Conjugate Gradient Method (for choosing a descent direction)

Consider the optimization problem

$$\min \ f(x) = \tfrac{1}{2} x^T A x - b^T x$$
$$\text{s.t. } x \in \mathbb{R}^n.$$

where A is a given symmetric positive definite $n \times n$ matrix and $b \in \mathbb{R}^n$ a given vector.

This objective is a convex quadratic function so is bounded below. Furthermore, the unique solution is the stationary point of f:

$$\nabla f(x) = Ax - b$$

Notice that when $\nabla f(x) = 0$, $Ax = b$. So, this optimization problem is a fancy way to find the solution to a (positive definite) set on n linear equations in n unknowns.

Aside:

$$f(x) = \frac{1}{2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{k\ell} x_k x_\ell - \sum_{k=1}^{n} b_k x_k$$

$$\frac{\partial f}{\partial x_i} = \frac{1}{2} \sum_{k=1}^{n} a_{ki} x_k + \frac{1}{2} \sum_{\ell=1}^{n} a_{i\ell} x_\ell - b_i$$

$$= \sum_{\ell=1}^{n} a_{i\ell} x_\ell - b_i$$

$$= a_i x - b_i \qquad A = \begin{bmatrix} — a_1^T — \\ \vdots \\ — a_n^T — \end{bmatrix}$$

$$\nabla f(x) = Ax - b$$

Consider any set of conjugate directions $\{p_0, p_1, \ldots, p_\ell\}$ $(p_i^T A p_j = 0,\ i \neq j)$

Then we can find the minimizer in at most n steps by successive exact line searches along descent directions p_k (Theorem 5.1)

$$x_{k+1} = x_k + \alpha_k p_k \qquad \alpha_k = \operatorname*{argmin}_{\beta} f(x_k + \beta p_k)$$

$$\Rightarrow \quad A(x_k + \alpha_k p_k) = b$$

$$\Rightarrow \quad A \alpha_k p_k = b - A x_k$$

$$\alpha_k p_k^T A p_k = p_k^T (b - A x_k)$$

$$\alpha_k = \frac{-r_k^T p_k}{p_k^T A p_k}$$

A conjugate gradient algorithm for "linear systems"

1. Given: x_0

2. set: $r_0 = Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$

3. Compute next iterate

$$\alpha_k \leftarrow -\frac{r_k^T p_k}{p_k^T A p_k}$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k$$

4. Updates

$$r_{k+1} \leftarrow A x_{k+1} - b$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$$

$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$$

$$k \leftarrow k+1$$

Why? Each new conjugate direction is chosen to be:

$$p_{k+1} = -r_{k+1} + \beta_{k+1} p_k \quad \text{for some } \beta_{k+1}.$$

require $p_{k+1} A p_k = 0 \Rightarrow \beta_{k+1}$ as shown!

5. If $r_k = 0$ then stop
otherwise goto step 3

Notice that each successive conjugate direction depends only on the previous direction, A, b, and current residual r.

Contrast this with a Gram-Schmidt method which requires knowledge of all previous directions.

Key Ideas on Convergence

Theorem 5.4: If A has r distinct eigenvalues, then CG terminates in no more than r iterations.

Theorem 5.5: If A has eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, then CG satisfies

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \| x_0 - x^* \|_A^2$$

example $\lambda = 1, 1, 1, 2, 3, 5, 8$ $(n=7)$

$$\| x_1 - x^* \|^2 \le (7/9) \| x_0 - x^* \|^2$$
$$\| x_2 - x^* \|^2 \le (4/6) \| x_0 - x^* \|^2$$
$$\| x_3 - x^* \|^2 \le (2/4) \| x_0 - x^* \|^2$$
$$\| x_4 - x^* \|^2 \le (1/3) \| x_0 - x^* \|^2$$
$$\| x_5 - x^* \|^2 \le (0) \| x_0 - x^* \|^2$$

Using the following properties of conjugate directions:

$$\left.\begin{aligned} r_k^T p_j &= 0 \\ r_k^T r_j &= 0 \\ p_k^T A p_j &= 0 \end{aligned}\right\} \quad j = 1, 2, \ldots, k-1$$

We can rewrite the algorithm into a more symmetric and useful form:

1. Given: x_0

2. Set: $r_0 = Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$

3. Compute next iterate

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k} = \frac{\|r_k\|^2}{\|p_k\|_A^2}$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k$$

4. Updates

$$r_{k+1} \leftarrow A x_{k+1} - b$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} = \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$$

$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$$

$$k \leftarrow k+1$$

5. If $r_k = 0$ then stop
   Otherwise goto step 3

Nonlinear Conjugate Gradient Method

- Step length α_k to be determined by line search
- Residual r_k to be replaced by the gradient $\nabla f(x_k)$

Fletcher-Reeves Algorithm

1. Given: x_0, $\varepsilon > 0$.
2. Set: $p_0 = -\nabla f(x_0)$, $k \leftarrow 0$
3. If $\nabla f(x_k) < \varepsilon$ then stop
4. Compute $\alpha_k = \arg\min_{\beta} f(x_k + \alpha_k p_k)$ $\leftarrow$ This line search is performed using strong Wolfe conditions with

   $$0 < c_1 < c_2 < 1/2$$

   to ensure that p_k is a descent direction.
5. Updates

$$x_{k+1} = x_k + \alpha_k p_k$$

$$\beta_{k+1} = \frac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)}$$

$$p_{k+1} = -\nabla f(x_{k+1}) + \beta_{k+1} p_k$$

$$k \leftarrow k+1$$

6. Go to step 3

This conjugate gradient method will require periodic restarts. It is natural to choose every n iterations $p_k = -\nabla f(x_k)$

Polak-Ribière Method

$$\beta_{k+1} = \frac{\nabla f(x_{k+1})^T \left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)}{\nabla f(x_k)^T \nabla f(x_k)}$$

(natural restart)

$$\beta_{k+1} \leftarrow \max\{\beta_{k+1}, 0\}$$