1 Lecture 12: Properties of the CG Algorithm

We showed that it is relatively easy to write an algorithm that satisfies properties 1 and 2 described in lecture 10. Namely we showed that if the search directions are $A$-orthogonal, then it naturally follows that the $k^{th}$ iterate can be written as $x_k = x_{k-1} + \alpha_k p_k$ while simultaneously maintaining that $x_k$ is the minimizer of $\phi$ over the span of all previous search directions. We still need to show that we satisfy property 3: $\text{span}\{p_1, p_2, \ldots, p_k\} = \mathcal{K}^k$. This is important because we know that the solution will be in the Krylov space after a reasonable number of iterations. Also, we would then be able naturally describe the convergence behavior of the method in terms of polynomials small over the spectrum of $A$.

We will inductively prove 4 properties of the CG algorithm described in the previous lecture. If the residual at the previous iteration is non-zero ($r_{k-1} \neq 0$), then:

- A. $\text{span}\{p_1, p_2, \ldots, p_k\} = \text{span}\{b, r_1, \ldots, r_{k-1}\} = \mathcal{K}^k$.
- B. $x_k = Q^k \lambda_k$ minimizes $\phi$ over $\mathcal{K}^k$.
- C. $p_i^T A p_j = 0$ if $i \neq j$ for $1 \leq i, j \leq k$.
- D. $r_i^T r_j = 0$ if $i \neq j$ for $1 \leq i, j \leq k$.

We will prove this inductively. First, this is trivially true when $k = 1$. Assume then that A-D are true for iteration $k - 1$ and that $r_{k-1} \neq 0$. We will now show that A-D all hold for the $k^{th}$ iterate as well.

**Property C.** First, we will show C holds. Given that $r_{k-1} \neq 0$ and given the argument put forth in section 3.1 of lecture 12. We can conclude that $p_k^T A p_j = p_j^T A p_k = 0$ for $1 \leq j \leq k - 1$. And that is enough because the other indices hold from the induction assumption.

**Property A.** First, we want to show that the search directions are linearly independent. We know that search directions 1 through $k-1$ are linearly independent from our induction hypothesis (since $\mathcal{K}^{k-1}$ is of dimension $k - 1$). We then just need to show that $p_k \notin \text{span}\{p_1, p_2, \ldots, p_{k-1}\}$. We can do this by contradiction. Assume $p_k \in \text{span}\{p_1, p_2, \ldots, p_{k-1}\}$. Now, since we just showed that property C holds at iteration $k$, we can conclude that although we construct the $k^{th}$ iterate as $x_k = x_{k-1} + \alpha_k p_k$, we can still say that $x_k = P^k \lambda_k$ where $\lambda_k$ is chosen to minimize $\phi$ over $\text{span}\{p_1, p_2, \ldots, p_k\}$. However, if $p_k \in \text{span}\{p_1, p_2, \ldots, p_{k-1}\}$ then $\text{span}\{p_1, p_2, \ldots, p_k\} = \text{span}\{p_1, p_2, \ldots, p_{k-1}\}$. Therefore, by property A at iteration $k - 1$ (which we are assuming is true in our induction hypothesis) we can conclude that $x_k = x_{k-1}$ since they are minimizers of $\phi$ (which are unique since $A$ is full rank). Therefore, it must be true that $\alpha_k p_k = 0$. First, assume $p_k \neq 0$. In this case, $\alpha_k = \frac{r_{k-1}^T p_k}{p_k^T A p_k}$ must be equal to zero. Now, by the definition of $p_k$ given in the previous lecture, we can see that $r_{k-1}^T p_k = p_k^T A p_k$. However, we just assumed $p_k^T A p_k \neq 0$ so this must not be possible. On the other hand, assume that $p_k = 0$. In this case (from the least squares definition of
that means Now, span $r$ therefore since we can say $R$ I.e. $R$

$$Ax$$

therefore we can say $r$ true that $r$

$$\{$$

However, we still need to show they span $K$. Therefore, it must be true that the search directions $p_i$ are linearly independent for $i = 1, 2, \ldots, k$. However, we still need to show they span $K^k$. First, the residuals are all of the form

$$r_i = r_{i-1} - \alpha_i Ap_i$$

therefore we can say

$$Ap_i = \frac{1}{\alpha_i} (r_{i-1} - r_i), \ i = 1, 2, \ldots, k - 1$$

In other words, \{Ap_1, Ap_2, \ldots, Ap_{k-1}\} $\subset$ span $\{b, r_1, \ldots, r_{k-1}\}$. Therefore, since $p_k = r_{k-1} - Ap^{k-1}z^{k-1}$ (see lecture 11), $p_k \in$ span $\{b, r_1, \ldots, r_{k-1}\}$. More specifically, we can say $P^k = [b, r_1, \ldots, r_{k-1}]M$ with $M \in \mathbb{R}^{k \times k}$ and det$(M) \neq 0$ (since the columns of $P^k$ are linearly independent). Therefore,

$$\text{span}\{p_1, p_2, \ldots, p_k\} = \text{span}\{b, r_1, \ldots, r_{k-1}\}.$$ 

Now, $r_{k-1} = b - Ax_{k-1}$ and $b \in K^k$ and $x_{k-1} \in K^{k-1}$ (by inductive hypothesis) and therefore $Ax_{k-1} \in K^k$ so since $r_{k-1}$ is a sum of two vectors in $K^k$ it is also in $K^k$. Therefore, $p_k = r_{k-1} - Ap^{k-1}z^{k-1} \in K^k$ since it is the sum of two vectors in $K^k$. Therefore,

$$\text{span}\{p_1, p_2, \ldots, p_k\} \subset K^k$$

and since we showed that the $p_i$ are linearly independent, we can conclude that property A holds for the $k$th iteration.

**Property B.** This follows by the definition of $\alpha_k$ given that we have already established C (which lets us use only the previous iterate and the new search direction when finding a minimizer of all previous directions) and A.

**Property D.** If we minimize $\phi$ over a set of the form span$\{v_1, v_2, \ldots, v_k\}$ (i.e. $x_k = V^k y_k$, $y_k \in \mathbb{R}^{k \times k}$) that the residual $r_k = b - Ax_k$ is orthogonal to the columns of the matrix $V^k = [v_1, \ldots, v_k]$.

I.e.

$$\left(V^k\right)^T r_k = 0, \text{ therefore } \left(P^k\right)^T r_k = 0$$

Now, span$\{p_1, p_2, \ldots, p_k\} = \text{span}\{b, r_1, \ldots, r_{k-1}\}$ from property B (which we just showed), so that means

$$r_k^T r_i = 0, \ i = 1, 2, \ldots, k - 1$$

Therefore since we can say $r_j^T r_i = 0$ for $i, j = 1, 2, \ldots, k - 1$ from our induction hypothesis, it is true that $r_j^T r_i = 0$ for $i, j = 1, 2, \ldots, k$. Therefore property D holds by induction.
2 Lecture 13: Connection with Lanczos Version

The Lanczos version of the algorithm can be shown to be equivalent to the energy based version.