34 Krylov subspaces, and the optimality of CG

(**DRAFT**: in particular, more exercises will be added in time.)

In Theorem 33.1 we proved some important properties of the CG algorithm. We didn't get to Part (d). Although it easily follows from Parts (a)–(c), we are going to take a different approach.

34.1 Krylov subspaces

(Parts of this presentation are borrowed from Wikipedia!)

Given $\mathbf{b}$, a matrix A and vector $\mathbf{b}$, the *Krylov subspace* generated by A and $\mathbf{b}$ is

$$\mathcal{K}_n = \operatorname{span}\{\mathbf{b}, A\mathbf{b}, A^2\mathbf{b}, \dots, A^{n-1}\mathbf{b}\}. \tag{34.1}$$

It's importance originates from the fact that, due the Cayley-Hamilton Theorem, one can express the inverse of a matrix as a linear combinations of its powers. The Cayley-Hamilton theorem states that, if $p(\lambda)$ is the characteristic polynomial of A, i.e., $p(\lambda) = \det(\lambda I - A)$, then $p(A) = 0$.

To be clear, we don't mean that we substitute A for λ in the expression $\lambda I - A$. We mean we write out the polynomial

$$p(\lambda) = \lambda^m + c_{m-1}\lambda^{m-1} + \cdots + c_1\lambda + (-1)^m \det(A).$$

Then we substitute A for λ in the above polynomial. Then $p(A) = 0$ can be rearranged to get:

$$A(A^{m-1} + c_{m-1}A^{m-2} + \cdots + c_1) = -(-1)^m \det(A) I.$$

Dividing by $(-1)^{m-1}\det(A)$ gives a formula for A^{-1}. This means, that the solution, $\mathbf{x}$, to $A\mathbf{x} = \mathbf{b}$ can be written as a linear combination of the vectors in $\mathcal{K}_m$.

34.2 CG, with $\mathbf{x}^{(0)} = \mathbf{0}$

If we initiate the CG algorithm with $\mathbf{x}^{(0)} = \mathbf{0}$. So the algorithm becomes: set $\mathbf{r}^{(0)} = \mathbf{p}^{(0)} = \mathbf{b}$.
For $k = 1, 2, \dots$, set

$$\begin{aligned}
\mathbf{x}^{(k)} &= \mathbf{x}^{(k-1)} + a_{k-1}\mathbf{p}^{(k-1)}.\\
\mathbf{r}^{(k)} &= \mathbf{r}^{(k-1)} - a_{k-1}A\mathbf{p}^{(k-1)}\\
\mathbf{p}^{(k)} &= \mathbf{r}^{(k)} + b_{k-1}\mathbf{p}^{(k-1)}
\end{aligned}$$

where

$$a_{k-1} = \frac{(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)})}{(\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)})}, \quad b_{k-1} = \frac{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})}{(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)})}.$$

Clearly, $\mathbf{r}^{(0)} \in \mathcal{K}_0$, and $\mathbf{p}^{(0)} \in K_0$. Also, since $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + a_0\mathbf{p}^{(0)}$, and a_0 is just a scalar, we get that $\mathbf{x}^{(1)} \in \mathcal{K}_0$ too.

Next, using that $\mathbf{r}^{(1)} = \mathbf{r}^{(0)} - a_0 A\mathbf{p}^{(0)}$, we see that $\mathbf{r}^{(1)} \in \mathcal{K}_1$. Proceeding inductively, we get that

$$\begin{aligned}
\mathcal{K}_n &= \operatorname{span}\{\mathbf{x}^{(1)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}\\
&= \operatorname{span}\{\mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \dots, \mathbf{r}^{(n-1)}\}\\
&= \operatorname{span}\{\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(n-1)}\}.
\end{aligned}$$

Note that, for example, $\mathbf{r}^{(n)}$ is orthogonal to $\mathcal{K}_n$.

34.3 Optimality of CG

The presentation of this section is borrowed from [TB97, Lecture 38].

Theorem 34.1 (Thm. 38.2 of [TB97])**.** *Suppose that A is s.p.d. and the CG algorithm generates the sequence $\{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots\}$. If $\mathbf{r}^{(n-1)} \neq 0$, then $\mathbf{x}^{(n)}$ is the unique vector $\mathcal{K}_n$ for which $\|\mathbf{e}^{(n)}\|_A$ is minimised. Furthermore, $\|\mathbf{e}^{(n)}\|_A \leqslant \|\mathbf{e}^{(n-1)}\|_A$.*

34.4 Exercises

Exercise 34.1. Show that any $x \in \mathcal{K}_n$ can be expressed as $p(A)\mathbf{b}$, where p is some polynomial of degree at most $n - 1$.

THE END.