Homework 9 Foundations of Computational Math 1 Fall 2012

Problem 9.1

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix, $C \in \mathbb{R}^{n\times n}$ be a symmetric nonsingular matrix, and $b \in \mathbb{R}^n$ be a vector. The matrix $M = C^2$ is therefore symmetric positive definite. Also, let $\tilde{A} = C^{-1}AC^{-1}$ and $\tilde{b} = C^{-1}b$.

The preconditioned Steepest Descent algorithm to solve $Ax = b$ is:

A, M are symmetric positive definite
x_0 arbitrary; $r_0 = b - Ax_0$; solve $Mz_0 = r_0$

do $k = 0, 1, \ldots$ until convergence

$$
\begin{aligned}
w_k &= Az_k \\
\alpha_k &= \frac{z_k^T r_k}{r_k^T w_k} \\
x_{k+1} &\leftarrow x_k + z_k\alpha_k \\
r_{k+1} &\leftarrow r_k - w_k\alpha_k \\
\text{solve } Mz_{k+1} &= r_{k+1}
\end{aligned}
$$

end

The Steepest Descent algorithm to solve $\tilde{A}\tilde{x} = \tilde{b}$ is:

$\tilde{A}$ is symmetric positive definite
$\tilde{x}_0$ arbitrary; $\tilde{r}_0 = \tilde{b} - \tilde{A}\tilde{x}_0$; $\tilde{v}_0 = \tilde{A}\tilde{r}_0$

do $k = 0, 1, \ldots$ until convergence

$$
\begin{aligned}
\tilde{\alpha}_k &= \frac{\tilde{r}_k^T \tilde{r}_k}{\tilde{r}_k^T \tilde{v}_k} \\
\tilde{x}_{k+1} &\leftarrow \tilde{x}_k + \tilde{r}_k\tilde{\alpha}_k \\
\tilde{r}_{k+1} &\leftarrow \tilde{r}_k - \tilde{v}_k\tilde{\alpha}_k \\
\tilde{v}_{k+1} &\leftarrow \tilde{A}\tilde{r}_{k+1}
\end{aligned}
$$

end

Show that given the appropriate consistency between initial guesses the preconditioned steepest descent recurrences to solve $Ax = b$ can be derived from the steepest descent recurrences to solve $\tilde{A}\tilde{x} = \tilde{b}$.

Solution:

This is easily shown via induction.

We have

$$M = C^2, \quad \tilde{A} = C^{-1}AC^{-1}, \quad \tilde{b} = C^{-1}b$$

$$\tilde{A}\tilde{x} = \tilde{b} \rightarrow \tilde{x} = Cx$$

Let $k = 0$ and assume that $\tilde{x}_0 = Cx_0$ (the consistency required). We have:

$$\tilde{r}_0 = \tilde{b} - \tilde{A}\tilde{x}_0 = C^{-1}r_0$$

$$\begin{aligned}\tilde{v}_0 = \tilde{A}\tilde{r}_0 &= C^{-1}AC^{-1}\tilde{r}_0 = C^{-1}AC^{-1}C^{-1}r_0 \\ &= C^{-1}AM^{-1}r_0 = C^{-1}Az_0\end{aligned}$$

where we have defined $M^{-1}r_0 = z_0$ as required by preconditioned steepest descent.

Since M and C are symmetric, we also have

$$\begin{aligned}\tilde{\alpha}_0 = \frac{\tilde{r}_0^T\tilde{r}_0}{\tilde{r}_0^T\tilde{v}_0} &= \frac{r_0^TC^{-T}C^{-1}r_0}{r_0^TC^{-T}C^{-1}Az_0} \\ &= \frac{r_0^TC^{-T}C^{-T}r_0}{r_0^TC^{-T}C^{-T}Az_0} \\ &= \frac{r_0^TM^{-T}r_0}{r_0^TM^{-T}Az_0} \\ &= \frac{z_0^Tr_0}{z_0^TAz_0} = \alpha_0\end{aligned}$$

where α_0 is as defined in preconditioned steepest descent.

Note it follows that

$$\begin{aligned}\tilde{x}_1 &\leftarrow \tilde{x}_0 + \tilde{r}_0\tilde{\alpha}_0 \\ Cx_1 &\leftarrow Cx_0 + C^{-1}r_0\alpha_0 \\ x_1 &\leftarrow x_0 + M^{-1}r_0\alpha_0 \\ \therefore \quad x_1 &\leftarrow x_0 + z_0\alpha_0\end{aligned}$$

and the first two iterates are related via the change of variables C as desired.

Now assume that

$$\tilde{r}_k = \tilde{b} - \tilde{A}\tilde{x}_k = C^{-1}r_k$$

$$\begin{aligned}\tilde{v}_k &= C^{-1}Az_k \\ \tilde{\alpha}_k &= \alpha_k \\ \tilde{x}_k &= Cx_k\end{aligned}$$

where $z_k = M^{-1}r_k$. We must show that this induction assumption implies the same relationships for $k+1$ and verify that the x_{k+1} and $\tilde{x}_{k+1}$ have the desired relationship.

We have

$$\begin{aligned}\tilde{r}_{k+1} &= \tilde{r}_k - \tilde{v}_k\tilde{\alpha}_k \\ &= C^{-1}r_k - C^{-1}Az_k\alpha_k \\ C\tilde{r}_{k+1} &= r_k - Az_k\alpha_k = r_{k+1}\end{aligned}$$

as desired.

We also have

$$\begin{aligned}\tilde{v}_{k+1} &= \tilde{A}\tilde{r}_{k+1} = C^{-1}AC^{-1}C^{-1}r_{k+1} \\ &= C^{-1}AM^{-1}r_{k+1} = C^{-1}Az_{k+1}\end{aligned}$$

$$\begin{aligned}\tilde{\alpha}_{k+1} &= \frac{\tilde{r}_{k+1}^T\tilde{r}_{k+1}}{\tilde{r}_{k+1}^T\tilde{v}_{k+1}} \\ &= \frac{r_{k+1}^T M^{-T} r_{k+1}}{r_{k+1}^T M^{-T} Az_{k+1}} \\ &= \frac{z_{k+1}^T r_{k+1}}{z_{k+1}^T Az_{k+1}} = \alpha_{k+1}\end{aligned}$$

as desired.

Finally, we can then verify the relationship between the iterates:

$$\begin{aligned}\tilde{x}_{k+1} &= \tilde{x}_k + \tilde{r}_k\tilde{\alpha}_k \\ &= Cx_k + C^{-1}r_k\alpha_k \\ C^{-1}\tilde{x}_{k+1} &= x_k + M^{-1}r_k\alpha_k = x_k + z_k\alpha_k = x_{k+1}\end{aligned}$$

$$\therefore \quad \tilde{x}_{k+1} = Cx_{k+1}$$

Problem 9.2

Consider the generic Conjugate Direction algorithm for solving the minimization problem

$$\min_{x\in\mathbb{R}^n} f(x)$$

where $f(x) = x^TAx - x^Tb$, $b \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite.

Denote the A-orthogonal direction vectors $d_0, d_1, \ldots$ and let $r_k = b - Ax_k$. Show that

$$\frac{d_{k-1}^T r_0}{d_{k-1}^T Ad_{k-1}} = \frac{d_{k-1}^T r_{k-1}}{d_{k-1}^T Ad_{k-1}}$$

Solution:

For the generic CD algorithm we have

$$x_{k-1} = x_0 + \alpha_0 d_0 + \cdots + \alpha_{k-2} d_{k-2}$$

$$\begin{aligned} r_{k-1} &= b - Ax_{k-1} = b - A\left(x_0 + \alpha_0 d_0 + \cdots + \alpha_{k-2} d_{k-2}\right) \\ &= r_0 - \left(\alpha_0 A d_0 + \cdots + \alpha_{k-2} A d_{k-2}\right) \end{aligned}$$

$$d_{k-1}^T r_{k-1} = d_{k-1}^T r_0 - \left(\alpha_0 d_{k-1}^T A d_0 + \cdots + \alpha_{k-2} d_{k-1}^T A d_{k-2}\right) = d_{k-1}^T r_0$$

by the $A-$orthogonality of the d_i vectors.

Problem 9.3

When solving $Ax = b$ or equivalently the associated quadratic definite minimization problem using CG, we have

$$x_{k+1} = x_0 + \alpha_0 p_0 + \cdots + \alpha_k p_k$$

where the p_j are $A-$orthogonal vectors. It can be shown that

$$span[p_0, \ldots, p_k] = span[r_0,\ Ar_0, \ldots, A^k r_0]$$

where $r_0 = b - Ax_0$ and x_0 is the initial guess at the solution $x^* = A^{-1}b$. Therefore,

$$x_{k+1} = x_0 + \gamma_0 r_0 + \gamma_1 A r_0 + \cdots + \gamma_k A^k r_0 = x_0 - P_k(A) r_0$$

where $P_k(A) = \gamma_0 I + \gamma_1 A + \cdots + \gamma_k A^k$ is a matrix that is called a matrix polynomial of evaluated at A. (A space whose span can be defined by a matrix polynomial is called a Krylov space.)

Denote $d_j = A^j r_0$ for $j = 0,\ 1, \ldots$ and determine the relationship between the coefficients $\alpha_0, \ldots, \alpha_k$ and the coefficients $\gamma_0, \ldots, \gamma_k$.

Solution: We have

$$x_{k+1} = x_0 + \alpha_0 p_0 + \cdots + \alpha_k p_k = x_0 + \gamma_0 r_0 + \gamma_1 A r_0 + \cdots + \gamma_k A^k r_0 = x_0 + \gamma_0 d_0 + \gamma_1 d_1 + \cdots + \gamma_k d_k$$

where $d_i = A^i r_0$. Define the $n \times k+1$ matrices and vectors

$$D_k = \begin{pmatrix} d_0 & d_1 & \ldots & d_k \end{pmatrix} \quad \text{and} \quad P_k = \begin{pmatrix} p_0 & p_1 & \ldots & p_k \end{pmatrix}$$

$$a = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_k \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_k \end{pmatrix}$$

and note

$$x_{k+1} = x_0 + P_k a = x_0 + D_k c = x_0 + v.$$

Since the d_i and p_i are two different bases for the same space the p_i have a unique linear combination in terms of the d_i, i.e.,

$$d_i = \mu_{0i} p_0 + \cdots + \mu_{ki} p_k$$

which can be written in matrix form

$$P_k M = D_k$$

where M is an $k+1 \times k+1$ nonsingular matrix

$$M = \begin{pmatrix} \mu_{00} & \cdots & \mu_{0k} \\ \vdots & \ddots & \vdots \\ \mu_{k0} & \cdots & \mu_{kk} \end{pmatrix}.$$

We can solve for M

$$\begin{aligned} P_k M &= D_k \\ P_k^T A P_k M &= P_k^T A D_k \\ S_k M &= P_k^T A D_k \\ M &= S_k^{-1} P_k^T A D_k \end{aligned}$$

where S_k is a diagonal nonsingular matrix with main diagonal elements $p_j^T A p_j$, $0 \le j \le k$. Therefore,

$$\mu_{ij} = \frac{p_i^T A d_j}{p_i^T A p_i}$$

$0 \le i,\ j \le k$ and we have

$$v = P_k a = D_k c = P_k M c \rightarrow M c = a.$$

Problem 9.4

Recall the basic CD/CG properties given that at step k it is assumed CG has not converged,

- $x_k = \alpha_0 d_0 + \cdots + \alpha_{k-1} d_{k-1}$ is optimal (inherited from CD), i.e.,

$$\forall x \in x_0 + span[d_0, d_1, \ldots, d_{k-1}], \quad \|x_k - A^{-1}b\|_A \le \|x - A^{-1}b\|_A$$

- $< d_k,\ d_j >_A = 0$ $i \neq j$ for $0 \le i, j \le k-1$ (inherited from CD).
- $< r_k,\ d_j > = 0$ for $0 \le j \le k-1$ (inherited from CD).
- $< r_k,\ r_j > = 0$ for $0 \le j \le k-1$ (CG-specific).

- $span[d_0, d_1, \ldots, d_k] = span[r_0, r_1, \ldots, r_k]$ (CG-specific).
- $span[r_0, r_1, \ldots, r_k] = span[r_0, Ar_0, \ldots, A^k r_0]$ (CG-specific).

Given the inherited properties prove the three CG-specific properties.

Solution:

It is easy to see given $r_0 = d_0$ that

$$span[d_0,\ d_1] = span[r_0,\ r_1] = span[r_0,\ Ar_0]$$
$$r_1^T d_0 = r_1^T r_0 = 0$$

Now assume

$$span[d_0, d_1, \ldots, d_{k-1}] = span[r_0, r_1, \ldots, r_{k-1}] = span[r_0, \ldots, A^{k-1} r_0]$$
$$< r_{k-1},\ r_j >= 0, \ \ 0 \le j < k-1$$

and use induction.

We have

$$r_k \perp span[d_0, d_1, \ldots, d_{k-1}] \ \text{ from CD}$$
$$\forall j \le k-1, \ \ r_j \in span[d_0, d_1, \ldots, d_j] \ \text{ by induction hypothesis}$$
$$\therefore \ \text{ for } \ 0 \le j \le k-1, \quad < r_k,\ r_j >= 0$$

$$span[d_0, d_1, \ldots, d_{k-1}] = span[r_0, r_1, \ldots, r_{k-1}] \ \text{ by induction hypothesis}$$
$$(d_k, d_{k-1}) \text{ and } (r_k, d_{k-1}) \text{ are each linearly independent from CD properties}$$
$$\text{and } \beta_{k-1} \ne 0 \text{ by assumption}$$

$$d_k = r_k + \beta_{k-1} d_{k-1} \to d_k \in span[r_k,\ d_{k-1}] \ \text{ and } \ r_k \in span[d_k,\ d_{k-1}]$$

$$\therefore \ \ span[d_0, d_1, \ldots, d_{k-1}, d_k] = span[d_0, d_1, \ldots, d_{k-1}, r_k]$$
$$= span[r_0, r_1, \ldots, r_{k-1}, r_k]$$

We have

$$d_{k-1},\ \ r_{k-1} \in span[r_0, Ar_0, \ldots, A^{k-1}r_0] \ \text{ by induction hypothesis}$$

$$r_k = r_{k-1} - \alpha_{k-1}Ad_{k-1}$$
$$r_k = \gamma_0 r_0 + \cdots + \gamma_{k-1}A^{k-1}r_0 - \alpha_{k-1}A\left(\mu_0 r_0 + \cdots + \mu_{k-1}A^{k-1}r_0\right)$$
$$= \rho_0 r_0 + \cdots + \rho_{k-1}A^{k-1}r_0 + \rho_k A^k r_0$$

$$r_k \notin span[r_0, \ldots, r_{k-1}],\ \ A^k r_0 \notin span[r_0, \ldots, r_{k-1}]$$
$$\therefore r_k \in span[r_0, \ldots, r_{k-1}, A^k r_0]$$
$$\text{and }\ span[r_0, \ldots, r_{k-1}, A^k r_0] = span[r_0, \ldots, r_{k-1}, r_k]$$
$$span[r_0, \ldots, r_{k-1}, r_k] = span[r_0, \ldots, A^k r_0] \qquad \square$$