Discussion 6

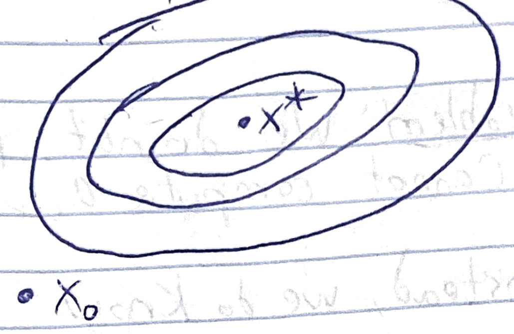
Conjugate Direction:

We have $f(x) = \frac{1}{2}x^T Q x - b^T x$, $\quad Q = Q^T,\ Q > 0$

$\nabla f(x) = Qx - b$

The level curves look like

The minimizer is $x^* = Q^{-1}b$



Suppose we start at x_0.

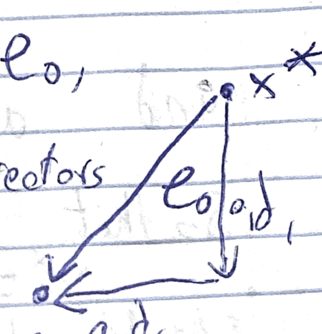
Idea: Suppose we have n orthogonal directions $d_0, d_1, \dots, d_{n-1}$, $\langle d_i, d_j \rangle = d_i^T d_j = 0$ for $i \neq j$

For example, in $\mathbb{R}^2$, take $d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $d_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Can we take one step in each direction to find x^*?

Can we write $x^* = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$

We consider error vector $x_0 - x^* = e_0$,
we can decompose it in terms of $d_0, d_1, \dots, d_{n-1}$ (since orthogonal vectors are linearly independent)



$e_0 = a_0 d_0 + a_1 d_1 + \dots + a_{n-1} d_{n-1}$

How to find $a_0, a_1, \dots, a_{n-1}$?

If we multiply by d_i^T, we have

$$d_i^T e_0 = a_0 d_i^T d_0 + \dots + a_{n-1} d_i^T d_{n-1}$$

All terms on right hand side except $a_i d_i^T d_i$ are zero (because $d_0, \dots, d_{n-1}$ are orthogonal)

So we have

$$d_i^T e_0 = a_i d_i^T d_i \quad \text{or} \quad a_i = \frac{d_i^T e_0}{d_i^T d_i}$$

Problem: We do not know e_0 so we cannot compute a_i.

Instead, we do know

$$Q e_0 = Q(x_0 - x^*) = Q x_0 - b = g_0$$

Idea: Instead of orthogonal, choose $d_0, d_1, \dots, d_{n-1}$ to be Q-conjugate, meaning $d_i^T Q d_j = 0$ for $i \neq j$.

Then if we write

$$x_0 - x^* = e_0 = a_0 d_0 + \dots + a_{n-1} d_{n-1}$$

and multiply by $d_i^T Q$, we get

$$d_i^T Q e_0 = a_0 d_i^T Q d_0 + \dots + a_{n-1} d_i^T Q d_{n-1} = a_i d_i^T Q d_i$$

and

$$a_i = \frac{d_i^T Q e_0}{d_i^T Q d_i} = \frac{d_i^T g_0}{d_i^T Q d_i}$$

So that

$$x^* = x_0 - a_0 d_0 - \dots - a_{n-1} d_{n-1}$$

We then have an algorithm

Conjugate Direction Algorithm

Given Q-conjugate directions $d_0, \dots, d_{n-1}$
and starting point x_0,

For $i = 0, 1, \dots, n-1$

$$\alpha_i = -a_i = \frac{-d_i^T g_0}{d_i^T Q d_i}$$

$$x_{i+1} = x_i + \alpha_i d_i$$

This converges in at most n steps.

Note: the book uses $\alpha_i = \frac{-d_i^T g_i}{d_i^T Q d_i}$ where $g_i = \nabla f(x_i)$

This is the same number, it turns out.

Note $x_i = x_0 + \alpha_0 d_0 + \dots + \alpha_{i-1} d_{i-1}$
and $x^* = x_0 + \alpha_0 d_0 + \dots + \alpha_{n-1} d_{n-1}$

We have $d_i^T g_i = d_i^T Q (x_i - x^*)$

$$= d_i^T Q \left((x_i - x_0) + (x_0 - x^*) \right)$$

$$= d_i^T Q (x_i - x_0) + d_i^T Q (x_0 - x^*)$$

$$= d_i^T Q (\alpha_0 d_0 + \dots + \alpha_{i-1} d_{i-1}) + d_i^T g_0$$

Note, as d_i is Q-conjugate to $d_0, \dots, d_{i-1}$, the left term is zero, so

$$d_i^T g_i = d_i^T g_0$$

and

$$\alpha_i = \frac{-d_i^T g_i}{d_i^T Q d_i} = \frac{-d_i^T g_0}{d_i^T Q d_i}$$

How to produce Q-conjugate Directions

The same way we make orthogonal directions, Gram-Schmitt!

Gram Schmitt works as follows:
Given $u_0, u_1, \dots, u_{n-1}$ linearly independent,
we make orthogonal vectors using the following algorithm

$d_0 = u_0$
for $i = 1, \dots, n-1$

$$d_i = u_i - \frac{u_i^T d_0}{d_0^T d_0} d_0 - \dots - \frac{u_i^T d_{i-1}}{d_{i-1}^T d_{i-1}} d_{i-1}$$

Intuitively, we subtract off component of u_i parallel to each of $d_0, d_1, \dots, d_{i-1}$

To make this work for Q-conjugate directions, replace standard inner product $u^T v$ with Q-innerproduct $u^T Q v$

So we have

$d_0 = u_0$
for $i = 1, \dots, n-1$

$$d_i = u_i - \frac{u_i^T Q d_0}{d_0^T Q d_0} d_0 - \dots - \frac{u_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}} d_{i-1}$$

After n steps, we have n Q-conjugate directions.
Then to solve our optimization problem,
we can apply general approach:

1) From any linearly independent set $u_0, \dots, u_{n-1}$, produce Q-conjugate $d_0, d_1, \dots, d_{n-1}$
2) Use conjugate direction algorithm with these directions.

Conjugate Gradient Algorithm

The conjugate gradient algorithm is exactly this approach with
$u_0 = -g_0, u_1 = -g_1, \ldots, u_{n-1} = -g_{n-1}$

This algorithm would appear as follows:

Given starting point x_0,
Let $u_0 = -g_0 = -\nabla f(x_0)$
Let $d_0 = u_0$ ← First step of Gram-Schmitt
Let $\alpha_0 = \dfrac{-g_0^T d_0}{d_0^T Q d_0}$ ← From conjugate direction
Let $x_1 = x_0 + \alpha_0 d_0$
For $i = 1, \ldots, n-1$
Let $u_i = -g_i = -\nabla f(x_i)$
Let $d_i = u_i - \dfrac{u_i^T Q d_0}{d_0^T Q d_0} d_0 - \cdots - \dfrac{u_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}} d_{i-1}$ ← Gram Schmitt
Let $\alpha_i = \dfrac{-g_i^T d_i}{d_i^T Q d_i}$ ← From Conjugate Direction
Let $x_{i+1} = x_i + \alpha_i d_i$

This converges in n steps.

Note, this algorithm is different from the one in your book because many terms in the expression

$$d_i = u_i - \frac{u_i^T Q d_0}{d_0^T Q d_0} d_0 - \cdots - \frac{u_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}} d_{i-1}$$

are zero. In fact, $u_i^T Q d_k = 0$ for $k < i-1$, See book for proof. So we can write

$$d_i = u_i - \frac{u_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}} d_{i-1}$$

If we let $\beta_{i-1} = \frac{-u_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}} = \frac{g_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}}$,

then our algorithm is

Given starting point x_0,

Let $u_0 = -g_0 = -\nabla f(x_0) = d_0$

Let $\alpha_0 = -\frac{g_0^T d_0}{d_0^T Q d_0}$

Let $x_1 = x_0 + \alpha_0 d_0$

For $i = 1, \dots, n-1$

  Let $u_i = -g_i = -\nabla f(x_i)$

  Let $\beta_{i-1} = \frac{g_i^T Q d_{i-1}}{d_{i-1}^T Q d_{i-1}}$

  Let $d_i = u_i + \beta_{i-1} d_{i-1}$

  Let $\alpha_i = \frac{-g_i^T d_i}{d_i^T Q d_i}$

  Let $x_{i+1} = x_i + \alpha_i d_i$.