Table of Contents

The Quadratic Form

$$f(\underline{x}) = \frac{1}{2}\,\underline{x}^T A \underline{x} - \underline{b}^T \underline{x} + c$$

A is a matrix, $\underline{b}$ and $\underline{x}$ are vectors, c is a scalar constant

$$f'(\underline{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(\underline{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\underline{x}) \end{pmatrix} = \frac{1}{2} A^T \underline{x} + \frac{1}{2} A \underline{x} - \underline{b}$$

Since we concentrate on positive-definite symmetric matrices A the gradient of $f(\underline{x})$ has the form

$$f'(\underline{x}) = A\underline{x} - \underline{b}$$

↳ Pictures of the Graph

If we now minimize $f(\underline{x})$ we'll get

$$\underline{0} = A\underline{x} - \underline{b}$$

We see that minimizing the quadratic form is nothing else than searching the solution point $\underline{x}$ of our system of linear equations.

The Method of Steepest Descent

Idea:

- Search the direction in which f decreases most quickly.
- Follow this direction to some special point.
- Iterate these steps thus moving to the minimum point.

The direction of steepest descent simply is $-f'(x_i)$.
This is because as the gradient points to the direction in which f increases most quickly you just have to move to opposite direction.

We now define

$e_i := x_i - x$ "error term"

$r_i = -f'(x_i) = b - Ax_i = Ax - Ax_i = A(x - x_i) = -Ae_i$ "residual"

We also define

$x_{i+1} = x_i + \alpha \cdot r_i$

We now minimize f in each iteration setting the directional derivative $\frac{\partial f(x_{i+1})}{\partial \alpha}$ to zero.

$$\frac{\partial}{\partial \alpha} f(x_{i+1}) = \left(\frac{\partial}{\partial x_i} f(x_{i+1})\right) \cdot \left(\frac{\partial}{\partial \alpha} x_{i+1}\right) = -r_{i+1}^T \cdot r_i = 0$$

We receive for α

$$r_{i+1}^T \cdot r_i = 0 = (b - Ax_{i+1})^T \cdot r_i = (b - Ax_i - \alpha A r_i)^T \cdot r_i = (b - Ax_i)^T r_i - \alpha (Ar_i)^T \cdot r_i$$

$$\alpha = \frac{r_i^T \cdot r_i}{r_i^T \cdot A \cdot r_i}$$

So the method of Steepest Descent all in all is

$$r_i = b - Ax_i \qquad (1)$$

$$\alpha_{(i)} = \frac{r_i^T \cdot r_i}{r_i^T A r_i} \qquad (2)$$

$$x_{i+1} = x_i + \alpha_{(i)} r_i \qquad (3)$$

In order to achieve more efficiency we also use

$$r_{i+1} = r_i - \alpha_{(i)} A r_i$$

Thus less matrix-vector products per iteration are required.

Conjugate Gradients

We now try to make this method more efficient. A way to achieve in this will be making steps of the right size along the search line which means that we will only make at most one step to each direction. The result would obviously be that we will be done after n iterations.

The trick is to see the error e_0 as a linear combination of n search direction vectors.

$$e_0 = \sum_{j=0}^{n-1} \alpha_j d_j$$

In each iteration one component will be taken out and so we have

$$e_i = \sum_j \alpha_j d_j$$

Now we can choose the search directions d_i such that every new d_j is conjugate to the next error term.

$$d_i^T A e_{i+1} = 0$$

We now have to find a way to construct those d_i's efficiently.

Things we know and things we want

$$x_{i+1} = x_i + \alpha d_i$$

$$d_i^T A e_{i+1} = 0$$

Now we also choose the d_j's such that they are A-orthogonal to each other.

$$d_i^T A d_j = 0 \quad , \; i \neq j$$

One way to do this is the Gram-Schmidt Conjugation but since this algorithm requires $O(n^3)$ steps we decide that it's not the best solution.
Another way is much more efficient and we now will get to know why the method is called CG.
Idea:
Why don't we construct our new search directions from the residual?
We already know that it's orthogonal to previous search directions because

$$-d_i^T \cdot A \cdot e_j = -\sum_{j=i}^{n-1} \alpha_{(j)} d_i^T \cdot A \cdot d_j$$

$$d_i^T \cdot r_j = 0 \quad , \; i < j$$

From this we can conclude that

$$r_i^T \cdot r_j = 0 \quad , \; i \neq j$$

and since
$$r_{i+1} = -Ae_{i+1} = -A(e_i + \alpha d_i) = r_i - \alpha A d_i$$
we know that
$$\operatorname{span}\{r_0, r_1, \ldots, r_{n-1}\} = \operatorname{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{n-1} r_0\}$$

This subspace is called a Krylov subspace.
We will refer to it later.

What we now try to do is fullfilling the following equation

$$d_{i+1} = r_{i+1} + \beta_{(i)} d_i$$

We simply accept at this point that we can compute $\beta_{(i)}$ by using

$$\beta_{(i+1)} = \frac{r_{i+1}^T \cdot r_{i+1}}{r_i^T \cdot r_i}$$

Another thing we accept is that

$$\alpha_{(i)} = \frac{r_i^T \cdot r_i}{r_i^T \cdot A d_i}$$

So the whole Method of CG is

$$d_0 = r_0 = b - A x_0$$

$$\alpha_{(i)} = \frac{r_i^T \cdot r_i}{r_i^T \cdot A d_i}$$

$$x_{i+1} = x_i + \alpha_{(i)} d_i$$

$$\beta_{(i+1)} = \frac{r_{i+1}^T \cdot r_{i+1}}{r_i^T \cdot r_i}$$

$$d_{i+1} = r_{i+1} + \beta_{(i+1)} d_i$$

$$r_{i+1} = r_i - \alpha_{(i)} \cdot A d_i$$

Complexity Analysis

Definition:

$$\|e\|_A = (e^T A e)^{\frac{1}{2}} \quad \text{energy norm}$$

Minimizing this norm of course is the same as minimizing f. (Actually not "of course" but it's a property of the quadratic form).

The subspace mentioned before now will be defined as

$$\mathcal{D}_i = \operatorname{span}\{r_0, Ar_0, \dots, A^{i-1}r_0\}$$

$$= \operatorname{span}\{d_0, Ad_0, \dots, A^{i-1}d_0\} = \operatorname{span}\{Ae_0, A^2e_0, \dots, A^ie_0\}$$

Now we can express the error e_0 as

$$e_0 = \sum_{j=1}^{n} \xi_j v_j \quad , v_j \text{ eigenvectors, unit length}$$

and $e_i = \sum_j \xi_j v_j = \left(I + \sum_{j=1}^{i} \psi_j A^j \right) e_0$

This leads to $e_i = P_i(A) e_0 \quad , P_i(0) = 1$

$$= \sum_j \xi_j P_i(\lambda_j) v_j$$

$$Ae_i = \sum_j \xi_j P_i(\lambda_j) \lambda_j v_j$$

$$\|e_i\|_A^2 = e_i^T A e_i = \sum_j \xi_j^2 [P_i(\lambda_j)]^2 \lambda_j$$

Now what CG actually does is picking a polynomial P_i of degree i that minimizes the error term. This polynomial is optimal. $\Delta(A)$ is our set of eigenvalues.

$$\|e_i\|_A^2 \leq \min_{P_i} \max_{\lambda \in \Delta(A)} [P_i(\lambda)]^2 \cdot \sum_j \xi_j^2 \lambda_j$$

$$= \min_{P_i} \max_{\lambda \in \Delta(A)} [P_i(\lambda)]^2 \|e_0\|_A^2$$

At this point we'll go back to the beginning.
In Steepest Descend we took the residual as our search direction.