CONJUGATE GRADIENT METHODS

CG methods

solving large linear systems of equations

solve nonlinear optimization problems

- ✓performance of the linear conjugate gradient method
- ✓Preconditioning
- ✓no matrix storage and are faster than the steepest descent method

Linear CG method

$$Ax = b$$

A is an $n \times n$ symmetric positive definite matrix

Linear CG method

$$Ax = b$$

A is an $n \times n$ symmetric positive definite matrix

$$\min \phi(x) \stackrel{\text{def}}{=} \frac{1}{2}x^T Ax - b^T x$$

$$\nabla\phi(x) = Ax - b \stackrel{\text{def}}{=} r(x)$$

at $x = x_k$

$$r_k = Ax_k - b$$

conjugate direction method

set of nonzero vectors $\{p_0, p_1, \ldots, p_l\}$ is said to be *conjugate* with respect to the symmetric positive definite matrix A if

$$p_i^T A p_j = 0, \qquad \text{for all } i \neq j$$

linearly independent

conjugate direction method

set of nonzero vectors $\{p_0, p_1, \ldots, p_l\}$ is said to be *conjugate* with respect to the symmetric positive definite matrix A if

$$p_i^T A p_j = 0, \qquad \text{for all } i \neq j$$

linearly independent

Given a starting point $x_0 \in \mathbb{R}^n$ and a set of conjugate directions $\{p_0, p_1, \ldots, p_{n-1}\}$

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

α_k is the one-dimensional minimizer of the quadratic function $\phi(\cdot)$ along $x_k + \alpha p_k$

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

conjugate direction method

For any $x_0 \in \mathbb{R}^n$ *the sequence* $\{x_k\}$ *generated by the conjugate direction algorithm converges to the solution* x^* *of the linear system in at most* n *steps.*

conjugate direction method

For any $x_0 \in \mathbb{R}^n$ the sequence $\{x_k\}$ generated by the conjugate direction algorithm converges to the solution x^* of the linear system in at most n steps.

PROOF:

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

conjugate direction method

For any $x_0 \in \mathbb{R}^n$ the sequence $\{x_k\}$ generated by the conjugate direction algorithm converges to the solution x^* of the linear system in at most n steps.

PROOF:

$p_k^T A$

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

$$\Downarrow$$

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

α_k

conjugate direction method

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

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

conjugate direction method

$p_k^T A$

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

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

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

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

conjugate direction method

$p_k^T A$

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

$$\Downarrow$$

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

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

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

$$\boxed{\sigma_k = \alpha_k}$$

conjugate direction method

$p_k^T A$

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

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

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

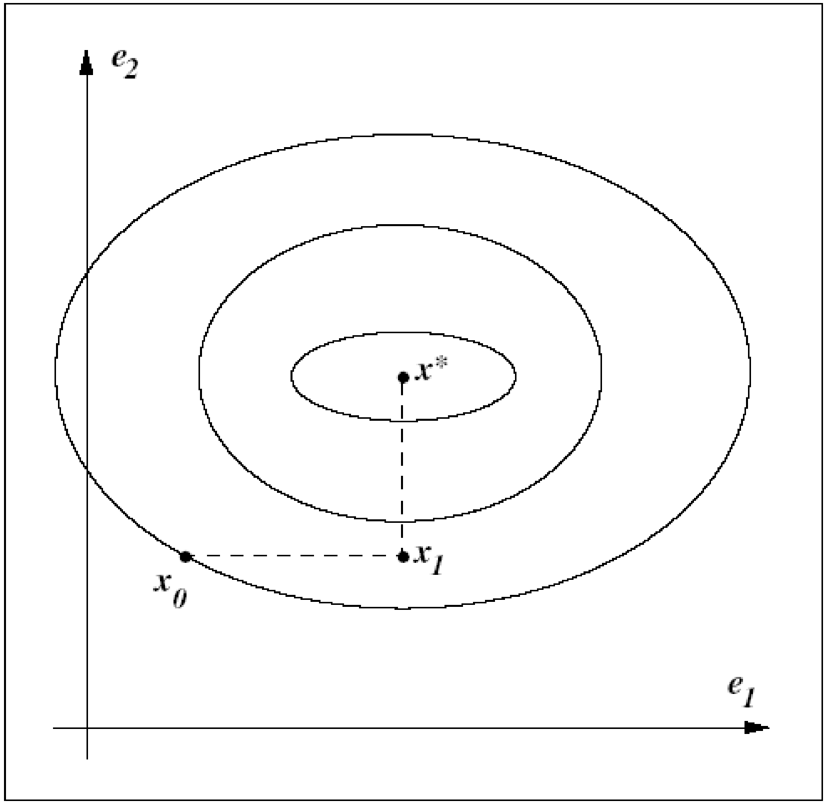
$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k} \qquad \alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

$$\boxed{\sigma_k = \alpha_k}$$

Linear CG method

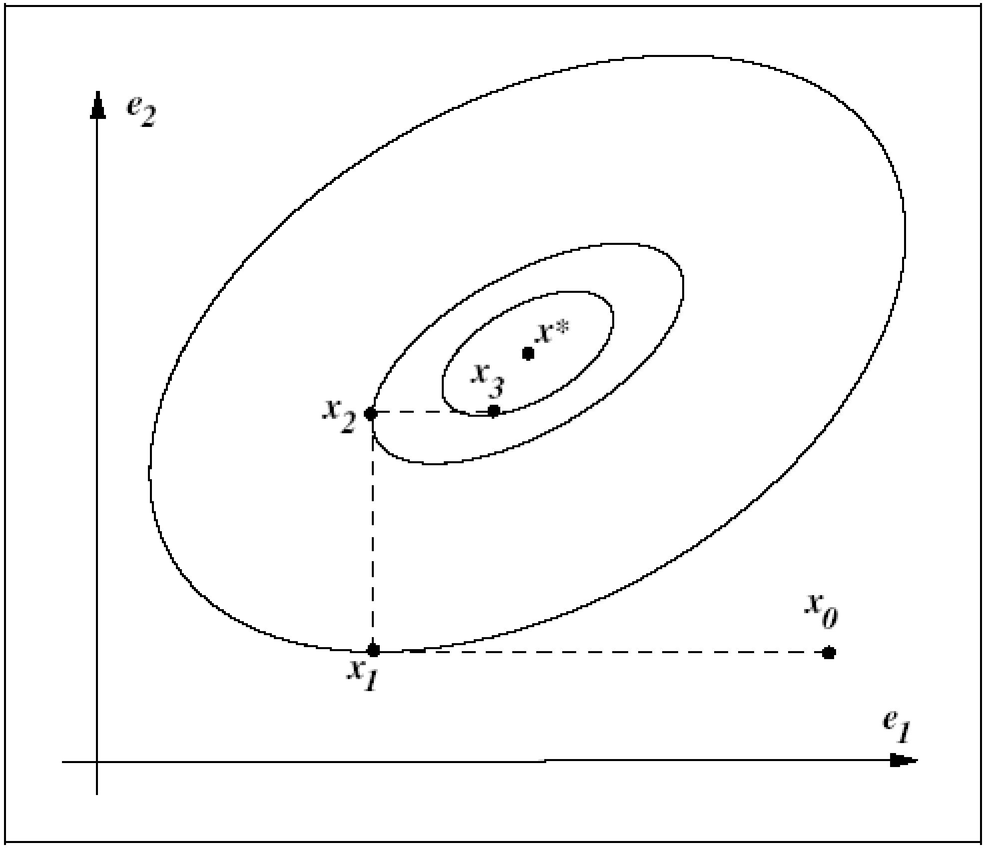
$$\phi(x) \stackrel{\text{def}}{=} \frac{1}{2}x^T A x - b^T x$$

Diagonal A



Linear CG method

Not diagonal A



Transform the problem to make *A diagonal and then minimize along the coordinate directions*

Linear CG method

$$\hat{x} = S^{-1}x$$

where S is the $n \times n$ matrix defined by

$$S = [p_0 \; p_1 \; \cdots \; p_{n-1}],$$

where $\{p_0, p_2, \ldots, p_{n-1}\}$ is the set of conjugate directions with respect to A

$$\hat{\phi}(\hat{x}) \stackrel{\text{def}}{=} \phi(S\hat{x}) = \tfrac{1}{2}\hat{x}^T(S^T A S)\hat{x} - (S^T b)^T \hat{x}$$

diagonal

performing *n one-dimensional minimizations along the coordinate directions*

Linear CG method

✓the *ith coordinate direction in ˆx-space corresponds* to the direction p_i *in x-space*

✓the coordinate search strategy applied to $\hat{\varphi}$ *is* equivalent to the conjugate direction algorithm

Theorem (Expanding Subspace Minimization).

Let $x_0 \in \mathbb{R}^n$ be any starting point and suppose that the sequence $\{x_k\}$ is generated by the conjugate direction algorithm Then

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

$$r_{k+1}^T A p_i = 0, \qquad \text{for } i = 0, 1, \ldots, k-1,$$

current residual r_k *is orthogonal to all previous search directions*

Linear CG method

✓compute a new vector p_k *by* using only the previous vector p_{k-1}
✓method requires little storage and computation

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

$p_{k-1}^T A$

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

Linear CG method

Algorithm (CG–Preliminary Version).

Given x_0;

Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$;

while $r_k \neq 0$

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

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

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

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

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

$$k \leftarrow k + 1;$$

end (while)

Linear CG method

Algorithm (CG).

Given x_0;

Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$;

while $r_k \neq 0$

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

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

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

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$

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

$$k \leftarrow k + 1;$$

end (while)

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

$$\alpha_k \leftarrow -\frac{r_k^T p_k}{p_k^T A p_k} \qquad p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$$

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

$$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{}$$

Linear CG method

If A has eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, we have that

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

NLCG

adapt the approach to minimize general convex functions

step length α_k : *a line search an approximate minimum of the* nonlinear function f *along* p_k

the residual r replaced by the gradient of the nonlinear objective f

NLCG

Algorithm (FR).

Given x_0;

Evaluate $\nabla f_0 = \nabla f(x_0)$;

Set $p_0 \leftarrow -\nabla f_0$, $k \leftarrow 0$;

while $\nabla f_k \neq 0$

Compute α_k and set $x_{k+1} = x_k + \alpha_k p_k$;

Evaluate ∇f_{k+1};

$$\beta_{k+1}^{FR} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k};$$

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{FR} p_k;$$

$$k \leftarrow k + 1;$$

end (while)