Conjugate Gradient Methods

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

February 26, 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:

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

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 ϕ 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 α_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 σ_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 σ_k coincide with the step lengths α_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

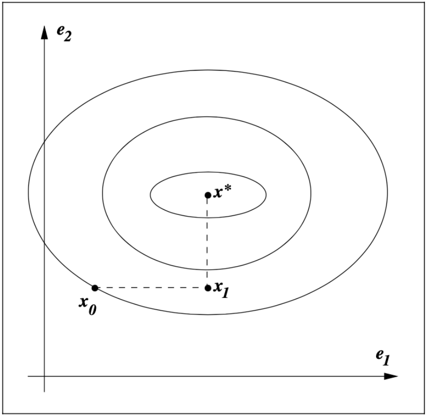

Figure: Successive minimizations along the coordinate directions find the minimizer of a quadratic with a diagonal Hessian in n iterations.

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