The conjugate gradient method

Nicholas Hu

The conjugate gradient method is an iterative method for solving the linear system $Ax = b$, where $A \in \mathbb{R}^{n \times n}$ is symmetric positive-definite.

Let $\langle x, y \rangle_A = \langle Ax, y \rangle$ be the inner product defined by A and $\|x\|_A = \sqrt{\langle x, x \rangle_A}$ be the induced norm. Given an initial guess $x^{(0)}$ for the solution x^*, the k^{th} iterate of the method is defined as

$$x^{(k)} = \operatorname*{arg\,min}_{x \in x^{(0)} + \mathcal{K}_k(A, r^{(0)})} \|x^* - x\|_A \,,$$

where $\mathcal{K}_k(A, r^{(0)})$ is the Krylov subspace $\operatorname{span}\{A^j r^{(0)}\}_{j=0}^{k-1}$ and $r^{(0)} = b - Ax^{(0)}$. (In other words, the A-norm of the error is minimized over the k^{th} affine Krylov subspace generated by the initial residual and translated by the initial guess.) [1]

Let us abbreviate $\mathcal{K}_k(A, r^{(0)})$ as $\mathcal{K}_k$ and write $r^{(k)} = b - Ax^{(k)}$ for the residual of the k^{th} iterate. The iterate $x^{(k)}$ is therefore the A-orthogonal projection of x^* onto $x^{(0)} + \mathcal{K}_k$, defined by the Galerkin conditions $x^{(k)} - x^{(0)} \in \mathcal{K}_k$ and $x^* - x^{(k)} \perp_A \mathcal{K}_k$; we note that the orthogonality condition is equivalent to $r^{(k)} \perp \mathcal{K}_k$.

Now suppose that $\{p^{(j)}\}_{j<k}$ is a basis of $\mathcal{K}_k$ and let $P_k = \begin{bmatrix} p^{(0)} & \cdots & p^{(k-1)} \end{bmatrix}$. Then $x^{(k)} = x^{(0)} + P_k y^{(k)}$, where

$$y^{(k)} = \operatorname*{arg\,min}_{y \in \mathbb{R}^k} \|x^* - (x^{(0)} + P_k y)\|_A \,.$$

If $p^{(k)}$ is such that $\{p^{(j)}\}_{j<k+1}$ is a basis of $\mathcal{K}_{k+1}$, we can express the next iterate $x^{(k+1)}$ in an analogous manner (that is, $x^{(k+1)} = x^{(0)} + P_{k+1} y^{(k+1)}$, where $P_{k+1} = \begin{bmatrix} P_k & p^{(k)} \end{bmatrix}$). Writing $y^{(k+1)} = \begin{bmatrix} \tilde{y}^{(k)} \\ \alpha_k \end{bmatrix}$ for some $\tilde{y}^{(k)} \in \mathbb{R}^k$ and $\alpha_k \in \mathbb{R}$, we see that

$$\begin{aligned} x^* - (x^{(0)} + P_{k+1} y^{(k+1)}) &= [x^{(k)} - (x^{(0)} + P_k \tilde{y}^{(k)})] + [(x^* - x^{(k)}) - \alpha_k p^{(k)}] \\ &= P_k (y^{(k)} - \tilde{y}^{(k)}) + [(x^* - x^{(k)}) - \alpha_k p^{(k)}] \,. \end{aligned}$$

Thus, if we select $p^{(k)}$ to be A-orthogonal to $p^{(j)}$ for all $j < k$, then by the Pythagorean theorem,

$$\|x^* - (x^{(0)} + P_{k+1} y^{(k+1)})\|_A^2 = \|P_k (y^{(k)} - \tilde{y}^{(k)})\|_A^2 + \|(x^* - x^{(k)}) - \alpha_k p^{(k)}\|_A^2 \,,$$

so the solution to the least squares problem for $y^{(k+1)}$ is given recursively by $\tilde{y}^{(k)} = y^{(k)}$ and $\alpha_k = \operatorname{proj}^A_{p^{(k)}}(x^* - x^{(k)})$. It follows that

$$\begin{aligned} x^{(k+1)} &= x^{(0)} + P_k y^{(k)} + \alpha_k p^{(k)} \\ &= x^{(k)} + \alpha_k p^{(k)}, \end{aligned} \tag{X}$$

where

$$\begin{aligned}\alpha_k &= \frac{\langle x^* - x^{(k)}, p^{(k)}\rangle_A}{\langle p^{(k)}, p^{(k)}\rangle_A} \\ &= \frac{\langle r^{(k)}, p^{(k)}\rangle}{\langle p^{(k)}, p^{(k)}\rangle_A}\,. \end{aligned}\tag{A}$$

This also implies that

$$r^{(k+1)} = r^{(k)} - \alpha_k A p^{(k)}. \tag{R}$$

To generate A-orthogonal vectors $p^{(j)}$ such that $\{p^{(j)}\}_{j<k}$ is a basis of $\mathcal{K}_k$ for each k, we notice that $r^{(k+1)} \perp_A \mathcal{K}_k = \operatorname{span}\{p^{(j)}\}_{j<k}$ because $r^{(k+1)} \perp \mathcal{K}_{k+1}$ and $A\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$. As a result, when $r^{(k+1)}$ is A-orthogonalized against $p^{(k)}$, the resulting vector will automatically be A-orthogonal to $p^{(j)}$ for all $j < k+1$, suggesting that we define

$$\begin{aligned} p^{(k+1)} &= r^{(k+1)} - \operatorname{proj}^A_{p^{(k)}} r^{(k+1)} \\ &= r^{(k+1)} + \beta_k p^{(k)}, \end{aligned}\tag{P}$$

where $p^{(0)} = r^{(0)}$ and

$$\beta_k = -\frac{\langle r^{(k+1)}, p^{(k)}\rangle_A}{\langle p^{(k)}, p^{(k)}\rangle_A}\,. \tag{B}$$

Referring back to the residual equation (R), we can show by induction that the $p^{(j)}$ thus defined will also constitute bases of successive Krylov subspaces. More precisely, suppose that the solution has not been found by the beginning of the k^{th} iteration, in the sense that $r^{(j)} \neq 0$ for all $j < k$. We claim then that $r^{(k-1)} \in \mathcal{K}_k$ and that $\{p^{(j)}\}_{j<k}$ is an A-orthogonal basis of $\mathcal{K}_k$.

Indeed, if $r^{(0)} \neq 0$, then $r^{(0)} \in \mathcal{K}_1 = \operatorname{span}\{r^{(0)}\}$ and $\{p^{(0)}\} = \{r^{(0)}\}$ is an A-orthogonal basis of $\mathcal{K}_1$. Now suppose that the claim holds up to the k^{th} iteration and that its hypothesis is satisfied at the beginning of the $(k+1)^{\text{th}}$ iteration. Then

$$r^{(k)} = r^{(k-1)} - \alpha_{k-1} A p^{(k-1)} \in \mathcal{K}_k + A\mathcal{K}_k \subseteq \mathcal{K}_{k+1}\,,$$

so

$$p^{(k)} = r^{(k)} + \beta_{k-1} p^{(k-1)} \in \mathcal{K}_{k+1} + \mathcal{K}_k \subseteq \mathcal{K}_{k+1}\,.$$

In addition, $p^{(k)} \neq 0$ because $r^{(k)} \perp \mathcal{K}_k$ and $r^{(k)} \neq 0$. Hence, by construction, $\{p^{(j)}\}_{j<k+1}$ is an A-orthogonal set of nonzero vectors in $\mathcal{K}_{k+1}$ and is moreover a basis thereof, since $\dim(\mathcal{K}_{k+1}) \leq k+1$.

An immediate consequence is that $\{r^{(j)}\}_{j<k}$ will be an orthogonal basis of $\mathcal{K}_k$ for all such iterations: if (say) $i < j < k$, then $r^{(i)} \in \mathcal{K}_{i+1} \subseteq \mathcal{K}_j$, and we know that $r^{(j)} \perp \mathcal{K}_j$. Furthermore, the iteration will break down exactly when $r^{(k)} \in \mathcal{K}_k$, or equivalently, $r^{(k)} = 0$, meaning that the solution was attained in the k^{th} iteration.

We can also derive alternative formulas for the scalars α_k and β_k that reduce the number of inner products in each iteration. First, using the fact that $r^{(k)} \perp \mathcal{K}_k = \operatorname{span}\{p^{(j)}\}_{j<k}$, we obtain

$$
\begin{aligned}
\alpha_k &= \frac{\langle r^{(k)}, p^{(k)} \rangle}{\langle p^{(k)}, p^{(k)} \rangle_A} \\
&= \frac{\langle r^{(k)}, r^{(k)} + \beta_{k-1} p^{(k-1)} \rangle}{\langle p^{(k)}, p^{(k)} \rangle_A} \\
&= \frac{\langle r^{(k)}, r^{(k)} \rangle}{\langle p^{(k)}, p^{(k)} \rangle_A} . \qquad \text{(A)}
\end{aligned}
$$

Hence

$$
\begin{aligned}
\beta_k &= -\frac{\langle r^{(k+1)}, p^{(k)} \rangle_A}{\langle p^{(k)}, p^{(k)} \rangle_A} \\
&= -\alpha_k \frac{\langle r^{(k+1)}, p^{(k)} \rangle_A}{\langle r^{(k)}, r^{(k)} \rangle} \\
&= \frac{\langle r^{(k+1)}, r^{(k+1)} - r^{(k)} \rangle}{\langle r^{(k)}, r^{(k)} \rangle} \\
&= \frac{\langle r^{(k+1)}, r^{(k+1)} \rangle}{\langle r^{(k)}, r^{(k)} \rangle} . \qquad \text{(B)}
\end{aligned}
$$

In summary,

> $$
> \begin{aligned}
> r^{(0)} &= b - Ax^{(0)} \\
> p^{(0)} &= r^{(0)} \\
> \\
> \alpha_k &= \frac{\langle r^{(k)}, r^{(k)} \rangle}{\langle p^{(k)}, p^{(k)} \rangle_A} && \text{A} \\
> x^{(k+1)} &= x^{(k)} + \alpha_k p^{(k)} && \text{X} \\
> r^{(k+1)} &= r^{(k)} - \alpha_k A p^{(k)} && \text{R} \\
> \beta_k &= \frac{\langle r^{(k+1)}, r^{(k+1)} \rangle}{\langle r^{(k)}, r^{(k)} \rangle} && \text{B} \\
> p^{(k+1)} &= r^{(k+1)} + \beta_k p^{(k)} && \text{P}
> \end{aligned}
> $$

---

1. The choice of this minimization problem can be partially motivated as follows. In view of the fact that $x^* = x^{(0)} + A^{-1} r^{(0)}$ and that A^{-1} is a polynomial in A of degree at most $n-1$, in the k^{th} iteration of the method, we seek an approximation to the solution of the form $x^{(0)} + p_{k-1}(A) r^{(0)}$, where p_{k-1} is a polynomial of degree at most $k-1$. This guarantees that the A-norm of the error decreases monotonically and that the solution is found in at most n iterations (in exact arithmetic). Although the choice of the objective function is not canonical, it turns out that this choice leads to a particularly tractable method. ↩