Definition - Conjugate Directions. Let Q be an $n \times n$ matrix and $Q \succ 0$. Any two non-zero vectors (directions) $d^{(1)}$, $d^{(2)} \in \mathbb{R}^n$ are said to be *conjugate vectors* or *conjugate directions* with respect to Q, if $(d^{(1)})^T Q d^{(2)} = 0$.

If $Q = I$, conjugacy reduces to orthogonality. More than two vectors are Q-conjugate if they are all mutually Q-conjugate.

Result. Let $\{d^{(0)}, d^{(1)}, \ldots, d^{(k)}\}$ be a set of $k+1$ non-zero vectors which are conjugate with respect to a given positive definite matrix Q. Then the vectors $d^{(0)}, d^{(1)}, \ldots, d^{(k)}$ are linearly independent.

Proof. To prove, we have to show that $\alpha_0 d^{(0)} + \alpha_1 d^{(1)} + \cdots + \alpha_k d^{(k)} = 0$ implies that each $\alpha_j = 0$. Consider:

$$\alpha_0 d^{(0)} + \alpha_1 d^{(1)} + \cdots + \alpha_k d^{(k)} = 0$$

and pre-multiply both sides by $(d^{(j)})^T Q$. Then by conjugacy, it reduces to

$$\alpha_j ((d^{(j)})^T Q d^{(j)}) = 0$$

But, since $Q \succ 0$, $(d^{(j)})^T Q d^{(j)} > 0$. Thus, $\alpha_j = 0$. We can do this for any j, and thus prove all α_j's are 0.

$\square$

For the problem $\min \frac{1}{2} x^T Q x - b^T x$ where $Q \succ 0$ and $Q^T = Q$, note that at x^*, $Qx^* = b$ by setting gradient to 0. And $x^* = Q^{-1} b$. We can solve this system of linear equations to find x^*, but we wanna do it without explicitly finding Q^{-1} so that it can be applied to general function where a Q might not exist. We gone prove that given n non-zero Q-conjugate vectors we can find x^* without having to compute Q^{-1}.

Result. Let $d^{(0)}, d^{(1)}, \ldots, d^{(n-1)}$ be a set of n non-zero vectors in $\mathbb{R}^n$ which are conjugate with respect to Q. Then x^*, which is the unique solution to the system $Qx = b$ is given by

$$x^* = \sum_{k=0}^{n-1} \left(\frac{(d^{(k)})^T b}{(d^{(k)}) Q d^{(k)}} \right) d^{(k)}$$

Proof. $d^{(0)}, d^{(1)}, \ldots, d^{(n-1)}$ are n linearly independent vectors in $\mathbb{R}^n$. Or, equivalently, they form a basis for $\mathbb{R}^n$. Therefore, there exist scalars

$\alpha_0, \alpha_1, \ldots \alpha_{n-1}$ such that:

$$x^* = \alpha_0 d^{(0)} + \alpha_1 d^{(1)} + \cdots + \alpha_{n-1} d^{(n-1)}$$

Now, pre-multiply with $(d^{(k)})^T Q$ and use conjugacy to determine α_k:

$$\alpha_k = \frac{(d^{(k)})^T Q x^*}{(d^{(k)})^T Q d^{(k)}}$$

but since x^* is solution of equation $Qx = b$, $Qx^* = b$.

$\square$

We will now rethink this as an iterative process:

Conjugate Direction Theorem. Let $\{d^{(0)}, d^{(1)}, \ldots, d^{(n-1)}\}$ be a set of n non-zero vectors in $\mathbb{R}^n$ which are conjugate with respect to Q. For any $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}$ generated according to

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k d^{(k)}, \\ \alpha_k &= -\frac{(g^k)^T d^{(k)}}{(d^{(k)})^T Q d^{(k)}}, \\ g^k &= Qx^k - b, \end{aligned}$$

converges to the unique solution x^* of the system $Qx = b$ exactly after n steps, i.e. $x^n = x^*$.

Proof. Since $d^{(0)}, d^{(1)}, \ldots, d^{(n-1)}$ form a basis of $\mathbb{R}^n$, there exist scalars $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ such that:

$$x^* - x^0 = \lambda_0 d^{(0)} + \lambda_1 d^{(1)} + \ldots + \lambda_{n-1} d^{(n-1)}$$

Again, pre-multiplying with $(d^{(k)})^T Q$, we get

$$\lambda_k = \frac{(d^{(k)})^T Q (x^* - x^0)}{(d^{(k)})^T Q d^{(k)}}$$

And, by following the iterative scheme:

$$\begin{aligned} x^1 - x^0 &= \alpha_0 d^{(0)} \\ x^2 - x^0 &= x^1 + \alpha_1 d^{(1)} - x^0 = \alpha_0 d^{(0)} + \alpha_1 d^{(1)} \\ &\vdots \\ x^k - x^0 &= \alpha_0 d^{(0)} + \alpha_1 d^{(1)} + \ldots + \alpha_{k-1} d^{(k-1)} \end{aligned}$$

Pre-multiplying by $(d^{(k)})^T Q$:

$$(d^{(k)})^T Q(x^k - x^0) = 0 \tag{1}$$

Now,

$$\lambda_k = \frac{(d^{(k)})^T Q(x^* - x^k + x^k - x^0)}{(d^{(k)})^T Q d^{(k)}}$$

but by 1:

$$\begin{aligned}
\lambda_k &= \frac{(d^{(k)})^T Q(x^* - x^k)}{(d^{(k)})^T Q d^{(k)}} \\
&= \frac{(d^{(k)})^T (Qx^* - Qx^k)}{(d^{(k)})^T Q d^{(k)}} \\
&= \frac{(d^{(k)})^T (b - Qx^k)}{(d^{(k)})^T Q d^{(k)}} \\
&= \frac{(d^{(k)})^T (-g^k)}{(d^{(k)})^T Q d^{(k)}} \\
&= \alpha_k
\end{aligned}$$

Also, $x^n - x^0 = \alpha_0 d^{(0)} + \alpha_1 d^{(1)} + \ldots + \alpha_{n-1} d^{(n-1)}$. But $\alpha_k = \lambda_k \Rightarrow x^n - x^0 = \lambda_0 d^{(0)} + \lambda_1 d^{(1)} + \ldots + \lambda_{n-1} d^{(n-1)} = x^* - x^0$. Thus, $x^n = x^*$.

$\square$

Conjugate gradient method for Quadratic case

Step 1 Choose $x^0 \in \mathbb{R}^n$, $d^{(0)} = -g^0 = b - Qx^0$. Set $k = 0$.

Step 2 Use the scheme:

$$\begin{aligned}
x^{k+1} &= x^k + \alpha_k d^{(k)} \\
\alpha_k &= \frac{-(g^k)^T d^{(k)}}{(d^{(k)})^T Q d^{(k)}} \\
d^{(k+1)} &= -g^{k+1} + \beta_k d^{(k)} \\
\beta_k &= \frac{(g^{k+1})^T Q d^{(k)}}{(d^{(k)})^T Q d^{(k)}} \\
g^k &= Qx^k - b
\end{aligned}$$

Step 3 Continue till we get $x^n = x^*$.

Note that we do not need all the conjugate directions from the start, we are generating them using β_k. β_k is such that new direction is conjugate to all previous directions: $(d^{(k)})^T Q d^{(i)} = 0, \ \forall \ i = 0, 1, \ldots, k$. This can be seen simply by pre-multiplying $d^{(k+1)}$ with $(d^{(k)})^T Q$ for $i = k$ and by induction for the rest of i.

Result. For conjugate gradient method, following hold:

- $(g^{k+1})^T d^{(k)} = 0$
- $\alpha_k = \dfrac{(g^k)^T g^k}{(d^{(k)})^T Q d^{(k)}}$
- $\beta_k = \dfrac{(g^{k+1})^T g^{k+1}}{(g^k)^T g^k}$
- $(d^{(k)})^T Q d^{(i)} = 0, (i = 0, 1, \ldots, k-1)$

General unconstrained problem

α_k is obtained from minimization of $\phi(\alpha) = f(x^k + \alpha d^{(k)})$. And after every $n-1$ iterations, the process is restarted (Powell's correction). The direction is reset to the steepest descent direction.

How is α_k obtained? Here's the sketch:

- We know that $\alpha = \dfrac{(d^{(k)})^T d^{(k)}}{(d^{(k)})^T Q d^{(k)}}$.
- Begin with an estimate $\hat{\alpha}_k$.
- Determine $\hat{f} \equiv f(x^k + \hat{\alpha}_k d^{(k)})$.
- Now, approximate $f(x)$ around $x = x^k$ with Taylor's series and evalu-

ate at $x^k + \hat{\alpha}_k d^{(k)}$:

$$
\begin{aligned}
f(x) &= f(x^k) + (\nabla f(x^k))^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k) \\
f(x^k + \hat{\alpha}_k) &= f(x^k) + (\nabla f(x^k))^T (\hat{\alpha}_k d^{(k)}) + \frac{1}{2}(\hat{\alpha}_k d^{(k)})^T \nabla^2 f(x^k)(\hat{\alpha}_k d^{(k)}) \\
\hat{f} &\approx f(x^k) + \hat{\alpha}_k (d^{(k)})^T \nabla f(x^k) + \frac{1}{2}\hat{\alpha}_k (d^{(k)})^T \nabla^2 f(x^k) d^{(k)}
\end{aligned}
$$

- From the actual value of $\hat{f}$ and approximated value of $\hat{f}$ we can produce an estimate for the denominator in calculation of α:

$$
(d^{(k)})^T \nabla^2 f(x^k) d^{(k)} = 2\frac{\hat{f} - f(x^k) - \hat{\alpha}_k (d^{(k)})^T \nabla f(x^k)}{\hat{\alpha}_k^2}
$$

- Use this to find α.