# Conjugate Gradient Methods

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

February 27, 2024

# Conjugate Gradient Methods

- They are among the most useful techniques for solving large linear systems of equations.
- They can be adapted to solve non-linear optimisation problems.
- The linear conjugate gradient method is an alternative to Gaussian elimination that is well suited for solving large scale problems.
- Linear conjugate gradient method was proposed by Hestenes and Stiefel in 1950.
- A Key feature of these algorithms is, they require no matrix storage and are faster than the steepest descent method.

# Linear Conjugate Gradient Method

The linear conjugate gradient method is an iterative method for solving linear system of equations

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

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

- The above problem of solving a linear system of equations can be equivalently stated as a minimisation problem:

$$\min_x \phi(x) := \frac{1}{2}x^T Ax - b^T x \tag{2}$$

**Remark**

Both (1) and (2) have the same unique solution.

# Linear Conjugate Gradient Method

- The equivalence of both the problems allows us to view conjugate gradient methods either as an algorithm for solving linear systems or as a technique for minimising convex quadratic functions.
- The residual $r$ of the linear system (1) is defined as:

$$r(x) := Ax - b \tag{3}$$

- Note that the gradient of $\phi$ is:

$$\nabla\phi = r(x) \tag{4}$$

- In particular at $x = x_k$

$$r_k = r(x_k) = Ax_k - b$$

# Conjugate Direction Methods

- Generates a set of vectors with a property known as conjugacy.
- The vectors are manufactured, in a very economical fashion.

**Conjugacy**

A set of non-zero vectors $\{p_0, p_1, \cdots, p_l\}$ is said to be conjugate with respect to the symmetric, positive definite matrix $A$ if

$$p_i^T A p_j = 0 \qquad \text{for,} \quad i \neq j \tag{5}$$

- Any set of vectors satisfying this property is also linearly independent.

# Conjugate Direction Methods

- The objective function $\phi(.)$ can be minimised in $n$ steps by successively minimising it along the individual directions in a conjugate set.
- Let $x_0 \in \mathbb{R}^n$ and a set of conjugate directions $\{p_0, p_1, \cdots, p_{n-1}\}$, the sequence of iterates is generated as:

$$x_{k+1} = x_k + \alpha_k p_k \tag{6}$$

- Where $\alpha_k$ is the one-dimensional minimiser of the quadratic function $\phi(.)$ along $x_k + \alpha p_k$, and can be obtained explicitly as:

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k} \tag{7}$$

# Convergence of Conjugate Direction Methods

**Theorem**

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 (1) in at most $n$ steps.

## Sketch of the Proof:

- Since the directions $\{p_i\}$ are linearly independent, they must span the whole space $\mathbb{R}^n$.
- Therefore, the difference between $x_0$ and the solution $x^*$ can be written in the following way:

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

for some choice of scalars $\sigma_k$.

# Convergence of Conjugate Direction Methods

- By premultiplying this expression by $p_k^T A$ and using the conjugacy property, we obtain:

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k} \tag{8}$$

- We now establish the result by showing that these coefficients $\sigma_k$ coincide with the step lengths $\alpha_k$.
- If $x_k$ is generated by the conjugate direction algorithm, we have

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

- By premultiplying this expression by $p_k^T A$ and using the conjugacy property, we have that

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

# Convergence of Conjugate Direction Methods

- Therefore,

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

- By comparing the above relation with (7) and (8), we find that $\sigma_k = \alpha_k$, giving the result.

**Remark**

If the matrix $A$ is diagonal, the contours of the function $\phi(.)$ are ellipses whose axes are aligned with the co-ordinate directions $e_1, e_2, \cdots, e_n$.

# Convergence of Conjugate Direction Methods

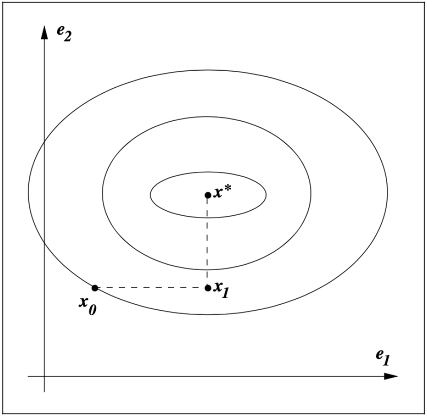
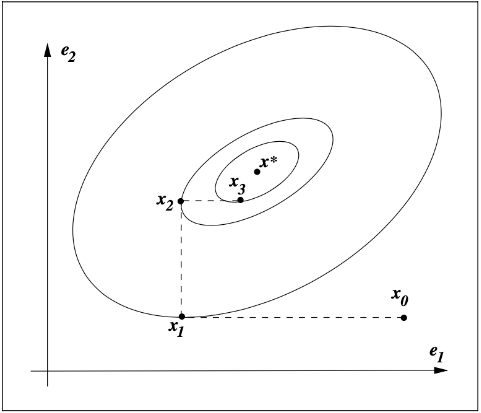



Figure: Successive minimizations along the coordinate directions find the minimizer of a quadratic with a diagonal Hessian in n iterations.

# Convergence of Conjugate Direction Methods

- Find the minimiser of this function by performing one-dimensional minimisations along the coordinate directions $e_1, e_2, \cdots, e_n$ in turn.
- When $A$ is not diagonal, its contours are still elliptical, but they are usually no longer aligned with the coordinate directions.
- Successive minimization along these directions in turn no longer leads to the solution in n iterations.

# Convergence of Conjugate Direction Methods



# Convergence of Conjugate Direction Methods

- The nice behaviour of Figure 1 can be obtained if we transform the problem to make $A$ diagonal and then minimize along the coordinate directions.
- We transform the problem by defining new variables $\hat{x}$ as:

$$\hat{x} = S^{-1}x \tag{9}$$

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

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

- The quadratic $\phi$ defined by (2) now becomes:

$$\hat{\phi}(\hat{x}) := \phi(S\hat{x}) = \frac{1}{2}\hat{x}^T(S^TAS)\hat{x} - (S^Tb)^T\hat{x}. \tag{10}$$

# Convergence of Conjugate Direction Methods

- By conjugacy property (5), the matrix $S^T A S$ is diagonal.
- The minimising value of $\hat{\phi}$ can be found by performing $n$ one-dimensional minimisations along the coordinate directions of $\hat{x}$.
- The coordinate search strategy applied to $\hat{\phi}$ is equivalent to the conjugate direction algorithm (6)-(7).
- The conjugate direction algorithm terminates in at most n steps.

# Convergence of Conjugate Direction Methods

- When the Hessian matrix is diagonal, each coordinate minimisation correctly determines one of the components of the solution $x^*$ .
- After $k$ one-dimensional minimisations, the quadratic has been minimized on the subspace spanned by $e_1, e_2, \ldots, e_k$.
- The following theorem proves this result for the general case in which the Hessian of the quadratic is not necessarily diagonal.

# Expanding Subspace Minimization

$$r_{k+1} = r_k + \alpha_k A p_k \tag{11}$$

**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 (6)-(7). Then

$$r_k^T p_i = 0, \qquad \text{for} \quad i = 0, 1, \ldots, k-1, \tag{12}$$

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

$$\{x | x = x_0 + \text{span}\{p_0, p_1, \cdots, p_{k-1}\}\} \tag{13}$$

- That is, the method minimizes $\phi$ piece-wise, one direction at a time.
- The current residual $r_k$ is orthogonal to all previous search direction.

## How to obtain conjugate directions??

- The discussion applies to a conjugate direction method (6)-(7) based on any choice of the conjugate direction set $\{p_0, p_1, \ldots, p_{n-1}\}$.
- There are many ways to choose the set of conjugate directions.
- The eigenvectors $\{v_1, v_2, \ldots, v_n\}$ of $A$ are mutually orthogonal as well as conjugate with respect to $A$.
- For large-scale applications computation of the complete set of eigenvectors requires an excessive amount of computation.
- To modify the Gram–Schmidt orthogonalisation process to produce a set of conjugate directions rather than a set of orthogonal directions.
- The Gram–Schmidt approach is also expensive, since it requires us to store the entire direction set.