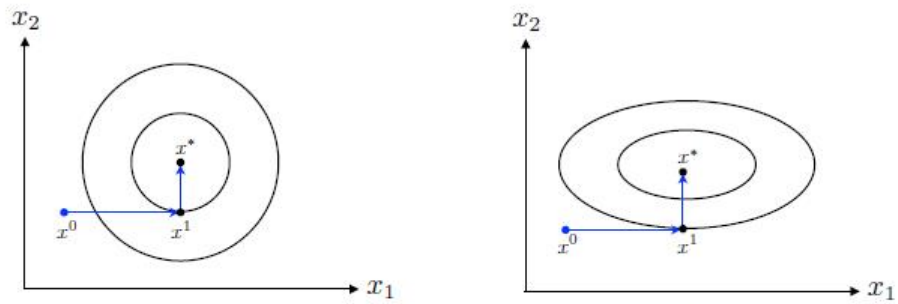
Geometric Ideas of Conjugate Gradient Method

Consider
$$\min\ f(x) = \frac{1}{2}x^TQx - b^Tx, \quad Q \succ 0, \quad x \in \mathbb{R}^n.$$

1. If $Q = I$ and $b = 0$, then $f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}(x_1^2 + \ldots + x_n^2)$, i.e., Q-conjugacy reduces to the usual orthogonality (see Figure(a)).

2. If $Q = \mathrm{diag}\,(a_1, \ldots, a_n)$ and $b = 0$, then $f(x) = \frac{1}{2}(a_1x_1^2 + \ldots + a_nx_n^2)$ (ellipsoid). For example, $f(x) = \frac{a_1}{2}x_1^2 + \frac{a_2}{2}x_2^2$ (see Figure(b)).



(a) (b)

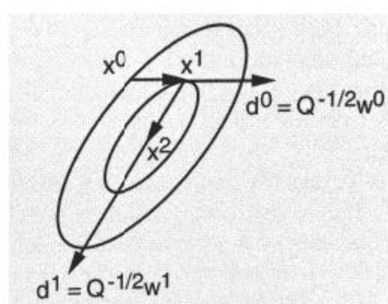
Figure : (a) Minimization along canonical orthogonal directions for $f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$.
(b) Minimization along n orthogonal directions for $f(x) = \frac{a_1}{2}x_1^2 + \frac{a_2}{2}x_2^2$.

3. If $Q \succ 0$ and $b = 0$, then Q is symmetric by definition and by the spectral decomposition $Q = PDP^T = P\sqrt{D}P^TP\sqrt{D}P^T = R^TR$. Then,
$$f(x) = \frac{1}{2}x^TQx = \frac{1}{2}x^TR^TRx = \frac{1}{2}(Rx)^T(Rx) = \frac{1}{2}y^Ty = \frac{1}{2}\|y\|^2$$

   where $y = Rx = Q^{\frac{1}{2}}x$. Therefore,
$$\min\ f(x) = \frac{1}{2}x^TQx \quad \overset{y=Rx}{\Longleftrightarrow} \quad \min\ f(y) = \frac{1}{2}y^Ty = \frac{1}{2}\|y\|^2.$$

   One can think of any conjugate direction method for minimizing $\frac{1}{2}x^TQx$ as a method that minimizes $\frac{1}{2}\|y\|^2$ by successive minimization along n orthogonal directions.

   If $\{w_0, \ldots, w_{n-1}\}$ is a set of orthogonal nonzero vectors, and
$$y_{k+1} = y_k + \alpha_k w_k \quad \text{for} \quad k = 0, 1, \ldots, n-1,$$
   then multiply by $R^{-1} = Q^{-\frac{1}{2}}$, i.e., $x = Q^{-\frac{1}{2}}y$, to get
$$x_{k+1} = x_k + \alpha_k d_k \quad \text{for} \quad k = 0, 1, \ldots, n-1,$$



   where $d_k = Q^{-\frac{1}{2}}w_k$. In this case orthogonality of $\{w_0, \ldots, w_{n-1}\}$ means $\langle w_i, w_j\rangle = 0$ $\forall\, i \neq j$, in which case $\langle d_i, d_j\rangle_Q = d_i^TQd_j = w_i^TQ^{-\frac{1}{2}}QQ^{-\frac{1}{2}}w_j = \langle w_i, w_j\rangle = 0\ \forall\, i \neq j$, so that $\{d_0, \ldots, d_{n-1}\}$ are Q-conjugate.

Proposition: Gram-Schmidt procedure. Let $\{v_1, v_2, \dots, v_k\}$ be a finite or countable set of linearly independent vectors. Construct $\{e_1, \dots, e_k\}$, an orthonormal set, with $\langle e_i, e_j \rangle = \delta_{ij}$. Then let

$$\begin{aligned}
u_1 &= v_1, & e_1 &= \tfrac{u_1}{\|u_1\|} \\
u_2 &= v_2 - \langle v_2, e_1 \rangle e_1, & e_2 &= \tfrac{u_2}{\|u_2\|} \quad \big(\langle e_1, u_2 \rangle = \langle e_1, v_2 \rangle - \langle v_2, e_1 \rangle \langle e_1, e_1 \rangle \big) \\
u_3 &= v_3 - \langle v_3, e_1 \rangle e_1 - \langle v_3, e_2 \rangle e_2, & e_3 &= \tfrac{u_3}{\|u_3\|} \\
&\vdots \\
u_k &= v_k - \textstyle\sum_{i=1}^{k-1} \langle v_k, e_i \rangle e_i, & e_k &= \tfrac{u_k}{\|u_k\|}
\end{aligned}$$

Then $\{e_1, \dots, e_n\}$ are orthonormal.

Proposition: Generating Q-conjugate vectors. Let $\{\xi_0, \xi_1, \dots, \xi_k\}$ be a set of linearly independent vectors, and let $\langle \xi_i, \xi_j \rangle_Q = \xi_i^T Q \xi_j$. Then let

$$\begin{aligned}
d_0 &= \xi_0 \\
d_1 &= \xi_1 - \tfrac{\langle \xi_1, d_0 \rangle_Q}{\langle d_0, d_0 \rangle_Q} d_0 \\
d_2 &= \xi_2 - \tfrac{\langle \xi_2, d_0 \rangle_Q}{\langle d_0, d_0 \rangle_Q} d_0 - \tfrac{\langle \xi_2, d_1 \rangle_Q}{\langle d_1, d_1 \rangle_Q} d_1 \\
&\vdots \\
d_k &= \xi_k - \textstyle\sum_{i=0}^{k-1} \tfrac{\langle \xi_k, d_i \rangle_Q}{\langle d_i, d_i \rangle_Q} d_i.
\end{aligned}$$

Then for $j < k$,

$$\begin{aligned}
d_j^T Q d_k &= \langle d_j, d_k \rangle_Q = \langle d_j, \xi_k \rangle_Q - \sum_{i=0}^{k-1} \frac{\langle \xi_k, d_i \rangle_Q}{\langle d_i, d_i \rangle_Q} \langle d_j, d_i \rangle_Q \\
&= \langle d_j, \xi_k \rangle_Q - \langle \xi_k, d_j \rangle_Q = 0.
\end{aligned}$$

Algorithm : Conjugate gradient method.

Step 0: Given x^0, set $k = 0$ and $d_0 = -g_0 = -\nabla f(x_0)$.

Step 1: For $k = 1, \dots, n-1$, do:

If $\|g_k\| \le \varepsilon$, stop.

Else, $\alpha_k = \arg\min_\alpha f(x^k + \alpha d_k) = -\frac{g_k^T d_k}{d_k^T Q d_k}$.

Step 2: Set $x^{k+1} = x^k + \alpha_k d_k$ where $d_{k+1} = -g_{k+1} + \beta_k d_k$ and $\beta_k = \dfrac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$.

Remark:

The interpretation of conjugate gradient method iteration is that

$$\begin{aligned}
d_{k+1} &= -g_{k+1} &&+ \beta_k d_k \\
&= \text{steepest descent} &&+ \text{preceding direction}.
\end{aligned}$$

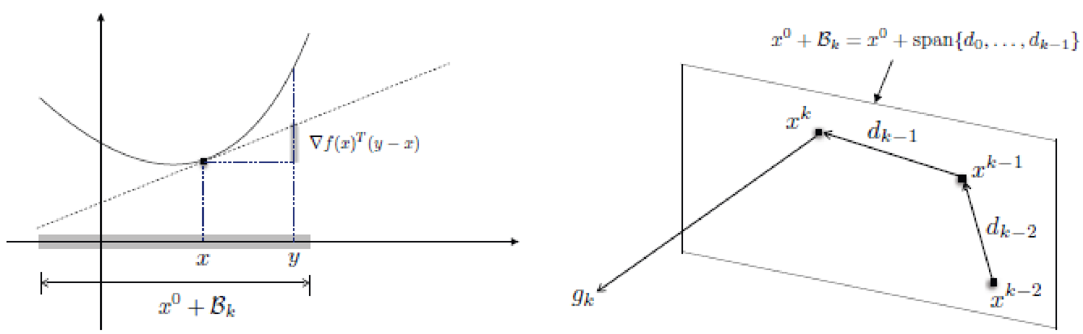
Theorem : Expanding subspace theorem. Let $\{d_i\}_{i=0}^{n-1}$ be a sequence of nonzero Q-conjugate vectors in $\mathbb{R}^n$. Then for any $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}$ generated according to $x^{k+1} = x^k + \alpha_k d_k$, with

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k} \quad \text{where} \quad g_k = Qx^k - b,$$

has the property that x^k minimizes $f(x) = \frac{1}{2}x^T Q x - b^T x$ on the line $x = x^{k-1} + \alpha d_{k-1}$ for $-\infty < \alpha < \infty$, as well as on the linear variety $x^0 + \mathcal{B}_k$, where $\mathcal{B}_k$ denotes the subspace of $\mathbb{R}^n$ spanned by $\{d_0, d_1, \ldots, d_{k-1}\}$.

Remark: One can think of x^k as the solution to

$$\begin{aligned} \min \quad & f(x) = \tfrac{1}{2}x^T Q x - b^T x, \quad Q \succ 0 \\ \text{s.t.} \quad & x = x^0 + \textstyle\sum_{i=0}^{k-1} \alpha^i d_i \ \big(= x^0 + \text{span}\{d_0, \ldots, d_{k-1}\} = x^0 + \mathcal{B}_k \big) \end{aligned}$$



(a) (b) Conjugate Direction Method

Figure: (a) The idea of Expanding Subspace Theorem. (b) Expanding subspaces along the conjugate directions.

Corollary: The gradients g_k for $k = 0, 1, \ldots, n$, satisfy

$$g_k^T \cdot d_i = 0 \quad \text{for} \quad i < k.$$

Corollary: $g_k^T d_k = g_0^T d_k, \ \forall\, k < n.$

Theorem : Validation theorem. The directions $d_0, d_1, \ldots, d_{n-1}$ where $d_{k+1} = -g_{k+1} + \beta_k d_k$ and $\beta_k = \dfrac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$, $\qquad d_0 = -g_0 = -\nabla f(x_0)$.

are Q-conjugate and descent at $x_0, x_1, \ldots, x_{n-1}$, respectively.

$$\uparrow$$

$$x^{k+1} = x^k + \alpha_k d_k$$

$$\uparrow$$

$$\alpha_k = -\tfrac{g_k^T d_k}{d_k^T Q d_k}.$$