Ecole Centrale Paris
Numerical Optimization and Applications
By Tahar Z. Boulmezaoud and Laurent Dumas 2011

Gradient conjugate method

Consider a quadratic function of the form

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c, \tag{1}$$

where A is symmetric definite positive matrix, $\boldsymbol{x}$ and $\boldsymbol{b}$ are column vectors and c a constant. In the sequel, when there is no confusion, we denote by $\mathbb{R}^n$ the space of all the colomun vectors of the form $(x_1, ..., x_n)^T$, $x_i \in \mathbb{R}$ for $1 \leq i \leq n$.

The *Gradient conjugate* method consists to start from an arbitrary $\boldsymbol{x}_0$, and then construct a sequence of directions $(\boldsymbol{d}_k)_{k\geq 0}$ and points $(\boldsymbol{x}_k)_{k\geq 0}$ as follows

- $\boldsymbol{d}_0 = -\nabla f(\boldsymbol{x}_0)$,
- For $k \geq 0$, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_0 + \boldsymbol{y}_k$ where $\boldsymbol{y}_k$ is solution of the optimization problem

$$\min_{\boldsymbol{y}\in W_k} f(\boldsymbol{x}_0 + \boldsymbol{y}) \text{ where } W_k = \text{span}\{\boldsymbol{d}_0, ..., \boldsymbol{d}_k\}. \tag{2}$$

- For $k \geq 0$, $\boldsymbol{d}_k$ is of the form

$$\boldsymbol{d}_{k+1} = -\nabla f(\boldsymbol{x}_{k+1}) + \sum_{i=0}^{k} \mu_i \boldsymbol{d}_i,$$

  where the coefficients $\mu_0, ..., \mu_k$ are chosen such that $\boldsymbol{d}_{k+1}$ are $\boldsymbol{d}_i$ and A-orthogonal (or *conjugate*) for any $i \leq k$, that is

$$\boldsymbol{d}_{k+1}^T A\boldsymbol{d}_i = 0.$$

We set

$$\boldsymbol{r}_k = -\nabla f(\boldsymbol{x}_k) \text{ for } k \geq 0.$$

1. Compute the gradient and the hessian matrix of f.
2. Is-it a convex function?
3. Find the minimum of f on $\mathbb{R}^n$. Is-it easy to compute?
4. Write the optimality conditions for the problem (2).
5. Show that
$$\boldsymbol{r}_{k+1}^T \boldsymbol{d}_i = 0 \text{ for any } i \leq k. \tag{3}$$
6. Deduce that $\boldsymbol{r}_{k+1} \in W_k^{\perp}$ for any $k \geq 0$.
7. Show that $W_k = \text{span}\{\boldsymbol{r}_0, ..., \boldsymbol{r}_k\}$ for any $k \geq 0$.
8. Deduce also that
   (a) if $\boldsymbol{r}_k = 0$ for some $k \geq 0$, then $\boldsymbol{r}_j = 0$ for any $j \geq k+1$,
   (b) there exists a unique integer $m \leq n$ such that $\boldsymbol{x}_m = \boldsymbol{x}$ ($\boldsymbol{r}_m = 0$) and $\boldsymbol{r}_{m-1} \neq 0$ or $m = 0$,
   (c) If $\boldsymbol{r}_k \neq \boldsymbol{0}$ for some $k \geq 0$, then $\dim W_k = k+1$.
9. Show that for $k < m$
$$\boldsymbol{y}_k = \sum_{i=0}^{k} \frac{\boldsymbol{r}_0^T \boldsymbol{d}_i}{\boldsymbol{d}_i^T A \boldsymbol{d}_i} \boldsymbol{d}_i.$$
(write $\boldsymbol{y} = \sum_{i=0}^{k} \eta_i \boldsymbol{d}_i$ and find η_i for each i).
10. Deduce that
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k, \ \lambda_k = \frac{\boldsymbol{r}_0^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T A \boldsymbol{d}_k} = \frac{\boldsymbol{r}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T A \boldsymbol{d}_k}.$$
11. Show that $W_k = \text{span}\{\boldsymbol{r}_0, \ A\boldsymbol{r}_0, ..., A^k \boldsymbol{r}_0\}$ for each $k \leq m$.
12. Show that
$$\boldsymbol{r}_{k+1}^T A \boldsymbol{d}_i = 0 \text{ for each } 0 \leq i < k.$$
13. Deduce that
$$\boldsymbol{d}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{d}_k,$$
with
$$\beta_k = -\frac{\boldsymbol{r}_{k+1}^T A \boldsymbol{d}_k}{\boldsymbol{d}_k A \boldsymbol{d}_k}$$