Lecture-9

Conjugate Direction Algorithm
(Solution of Linear System or
Minimization of A Quadratic
Function)

Conjugate Gradient

- Linear conjugate gradient: for solving linear systems $Ax=b$ with PD matrix, A.
  - Exact solution in *n steps* (Hestenes & Stiefel, 1950s)
  - Approximate solution in fewer than *n steps*
- Non-linear conjugate gradient: for solving large-scale non-linear optimization problems.
  - Fletcher and Reeves, 1964
  - Polk-Ribiere, 1969

Conjugate Gradient

A is symmetric PD. (1)

Or minimize the following function:

(2)

$r(x)$ is the residual

$S = \{p_0, p_1, \ldots, p_{n-1}\}$ The set S is conjugate wrt A if

$$p_i^T A p_j = 0 \qquad \forall i \neq j$$

Linear Independence

S is linearly independent

if $s_0 p_0 + s_1 p_1 + \ldots s_{n-1} p_{n-1} = 0$

then $s_0 = s_1 = s_2 = \ldots s_{n-1} = 0$

Conjugate set is also linearly independent.

$$p_i^T A p_j = 0 \qquad \forall i \neq j$$

Conjugate Direction Method

$$x_{k+1} = x_k + \alpha_k p_k \qquad \text{Line search}$$

$$p_i^T A p_j = 0 \qquad \forall i \neq j$$

$$f(x) = \frac{1}{2} x^T A x - b^T x \qquad \text{1D minimizer of a quadratic function}$$

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

Convergence Rate of Steepest Descent

$$\frac{d}{d\alpha} f(x_k - \alpha g_k) = \frac{d}{d\alpha}\left(\frac{1}{2}(x_k - \alpha g_k)^T Q (x_k - \alpha g_k) - b^T (x_k - \alpha g_k)\right) = 0$$

$$= -(x_k - \alpha g_k)^T Q g_k + b^T g_k = 0$$

$$-x_k^T Q g_k + \alpha g_k^T Q g_k + b^T g_k = 0$$

$$\alpha g_k^T Q g_k = x_k^T Q g_k - b^T g_k$$

$$\alpha = \frac{x_k^T Q g_k - b^T g_k}{g_k^T Q g_k}$$

$$\alpha = \frac{(x_k^T Q - b^T) g_k}{g_k Q g_k} \qquad \nabla f(x) = Qx - b$$

From Lecture-5

$$\alpha = \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k}$$

$$x_{k+1} = x_k - \alpha_k \nabla f_k \qquad x_{k+1} = x_k - \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k} \nabla f_k$$

Conjugate Direction Method

$$\alpha = \frac{x_k^T Q g_k - b^T g_k}{g_k^T Q g_k}$$

$$\alpha = \frac{(x_k^T A - b^T)(-p_k)}{(-p_k) A (-p_k)}$$

$$\alpha_k = -\frac{\nabla f_k^T p_k}{p_k^T A p_k} \qquad \nabla f(x) = Ax - b = r(x)$$

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k} \qquad p_i^T A p_j = 0 \qquad \forall i \neq j$$

Theorem 5.1

For any x^0 the sequence $\{x_k\}$ generated by the conjugate direction algorithm, converges to the solution x^* of the linear system in at most n steps.

- Sequence $\{x_k\}$
- Linearly independent vectors
- Conjugate vectors

Proof

$$x_{k+1} = x_k + \alpha_k p_k \qquad \alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

$$x_k = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \ldots + \alpha_{k-1} p_{k-1}$$

$$x_k - x_0 = \alpha_0 p_0 + \alpha_1 p_1 + \ldots + \alpha_{k-1} p_{k-1}$$

Proof

S is linearly independent

Therefore:

$$x^* - x_0 = \sigma_0 p_0 + \sigma_1 p_1 + \ldots \sigma_{n-1} p_{n-1}$$

$$p_k^T A(x^* - x_0) = p_k^T A(\sigma_0 p_0 + \sigma_1 p_1 + \ldots \sigma_{n-1} p_{n-1})$$

$$p_k^T A(x^* - x_0) = (0 + 0 + \ldots + \sigma_k p_k^T A p_k + \ldots + 0)$$ conjugate

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k}$$

Proof

$$x_{k+1} = x_k + \alpha_k p_k \qquad \alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

$$x_k = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \ldots + \alpha_{k-1} p_{k-1}$$

$$x_k - x_0 = \alpha_0 p_0 + \alpha_1 p_1 + \ldots + \alpha_{k-1} p_{k-1}$$

$$p_k^T A(x_k - x_0) = 0$$

$$p_k^T A x_k = p_k^T A x_0$$

$$p_k^T A(x^* - x_0) = p_k^T A(x^* - x_k) = p_k^T (b - A x_k) = -p_k^T r_k$$

$$p_k^T A(x^* - x_0) = -p_k^T r_k$$

Proof

$$p_k^T A(x^* - x_0) = -p_k^T r_k$$

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k} \qquad \alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

Therefore:

$$\sigma_k = \alpha_k$$

QED

Interpretation of Theorem 5.1

If A is a diagonal matrix, then we can minimize (1-D) the function along coordinate axes in n iterations.

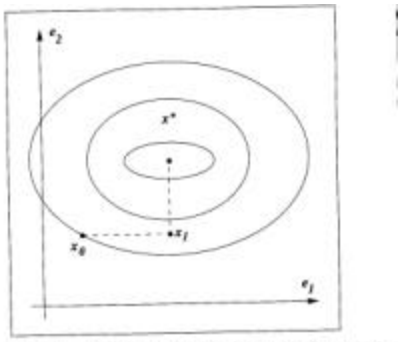
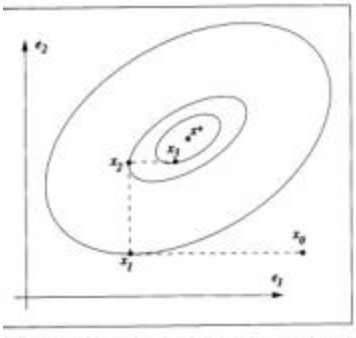


Figure 5.1 Successive minimizations along the coordinate directions minimizer of a quadratic with a diagonal Hessian in n iterations.

Interpretation of Theorem 5.1

If A is not a diagonal matrix, then we can not minimize the function along coordinate axes in n iterations.



igure 5.2 Successive minimization along coordinate axes does not find the solution ı n iterations, for a general convex quadratic.

Transformed Problem

Let

$$\hat{x} = S^{-1}x \qquad \text{where} \qquad S = [p_0, p_1, \ldots, p_{n-1}]$$

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

$$\hat{f}(\hat{x}) = f(S\hat{x}) = \frac{1}{2}\hat{x}^T (S^T AS)\hat{x} - (S^T b)^T \hat{x}$$

By conjugacy S^TAS is a diagonal matrix.

Now we can minimize along coordinate directions in transformed space.

However, each coordinate direction in transformed space correspond to the conjugate direction in the original space due to $\hat{x} = S^{-1}x$

Therefore, we conclude the conjugate direction algorithm converges in *n* steps.

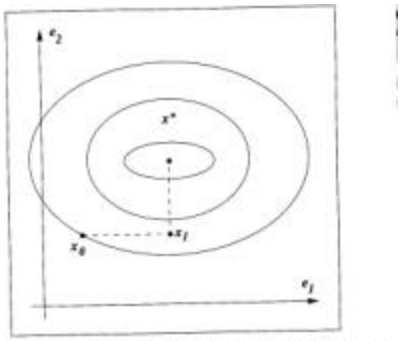


Figure 5.1 Successive minimizations along the coordinate directions minimizer of a quadratic with a diagonal Hessian in *n* iterations.

When Hessian is diagonal, each coordinate minimization correctly determines of the components of the solution x^*. Therefore, after k 1-D minimizations, the quadratic has been minimized on the subspace spanned by $e_1, e_2, \ldots, e_k$.

Theorem 5.2

Let x_0 be any starting point and suppose that the sequence $\{x_k\}$ is generated by the conjugate direction algorithm. Then

$$r_k^T p_i = 0 \quad \text{for} \quad i = 0, \ldots, k-1$$

and x_k is minimizer of $f(x) = \frac{1}{2} x^T A x - b^T x$ over the set

$$\left\{ x \mid x = x_0 + span \{ p_0, \ldots, p_{k-1} \} \right\} \qquad (3)$$

Proof

First show that a point $\tilde{x}$ minimizes f over the set (3) if and only if

$$r(\tilde{x})^T p_i = 0 \quad \text{for} \quad i = 0, \ldots, k-1$$

$$\left\{ x \mid x = x_0 + span \{ p_0, \ldots, p_{k-1} \} \right\}$$

Where

Let $h(\sigma) = f(x_0 + \sigma_0 p_0 + \ldots + \sigma_{k-1} p_{k-1})$

$\sigma = (\sigma_0, \sigma_1, \ldots \sigma_{k-1})$

Since $h(\sigma)$ is strictly convex quadratic, it has a unique minimizer:

$$\frac{\partial h}{\partial \sigma_i} = 0, \qquad i = 0, \ldots, k-1$$

$$\nabla f(x_0 + \sigma_0^* p_0 + \ldots + \sigma^*_{k-1} p_{k-1})^T p_i = 0 \qquad i = 0, \ldots, k-1$$ Chain rule

$r(x)$ is the residual

$$r(\tilde{x})^T p_i = 0 \qquad i = 0, \ldots, k-1$$

Proof

$$\nabla f(x) = Ax - b = r(x) \qquad x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = r_k + \alpha_k A p_k$$

$$r_k = r_{k-1} + \alpha_{k-1} A p_{k-1} \qquad \text{(A)}$$

Use induction:

True for k=1

From (A)

$$r_1 = r_0 + \alpha_0 A p_0$$

$$r_1^T p_0 = (r_0 + \alpha_0 A p_0)^T p_0$$

$$r_1^T p_0 = r_0^T p_0 + \alpha_0 p_0^T A p_0$$

$$r_1^T p_0 = 0$$

Because

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

Proof

$$r_k = r_{k-1} + \alpha_{k-1} A p_{k-1} \qquad \text{(A)}$$

Assume true for k-1

From (A)

$$p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0$$

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

Definition

And

$$p_i^T r_k = p_i^T r_{k-1} + \alpha_{k-1} p_i^T A p_{k-1} = 0 \qquad i = 0, \ldots, k-2$$

Conjugacy

$r_k^T p_i = 0$ for $i = 0, \ldots, k-2$

induction

Therefore $r_k^T p_i = 0$ for $i = 0, \ldots, k-1$ QED

How do we select conjugate directions

- Eigenvalues of *A* are mutually orthogonal and conjugate wrt to *A*.
- Gram-Schmidt process to produce conjugate directions instead of orthogonal vectors.

Basic Properties of the CG

Each direction is chosen to be a linear combination of the steepest descent direction and the previous direction.

$$p_k = -\nabla f_k + \beta_k p_{k-1}$$

$$p_k = -r_k + \beta_k p_{k-1}$$

$$p_{k-1}^T A p_k = -r_k p_{k-1}^T A + \beta_k p_{k-1}^T A p_{k-1}$$

$$\beta_k = \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

Algorithm 5.1

Given x_0;

set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$

While $r_k \neq 0$

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

$$x_{k+1} \leftarrow x_k + a_k p_k;$$

$$r_{k+1} \leftarrow Ax_{k+1} - b;$$

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

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

$$k \leftarrow k + 1;$$

end (*while*)

p_0 is steepest descent