Derivation of Exact Step Size in Steepest Descent

$$f(x_k) = \frac{1}{2} x_k^T Q x_k - b^T x$$

Steepest Descent

The quadratic function is

$$\boxed{f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x}}$$

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad Q = \begin{bmatrix} a & c \\ c & b \end{bmatrix}.$$

R

U

A

Explicitly,

$$f(x, y) = \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & c \\ c & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Expanding the quadratic term:

$$f(x, y) = \frac{1}{2}(ax^2 + 2cxy + by^2) - b_1x - b_2y = \frac{a}{2}x^2 + cxy + \frac{b}{2}y^2 - b_1x - b_2y.$$

The gradient is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} ax + cy - b_1 \\ cx + by - b_2 \end{pmatrix} = Q\mathbf{x} - \mathbf{b}.$$

Thus,

$$\nabla f(\mathbf{x}) = -(\mathbf{b} - Q\mathbf{x}).$$

Taking transpose,

$$[\nabla f(\mathbf{x})]^T = -(\mathbf{b}^T - \mathbf{x}^T Q),$$

so

$$-[\nabla f(\mathbf{x})]^T = \mathbf{b}^T - \mathbf{x}^T Q.$$

Since $Q = Q^T$ (symmetric).

$f(x) =$

Iterative Procedure

The steepest descent update rule is

α

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta_k \nabla f(\mathbf{x}_k),$$

where $\eta_k > 0$ is the step size.

Let $\mathbf{S} = \nabla f(\mathbf{x}_k)$ and let Q be the Hessian matrix

$$Q = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}.$$

$f(x_k - \eta_k S)$

To find the optimal η_k, define the line search function

α

$$g(\eta_k) = f(\mathbf{x}_k - \eta_k \nabla f(\mathbf{x}_k)) = f(\mathbf{x}_k - \eta_k \mathbf{S}).$$

Substitute into the quadratic form:

$$\begin{aligned} g(\eta_k) &= \frac{1}{2}(\mathbf{x}_k - \eta_k \mathbf{S})^T Q (\mathbf{x}_k - \eta_k \mathbf{S}) - \mathbf{b}^T(\mathbf{x}_k - \eta_k \mathbf{S}) \\ &= \frac{1}{2}\mathbf{x}_k^T Q \mathbf{x}_k - \eta_k \mathbf{S}^T Q \mathbf{x}_k + \frac{\eta_k^2}{2}\mathbf{S}^T Q \mathbf{S} - \mathbf{b}^T \mathbf{x}_k + \eta_k \mathbf{b}^T \mathbf{S} \\ &= \frac{\eta_k^2}{2}\mathbf{S}^T Q \mathbf{S} + \eta_k(\mathbf{b}^T \mathbf{S} - \mathbf{x}_k^T Q \mathbf{S}) + \frac{1}{2}\mathbf{x}_k^T Q \mathbf{x}_k - \mathbf{b}^T \mathbf{x}_k. \end{aligned}$$

Using the relation $-[\nabla f(\mathbf{x}_k)]^T = \mathbf{b}^T - \mathbf{x}_k^T Q$,

$$\mathbf{b}^T \mathbf{S} - \mathbf{x}_k^T Q \mathbf{S} = -\mathbf{S}^T \mathbf{S}.$$

Therefore,

$$g(\eta_k) = \frac{1}{2}(\mathbf{S}^T Q \mathbf{S})\eta_k^2 - (\mathbf{S}^T \mathbf{S})\eta_k + C,$$

where

$$C = \frac{1}{2}\mathbf{x}_k^T Q \mathbf{x}_k - \mathbf{b}^T \mathbf{x}_k.$$

$g'(\eta_k) = 2A\eta_k + B$

Thus,

$$\boxed{g(\eta_k) = A\eta_k^2 + B\eta_k + C,}$$

with

$$\underline{A = \frac{1}{2}\mathbf{S}^T Q \mathbf{S}}, \quad \underline{B = -\mathbf{S}^T \mathbf{S}}.$$

$2A\eta_k + B = 0$

$\eta_k = \frac{-B}{2A}$

$$S = \nabla f(x) = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$$

$$H := Q = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

Optimal Step Size

We minimize $g(\eta_k)$ with respect to η_k:

$$\frac{dg}{d\eta_k} = 2A\eta_k + B = 0.$$

Solving for η_k,

$$\boxed{\eta_k = -\frac{B}{2A} = \frac{\mathbf{S}^T\mathbf{S}}{\mathbf{S}^T Q \mathbf{S}}.}$$

Since $\mathbf{S} = \nabla f(\mathbf{x}_k)$,

$$\eta_k^* = \frac{[\nabla f(\mathbf{x}_k)]^T \nabla f(\mathbf{x}_k)}{[\nabla f(\mathbf{x}_k)]^T Q \nabla f(\mathbf{x}_k)} = \frac{\mathbf{S}^T\mathbf{S}}{\mathbf{S}^T Q \mathbf{S}}.$$

Here,

$$\mathbf{S} = \nabla f(\mathbf{x}_k) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}, \quad Q = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

is the Hessian matrix.

This is the exact optimal step size (exact line search) for quadratic functions in the method of steepest descent.