x_0 + \sigma_0^* p_0 + \cdots + \sigma_{k-1}^* p_{k-1}) p_i = 0 \qquad \text{for } i = 0, 1, \cdots, k-1 \qquad (2)$$

However, recall for for $\varphi(x) = \frac{1}{2} x^T A x - b^T x$

$$\nabla \varphi(x) = Ax - b = r(x)$$

$\therefore$ (2) is equivalent to

$$r(\tilde{x})^T p_i = 0 \qquad \text{for } i = 0, 1, \cdots, k-1$$

where $\tilde{x} = x_0 + \sigma_0^* p_0 + \cdots + \sigma_{k-1}^* p_{k-1}$

Bottom of p. 106 of N+W. Claim

$$r_1^T p_0 = 0$$

Note $r_1 = A x_1 - b$ and $x_1 = x_0 + \alpha_0 p_0$

$$\alpha_0 = - \frac{r_0^T p_0}{p_0^T A p_0}$$ (see (5.7) in N+W)

$$\therefore \quad r_1^T p_0 = (A x_1 - b)^T p_0$$

$$= (A(x_0 + \alpha_0 p_0) - b)^T p_0$$

$$= (A x_0 - b + \alpha_0 A p_0)^T p_0$$

$$= (r_0 + \alpha_0 A p_0)^T p_0$$

$$= r_0^T p_0 + \alpha_0 p_0^T A p_0$$

$$= r_0^T p_0 - \frac{r_0^T p_0}{p_0^T A p_0} p_0^T A p_0$$

$$= 0$$

N & W p. 107

Eigenvectors satisfy conjugacy property

A SPD $\Rightarrow$ $\exists$ an orthogonal matrix Q s.t.
(i.e. $Q Q^T = Q^T Q = I$)

$$Q^T A Q = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \qquad (1)$$

Let $Q = [q_1, q_2 \cdots q_n]$

(1) is equivalent to

$$A Q = Q \Lambda$$

j^{th} column of AQ is $A q_j$ and

j^{th} column of $Q\Lambda$ is $\lambda_j q_j$

So (1) $\Rightarrow$ $A q_j = \lambda_j q_j$

$\Rightarrow$ the matrix Q consists of eigenvectors of A that are normalizes so that

$$q_i^T q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

(This comes from $Q^T Q = I$

$$\Rightarrow \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} [q_1\ q_2 \cdots q_n] = I$$

$\Rightarrow$ (i,j) element of $Q^T Q$ is $q_i^T q_j$

and (i,j) element of I is $\begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$

For these normalized eigenvectors in the Q matrix

$$q_i^T A q_j = q_i^T (A q_j) = q_i^T (\lambda_j q_j) = \lambda_j q_i^T q_j$$
$$= \begin{cases} \lambda_j & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$

$\therefore$ The set of eigenvectors from Q

$$\{ q_1, q_2, \cdots q_n \}$$

is a set of conjugate vectors for A

Note not all eigenvectors are conjugate.

E.g. supose $A = I$

Then $q_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $q_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are both

eigenvectors of I since $I q_1 = q_1$ & $I q_2 = q_2$

But $q_1^T I q_2 = q_1^T q_2 = 1$ $\Rightarrow$ not conjugate.

This can only occur when A has eigenvalues $\lambda_i = \lambda_j$ for some $i \neq j$.