The Conjugate Gradient Method is an algorithm for finding a numerical solution of a system of linear equations of the form

$$Ax = b \tag{1}$$

where A is a symmetric, positive-definite matrix.

0.1 Quadratic Form

Finding conflicting definitions for a *quadratic form* when considering whether the function is homogeneous or not.

Definition 1. A *quadratic form* is a scalar quadratic function of a vector with the form

$$f(x) = x^T Ax - b^T x + c \tag{2}$$

where A is a matrix, x and b are vectors, and c is a scalar constant.

0.1.1 Derivative of a Quadratic Form

Let $f(x) = x^T Ax$ over the real numbers. Let $y(x) = Ax$ so that $f(x, y(x)) = x^T y(x)$, then

$$\begin{aligned} f'(x, y(x)) &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot y'(x) \\ &= y(x)^T + x^T \cdot A \\ &= (Ax)^T + x^T A \\ &= x^T A^T + x^T A \end{aligned}$$

Let $g(x) = \frac{1}{2} f(x)$ from (2) then

$$g'(x) = x^T A^T - b^T \tag{3}$$

Suppose $g'(x) = 0$ and A is symmetric, then solving the linear system $Ax = b$ is equivalent to finding the minimum/maximum of the quadratic form $g(x)$. If A is positive-definite, then this will be a minimum.

Claim 1. Let A be a symmetric, positive-definite matrix, then x is a solution for $Ax = b$ if and only if $g(x) = \frac{1}{2} x^T Ax - b^T x + c$ is minimized at x.

Proof. ... □

0.2 Method of Steepest Descent

Very intuitive. $\nabla f(x)$ points in the direction of steepest ascent at $(x, f(x))$. By equation (3), the steepest descent is then

$$-f'(x) = b - Ax. \tag{4}$$

What size of step α should we take in direction $-f'(x)$? Clearly the α which minimizes f along the line,

$$x_{(i+1)} = x_{(i)} + \alpha r_{(i)} \tag{5}$$

where $r_{(i)} = -f'(x_i)$ or the error transformed by A into the same space as b. α minimizes f when $\frac{d}{d\alpha}(f(x_i)) = 0$

$$\frac{d}{d\alpha} f(x_{(i)}) = r_{(i)} \cdot f'(x_{(i+1)})^T = 0$$

when $r_{(i)}$ and $f'(x_{(i+1)})$ are orthogonal. Solving for α,

$$\begin{aligned}
r_{(i)} \cdot f'(x_{(i+1)})^T &= 0 \\
r_{(i)} \cdot (b - Ax_{(i+1)})^T &= 0 \\
r_{(i)} \cdot (b - A(x_{(i)} + \alpha r_i))^T &= 0 \\
&\cdots \\
\alpha &= \frac{r_{(i)}^T r_{(0)}}{r_{(i)}^T A r_{(0)}} \qquad (6)
\end{aligned}$$