Chapter 9

Conjugate Gradient Methods

The problem

Solving linear systems. Let A be a symmetric definite positive matrix ($n \times n$) and let $\boldsymbol{b} \in \mathbb{R}^n$. Then, solving the linear system

$$A\,\boldsymbol{x} = \boldsymbol{b} \tag{1}$$

is equivalent to solve the minimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \varphi(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^T A \boldsymbol{x} - \boldsymbol{b}\boldsymbol{x}. \tag{2}$$

Definition. The error associated to the vector $\boldsymbol{x}$ as a solution of (1) will be represented by

$$r(\boldsymbol{x}) := A\,\boldsymbol{x} - \boldsymbol{b} = \nabla\varphi(\boldsymbol{x}). \tag{3}$$

We also use the notation $r_k := r(\boldsymbol{x}_k)$.

A conjugate set of vectors

Definition. Let A be a symmetric definite positive matrix ($n \times n$). A system of non-zero vectors $\{p_0, \ldots, p_{n-1}\}$ is said to be **conjugate** with respect to A if

$$p_i^T A \, p_j = 0, \quad \text{for all } i \neq j. \tag{4}$$

Lemma. A conjugate system always exists and it is always linearly independent (base).

Proof. Exercise.

The conjugate direction method

Definition. Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ and a conjugate system $\{p_0, \ldots p_{n-1}\}$ we generate the following finite sequence

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k p_k \tag{5}$$

where α_k is the one-dimensional minimizer of the quadratic function φ along the p_k-direction $(\boldsymbol{x}_k + \alpha p_k)$. In particular,

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A\ p_k}. \qquad \text{(exercise)}$$

The conjugate direction method

Theorem. For any $\boldsymbol{x}_0 \in \mathbb{R}^n$, the (finite) sequence $\{\boldsymbol{x}_k\}$ generated by the conjugate direction method (5) converges to the solution $x^\star$ of the linear system (1) in at most n steps.

Proof. Since the conjugate system is a base of $\mathbb{R}^n$ we have

$$\boldsymbol{x}^\star - \boldsymbol{x}_0 = \beta_0 \boldsymbol{p}_0 + \cdots + \beta_{n-1} \boldsymbol{p}_{n-1},$$

for some (unique) choice of the parameters β_ℓ, $\ell = 0, \ldots, n-1$. We claim that

$$\beta_\ell = \frac{\boldsymbol{p}_\ell^T A(\boldsymbol{x}^\star - \boldsymbol{x}_0)}{\boldsymbol{p}_\ell^T A\, \boldsymbol{p}_\ell}.$$

To see the claim we pre-multiply the above expression by $p_\ell^T A$ and use the conjugacy property.

The conjugate direction method

Proof (continue).
On the other hand, using (5) we have that

$$x_\ell = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_{\ell-1} p_{\ell-1}.$$

Again by pre-multiplying the above expression by $p_\ell^T A$ and using the conjugacy property we have

$$p_\ell^T A(\boldsymbol{x}_k - \boldsymbol{x}_0) = 0.$$

Therefore

$$p_\ell^T A(\boldsymbol{x}^\star - \boldsymbol{x}_0) = p_\ell^T A(\boldsymbol{x}^\star - \boldsymbol{x}_k + \boldsymbol{x}_k - \boldsymbol{x}_0) = p_\ell^T A(\boldsymbol{x}^\star - \boldsymbol{x}_k) = p_\ell^T (\boldsymbol{b} - \boldsymbol{A} x_\ell) = -p_\ell^T r_\ell$$

Now we have

$$\beta_\ell = \frac{p_\ell A(\boldsymbol{x}^\star - \boldsymbol{x}_0)}{p_\ell^T A\ p_\ell} = -\frac{r_\ell^T p_\ell}{p_\ell^T A\ p_\ell} = \alpha_\ell$$

All together implies that $x_\ell = x^\star$ for some $\ell = 0, \ldots, n-1$.

The conjugate direction method

Exercise. Proof (or at least illustrate) the above theorem with a geometric argument in the case of A being a diagonal matrix.

Hint: Notice, first, that in this case the level curves of φ *are just ellipses whose axes are aligned with the coordinate directions and then we can use the conjugate system as the coordinates directions themselves.*

The conjugate direction method

Definition. Given a system of vectors $\{v_1, \ldots v_k\}$ we denote by $< v_1, \ldots v_k >$ the subspace generated by the vectors of the system.

Theorem. Let $\mathbf{x}_0 \in \mathbb{R}^n$ and consider the (finite) sequence $\{\mathbf{x}_k\}$ generated by the conjugate direction method (5). Then

$$r_k^T p_\ell = 0 \text{ for } \ell = 0, 1, \ldots k-1. \tag{6}$$

Moreover $\mathbf{x}_k$ is the minimizer of φ over the set

$$\{\mathbf{x} \mid \mathbf{x} = \mathbf{x}_0 + < p_0, \ldots, p_{k-1} >\}. \tag{7}$$

The conjugate direction method

Proof. We claim that a vector $\hat{x}$ minimizes φ over the set (7) if and only if $r(\hat{x})^T p_\ell = 0$ for all $\ell = 0, 1, \ldots k-1$. To see the claim we argue as follows. Let

$$h(\sigma) = \varphi\left(x_0 + \sum_{\ell=0}^{k-1} \sigma_\ell p_\ell\right), \quad \sigma = (\sigma_0, \ldots, \sigma_{k-1})^T .$$

Its unique minimizer, $\sigma^\star$ (h is strictly convex quadratic), satifies

$$\frac{\partial h(\sigma^\star)}{\partial \sigma_\ell} = 0, \ \ell = 0, \ldots, k-1.$$

So (chain rule)

$$\nabla\varphi\left(x_0 + \sum_{\ell=0}^{k-1} \sigma_\ell^\star p_\ell\right) p_\ell = r(\hat{x})^T p_\ell = 0, \quad \ell = 0, \ldots, k-1.$$

Now we should proof (6) (for all $\ell = 0, 1, \ldots k-1$). We use induction.

The conjugate direction method

Proof.

- The case $k = 1$ follows form the previous arguments. Since $x_1 = x_0 + \alpha_0 p_0$ minimizes φ along p_0 we have that $r_1^T p_0 = 0$.
- Assume $r_{k-1}^T p_\ell = 0$ for all $\ell = 0, \ldots, k-2$. One can verify that $r_k = r_{k-1} + \alpha_{k-1} A\, p_{k-1}$. So

$$p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0$$

  where the last equality follows from the definition of α_{k-1}. For the rest of conjugate directions p_ℓ, $\ell = 0, \ldots, k-2$ we have

$$p_\ell^T r_k = p_\ell^T r_{k-1} + \alpha_{k-1} p_\ell^T A p_{k-1} = 0$$

  because of the induction hypothesis and the fact that p_ℓ is a conjugate vector with respect to A.

The conjugate gradient method

Definition. The conjugate gradient method is a conjugate direction method such in generating its set of conjugate vectors the new vector p_k can be computed using only the previous vector p_{k-1}. The algorithm gives for free the conjugacy property of p_k with respect to all previous $\{p_0, \dots p_{k-2}\}$ vectors of the conjugate system. More concretely, p_0 is the steepest descent direction at the initial point x_0 and then we define

$$p_k = -r_k + \beta_k p_{k-1},$$

where β_k is determined such that $p_k^T A \, p_{k-1} = 0$.

Lemma. Under the above notation we have

$$\beta_k = \frac{r_k^T A \, p_{k-1}}{p_{k-1}^T A \, p_{k-1}}. \qquad \text{(exercise)}$$

The conjugate gradient method

Lemma. Suppose that the k-th iterate generated by the conjugate gradient method is not the solution point $x^\star$. Then the following statements hold.

(a) $r_k^T r_\ell = 0$ for $\ell = 0, \ldots, k-1$.

(b) $< r_0, \ldots, r_k > = < p_0, \ldots, p_k > = < r_0, Ar_0, \ldots, A^k r_0 >$.

(c) $p_k^T A\, p_\ell = 0$ for $\ell = 0, \ldots, k-1$.

Therefore, the (finite) sequence $\{x_k\}$ converges to $x^\star$ in at most n steps.

The conjugate gradient method

Proof (of (a)). Because the conjugate gradient method is a conjugate method we already have that

$$r_k^T p_\ell = 0 \quad \text{for} \quad \ell = 0, \ldots, k-1 \quad \text{and for all} \quad k = 1, \ldots, n-1.$$

But now since

$$p_\ell = -r_\ell + \beta_\ell p_{\ell-1}$$

we conclude that $r_\ell \in < p_\ell, p_{\ell-1} >$. So,

$$r_k^T r_\ell = r_k^T (\hat{c}_1 p_\ell + \hat{c}_2 p_{\ell-1}) = 0 \quad \text{for all} \quad \ell = 1, \ldots, k-1.$$

By definition p_0 satisfies $r_k^T r_0 = -r_k^T p_0 = 0$ which finishes the proof.

The conjugate gradient method: Rate of convergence

Definition. We might consider the following choices of te parameters α_k and β_k above.

$$\hat{\alpha}_k = \frac{r_k^T p_k}{p_k^T A \, p_k} \quad \text{and} \quad \hat{\beta}_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

The conjugate gradient method: Rate of convergence

We might construct the following algorithm CG: Conjugated gradient.

Let $x_0 \in \mathbb{R}^n$. Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$. While $r_k \neq 0$.

$$\begin{aligned}
\hat{\alpha}_k &\leftarrow \frac{r_k^T p_k}{p_k^T A\, p_k} \\
x_{k+1} &\leftarrow x_k + \hat{\alpha}_k p_k \\
r_{k+1} &\leftarrow x_k + \hat{\alpha}_k p_k \\
\hat{\beta}_{k+1} &\leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} \\
p_{k+1} &\leftarrow -r_{k+1} + \hat{\beta}_{k+1} p_k \\
k &\leftarrow k+1 \quad \text{end (while)}
\end{aligned} \tag{8}$$

The conjugate gradient method: Rate of convergence

Theorem. If A has only $r \leq n$ distinct eigenvalues, then the CG method iteration will terminate at the solution in aat most r iterations.