Dr. J. E. Rayas-Sánchez

THE CONJUGATE GRADIENT METHOD ON QUADRATIC FUNCTIONS

Consider a quadratic objective function u: $\Re^n \rightarrow \Re$ given by

$$u(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c \tag{1}$$

where $\boldsymbol{x} \in \Re^n$ contains the n optimization variables, and $\boldsymbol{Q} \in \Re^{n\times n}$ is a positive definite matrix. We want to solve

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} u(\boldsymbol{x}) \tag{2}$$

The optimal solution $\boldsymbol{x}^*$ can be found by making $\nabla u(\boldsymbol{x}^*) = \boldsymbol{0}$. From (1)

$$\nabla u(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} \tag{3}$$

The conjugate gradient method imposes conjugate search directions, which means that any two search directions $\boldsymbol{d}_j$ and $\boldsymbol{d}_k$ satisfy $\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_k = 0$ for $j \neq k$, and the sequence of directions $\boldsymbol{d}_0, \boldsymbol{d}_1,\ldots, \boldsymbol{d}_k$, with $k \leq n-1$, are linearly independent.

If an exact line search is realized at each iteration, the following unidimensional problem is solved at the i-th iteration

$$\alpha_i = \arg\min_{\alpha} u(\boldsymbol{x}_i + \alpha \boldsymbol{d}_i) = \arg\min_{\alpha} u(\alpha) \tag{4}$$

where $\boldsymbol{x}_i$, $\boldsymbol{d}_i \in \Re^n$ are the current iterate and the current search direction, respectively, and $\alpha_i \in \Re$ is the one-dimensional minimizer at the i-th iteration. Once α_i is found, the next iterate is calculated by

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{d}_i \tag{5}$$

Applying the chain rule on the scalar unidimensional function $u(\alpha)$,

$$\frac{du}{d\alpha} = \nabla u(\boldsymbol{x})^{\mathrm{T}} \frac{d\boldsymbol{x}}{d\alpha} = (\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b})^{\mathrm{T}}\boldsymbol{d}_i = \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{b} \tag{6}$$

Solving problem (4) implies

$$\left.\frac{du}{d\alpha}\right|_{\alpha=\alpha_i} = 0 = \left.(\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{b})\right|_{\alpha=\alpha_i} = \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_i + \alpha_i\boldsymbol{d}_i) + \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{b} = \alpha_i\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i + \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}_i + \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{b} \tag{7}$$

Solving (7) for α_i and using (3),

$$\alpha_i = \frac{-\boldsymbol{d}_i^{\mathrm{T}}(\boldsymbol{Q}\boldsymbol{x}_i + \boldsymbol{b})}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i} = \frac{-\boldsymbol{d}_i^{\mathrm{T}}\nabla u(\boldsymbol{x}_i)}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i} \tag{8}$$

An exact solution to (4) at the i-th iteration is given by (8). Since $\boldsymbol{Q}$ is positive definite, hence $\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i >$ 0, and it is seen from (8) that if $\alpha_i > 0$, then $\boldsymbol{d}_i^{\mathrm{T}}\nabla u(\boldsymbol{x}_i) < 0$, which ensures that $\boldsymbol{d}_i$ is a down-hill direction.

Since the sequence of directions $\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_k$ are Q-conjugate, with $\boldsymbol{Q}$ positive definite, then the linear combination of n search directions $a_0\boldsymbol{d}_0 + a_1\boldsymbol{d}_1 + \ldots + a_{n-1}\boldsymbol{d}_{n-1}$ expands the whole space $\Re^n$. Hence, we can ensure that

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = a_0\boldsymbol{d}_0 + a_1\boldsymbol{d}_1 + \ldots + a_{n-1}\boldsymbol{d}_{n-1} \tag{9}$$

for some set of non-zero real coefficients $\{a_0, a_1, \ldots, a_{n-1}\}$.

On the other hand, from (5) we know that the i-th conjugate gradient iterate is given by

$$\boldsymbol{x}_i = \boldsymbol{x}_0 + \alpha_0\boldsymbol{d}_0 + \alpha_1\boldsymbol{d}_1 + \ldots + \alpha_{i-1}\boldsymbol{d}_{i-1} \tag{10}$$

From (9) and (10), if we prove that $a_k = \alpha_k$ for $k = 0, 1, \ldots n-1$, then the optimal solution must be obtained in n iterations, that is, $\boldsymbol{x}_n = \boldsymbol{x}^*$.

Pre-multiplying (9) by $\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}$, with $i \le n-1$, and considering that all directions are Q-conjugate,

$$\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0) = \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}a_i\boldsymbol{d}_i \tag{11}$$

Solving (11) for a_i,

$$a_i = \frac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0)}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i} \tag{12}$$

Similarly, pre-multiplying (10) by $\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}$, and considering that all directions are Q-conjugate,

$$\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_i - \boldsymbol{x}_0) = 0 \tag{13}$$

Since

$$\boldsymbol{x}_n - \boldsymbol{x}_0 = (\boldsymbol{x}_n - \boldsymbol{x}_i) + (\boldsymbol{x}_i - \boldsymbol{x}_0) \tag{14}$$

Pre-multiplying (14) by $\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}$, and using (11) and (13),

$$\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_n - \boldsymbol{x}_0) = \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_n - \boldsymbol{x}_i) \tag{15}$$

Substituting $\boldsymbol{x}_n = \boldsymbol{x}^*$ in (15), and using (3),

$$\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0) = \boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}_i) = \boldsymbol{d}_i^{\mathrm{T}}[(\nabla u(\boldsymbol{x}^*) - \boldsymbol{b}) - (\nabla u(\boldsymbol{x}_i) - \boldsymbol{b})] = -\boldsymbol{d}_i^{\mathrm{T}}\nabla u(\boldsymbol{x}_i) \tag{16}$$

Substituting (16) in (12),

$$a_i = \frac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}_0)}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i} = \frac{-\boldsymbol{d}_i^{\mathrm{T}}\nabla u(\boldsymbol{x}_i)}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{d}_i} \tag{17}$$

Since (17) is equal to (8), then $a_k = \alpha_k$ for $k = 0, 1, \ldots n-1$, which proves that the conjugate gradient optimization method solves a quadratic objective function in n iterations if an exact line search is realized at each iteration.