3.3.3. CONJUGATE GRADIENT METHODS

The conjugate gradient methods are based on the quadratic model of the objective function. Search directions are formed using gradient information so that the directions are conjugate when the objective function is quadratic. Because no matrix operations are needed, the method is suitable especially for large scale problems (= problems with a large number of variables).

Conjugate gradient method of Fletcher and Reeves

A quadratic objective function:

Minimize $f(\mathbf{x}) = \frac{1}{2}\, \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x}$, where the Hessian Q is positive definite.

Given a starting point $\mathbf{x_0}$, the first search direction is the direction of steepest descent:

$$\mathbf{p_0} = -\nabla f(\mathbf{x_0}) = -\mathbf{Q}\mathbf{x_0} - \mathbf{c}$$

Line search: minimize $F(\alpha) = f(\mathbf{x_0} + \alpha\mathbf{p_0})$

The optimal step size α_0 can be calculated from the stationarity condition that $F'(\alpha_0) = 0$, i.e. the directional derivative in the direction $\mathbf{p_0}$ at $\mathbf{x_0}+\alpha_0\mathbf{p_0}$ is 0:

$$F'(\alpha_0) = \mathbf{p_0}^T\nabla f(\mathbf{x_0}+\alpha_0\mathbf{p_0}) = 0 \qquad \Leftrightarrow$$

$$\mathbf{p_0}^T[Q(\mathbf{x_0}+\alpha_0\mathbf{p_0}) + \mathbf{c}] = 0 \qquad \Leftrightarrow$$

$$\alpha_0 = \frac{-\boldsymbol{p_0}^T\nabla f(\boldsymbol{x_0})}{\boldsymbol{p_0}^T\boldsymbol{Q}\boldsymbol{p_0}} = \frac{\|\nabla f(\boldsymbol{x_0})\|^2}{\boldsymbol{p_0}^T\boldsymbol{Q}\boldsymbol{p_0}}$$

Set $\mathbf{x_1} := \mathbf{x_0} + \alpha_0\mathbf{p_0}$

The next direction is

$$\mathbf{p_1} = -\nabla f(\mathbf{x_1}) + \beta_1\mathbf{p_0}$$

where the coefficient β_1 can be chosen so that the directions $\mathbf{p_0}$ and $\mathbf{p_1}$ are Q-conjugate, i.e. by solving β_1 from the condition

$$\mathbf{p_0}^T\mathbf{Q}\mathbf{p_1} = 0.$$

After some calculations we get

$$\beta_1 = \frac{\nabla f(\boldsymbol{x_1})^T\nabla f(\boldsymbol{x_1})}{\nabla f(\boldsymbol{x_0})^T\nabla f(\boldsymbol{x_0})} = \frac{\|\nabla f(\boldsymbol{x_1})\|^2}{\|\nabla f(\boldsymbol{x_0})\|^2}$$

The step length can again be determined from the condition $\mathbf{p_1}^T\nabla f(\mathbf{x_1}+\alpha_1\mathbf{p_1}) = 0$and set $\mathbf{x_2} := \mathbf{x_1} + \alpha_1\mathbf{p_1}$.

The procedure is continued so that direction $\mathbf{p_k}$ is required to be conjugate with the previous directions $\mathbf{p_0}, \ldots, \mathbf{p_{k-1}}$ (c.f. the Gram-Schmidt orthogonalization). This leads to simple formulas: the search direction at point $\mathbf{x_k}$ is

$\mathbf{p_k} = = -\nabla f(\mathbf{x_k}) + \beta_k \mathbf{p_{k-1}} = -\mathbf{g_k} + \beta_k \mathbf{p_{k-1}}$

where

$$\beta_k = \frac{\|\nabla f(\boldsymbol{x_k})\|^2}{\|\nabla f(\boldsymbol{x_{k-1}})\|^2} = \frac{\|\boldsymbol{g_k}\|^2}{\|\boldsymbol{g_{k-1}}\|^2} \qquad k=1,\ldots,n-1$$

The direction from point $\mathbf{x_k}$ is

$$\boldsymbol{p_k} = -\nabla f(\boldsymbol{x_k}) + \frac{\|\nabla f(\boldsymbol{x_k})\|^2}{\|\nabla f(\boldsymbol{x_{k-1}})\|^2}\, \boldsymbol{p_{k-1}}$$

The optimal step size (for the quadratic function), from the minimization of $f(\mathbf{x_k} + \alpha \mathbf{p_k})$ is

$$\alpha_k = \frac{-\boldsymbol{p_k}^T \nabla f(\boldsymbol{x_k})}{\boldsymbol{p_k}^T \boldsymbol{Q} \boldsymbol{p_k}}$$

It can be shown that directions $\mathbf{p_0}$, $\mathbf{p_1}$, ... , $\mathbf{p_{n-1}}$ are mutually conjugate.

Theoretically, the method reaches the minimum in at most n iterations, if exact arithmetic and exact line search are used. In practice, rounding errors may destroy conjugacy and iteration is continued. Optimality is checked at every iteration.

Generalization to the minimization of a general nonlinear function:

• The step size α_k must be computed with an iterative method because the above formula holds only for quadratic functions.

• A restart is recommended after every cycle of n line searches, using the steepest descent direction: $\mathbf{p_k} := -\nabla f(\mathbf{x_k})$.

• The line search must be performed with sufficient accuracy to ensure the descendent condition.

Quasi-Newton methods are usually more efficient than conjugate gradient methods when the problem dimension is moderate.

<u>Example 3.5.</u> Minimize $f(\mathbf{x}) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$ with the conjugate gradient method of Fletcher and Reeves, starting from the point $\mathbf{x_0} = (0, 0)$.