Conjugate Gradient - 1

1 Co-ordinate Descent Method

This method attempts to minimize the objective function, by minimizing the function along one (axis) direction at a time. For a function of the form,

$$f(X) = 4X_1^2 + X_2^2$$

this method tries to minimize the function along X_1 first and then along X_2 (i.e. along the axes), eventually reaching the minima (0,0) in exactly two iterations. However for a function of the form,

$$g(X) = 4X_1^2 + X_2^2 - 2X_1X_2$$

the same method takes multiple iterations to converge to the minimum.

- The reason for multiple iterations for $g(X)$ to converge can be attributed to the fact that the elliptically shaped contours of $g(X)$ do not have their axes lined up along co-ordinate axes, X_1 and X_2.
- Whenever the Hessian(H) is a diagonal matrix (as in case of $f(X)$, not for $g(X)$), co-ordinate descent method converges to the solution in exactly N iterations (N being the dimension of X).

It is also necessary to note that this method would fail for non-smooth functions as it might get stuck at a non-stationary point.
However, the method attracts substantial interest due it's inherent advantages of high computational efficiency, scalability, quicker convergence and ease of implementation.

2 Linear Conjugate Gradient Method

The Linear Conjugate Gradient Method is an iterative method to solve the system of equations $Ax = b$, (A is symmetric and positive definite) which can be equivalently formulated as a minimization problem of the quadratic form:

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

Hence, the method can well be extended to solve optimization problems.

3 Conjugate Gradient Method

Let us consider a minimization problem of the type

$$\phi(x) = \frac{1}{2}x^T Hx + c^T x, \tag{1}$$

where H is assumed to be symmetric and positive definite. Let $d^0, d^1, ..., d^{n-1}$ be n such linearly independent vectors (directions) and x be the iterative solution for the system. We have,

$$x = x^0 + \sum_{i=0}^{n-1} \alpha^i d^i, \tag{2}$$

Upon back substitution to the original problem, we get

$$\phi(\alpha) = \frac{1}{2}\alpha^T D^T HD\alpha + c^T D\alpha, \tag{3}$$

where D = $[d^0 d^1 .....d^{n-1}]$ and $\alpha = [\alpha^0 \alpha^1 .....\alpha^{n-1}]$.
Let $H^{'} = D^T HD$, then $H^{'}$ is a matrix with elements as $d^{iT} Hd^j$ $\forall$ i,j.

- For $H^{'}$ to be a diagonal matrix, $d^{iT} Hd^j = 0$ $\forall i\neq$ j. This condition ensures that $\phi(\alpha)$ is minimized.

This reduces to condition of H-orthogonality, i.e. the vectors (d^i) must be orthogonal to each other. Such set of vectors (d^i), which satisfy the relation $d^{iT} Hd^j = 0$ $\forall i\neq$ j. are referred to as H-conjugate vectors.

Proof of Existence:Let H be a symmetric matrix with n orthogonal eigen-vectors, if v_1 v_2 be two such distinct eigen vectors, then $v_1^T v_2 = 0$ and $Hv_1 = \lambda_1 v_1$, a multiplication by v_2^T results in $v_2^T Hv_1 = 0$. This suffices the proof of existence of H-conjugate vectors, for symmetric Hessian.

Gradient, Error and Residuals

- Residual(r_k) is defined as $-\nabla\phi(x) = -c - Hx$ (from 1).
- Error(e_k) = $x_k - x^*$, x^* is the true solution.

Let x_k be an iterative solution to a system, with α_k and d_k being the step length and direction at the k^{th} iterate, then $x_{k+1} = x_k + \alpha_k d_k$. Hence, $f(x_{k+1}) = f(x_k + \alpha_k d_k)$. Differentiating w.r.t α_k, we get $\nabla \mathrm{f}(\mathrm{x}_{k+1})^T d_k = 0$ (since, $f(x_{k+1})$ is some constant), equivalently,

$$g_{k+1}^T d_k = 0,$$

where $g_{k+1} = \nabla \mathrm{f}(\mathrm{x}_{k+1})$. Also, $g_k = He_k$. Hence, $(He_{k+1})^T d_k = 0$, which is equivalent to

$$e_{k+1}^T H^T d_k = 0, \tag{4}$$

which means error in $(k+1)^{th}$ iteration is H-conjugate to d^k $\forall$ k $\in$ [0,n-1].