Optimization

Màster de Fonaments de Ciència de Dades

Gerard Gómez

Lecture III. Alternating directions methods

Conjugate directions

- The main purpose of the alternating directions methods is to accelerate the convergence of the descent methods and, in this way, reduce the total number of iterations.
- One basic idea for these methods is the one related to conjugate directions which is a generalization of orthogonality.
- Two vectors $x, y \in \mathbb{R}^n$ are said to be conjugate directions with respect to the $n \times n$ symmetric positive definite matrix A if

$$x^T A y = 0$$

- If A is symmetric positive definite matrix, then it has n orthogonal eigenvectors. These n vectors are also mutually conjugate, since

$$x^T A y = x^T \lambda y = \lambda x^T y = 0$$

Thus, for every $n \times n$ symmetric positive definite matrix there is at least one set of n mutually conjugate directions

Conjugate directions

- **Remark:** Let $d_1, \ldots, d_m$ $(m \leq n)$ be m nonzero vectors mutually conjugate with respect to A, then these vectors are linearly independent.

  If this was not the case, then we could write

  $$d_m = \sum_{i=1}^{m-1} \alpha_i d_i$$

  from which it follows that

  $$(d_m)^T A d_m = 0$$

  that contradics the fact that $d_m \neq 0$ and that A is positive definite

Conjugate directions

- Let $v_1, \dots, v_k$ be k linearly independent vectors, then we can construct k mutually conjugate directions $d_1, \dots, d_k$, with respect to A, such that

$$< v_1, \dots, v_k > = < d_1, \dots, d_k >$$

The construction is similar to the Gram-Schmidt orthogonalization method. Define

$$\begin{aligned} d_1 &= v_1 \\ d_{i+1} &= v_{i+1} - \sum_{m=1}^{i} \frac{v_{i+1}^T A d_m}{d_m^T A d_m} d_m, \quad i = 1, \dots, k-1 \end{aligned}$$

Note that $d_m^T A d_m \neq 0$ since A is positive definite. Clearly

$$v_{i+1} \in < d_1, \dots, d_{i+1} > \quad \text{and} \quad d_{i+1} \in < v_1, \dots, v_{i+1} >$$

so $< v_1, \dots, v_{i+1} > = < d_1, \dots, d_{i+1} >$
Now we need to proof that if $d_1, \dots, d_i$ are mutually conjugate, then $d_{i+1}^T A d_j = 0$ for $j = 1, \dots, i$

$$d_{i+1}^T A d_j = v_{i+1}^T A d_j - \sum_{m=1}^{i} \frac{v_{i+1}^T A d_m}{d_m^T A d_m} d_m^T A d_j = v_{i+1}^T A d_j - \frac{v_{i+1}^T A d_j}{d_j^T A d_j} d_j^T A d_j = 0$$

since $d_m^T A d_j = 0$ except if $m = 1$

Conjugate directions

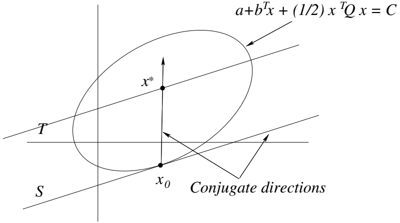
A geometric interpretation of conjugate vectors is the following. Let

$$f(x) = a + b^T x + \frac{1}{2} x^T A x$$

with A a symmetric positive definite matrix, be a quadratic function with a global minimum

$$f'(x^*) = 0 \quad \Rightarrow \quad x^* = -A^{-1} b$$

Then, the surfaces $f(x) = c$ (constant) are generally ellipsoids with center at x^*. Let x_0 be a point satisfying $f(x_0) = c$



Then the vector joining x_0 and x^* is conjugate with respect to A to every vector in the tangent hyperplane to the ellipsoid at x_0

Conjugate directions

Two affine spaces S and T $(S \neq T)$ are parallel if they are generated by the same set of vectors $z_1, \ldots, z_m$ but at different points: $x(S) \in S$, $x(T) \in T$ and $x(S) \neq x(T)S$.

Theorem

Let $x^*(S)$ and $x^*(T)$ be the points that minimize

$$f(x) = a + b^T x + \frac{1}{2} x^T A x$$

in two parrallel affine spaces S and T. Then $x^*(S) - x^*(T)$ and any direction contained in S and T are conjugate w.r.t. A

Proof: Let z be a direction of S and T, then

$$\frac{d}{d\alpha}[f(x^*(S)) + \alpha z]_{\alpha=0} = 0 \quad \Rightarrow \quad z^T[Ax^*(S) + b] = 0$$

$$\frac{d}{d\alpha}[f(x^*(T)) + \alpha z]_{\alpha=0} = 0 \quad \Rightarrow \quad z^T[Ax^*(T) + b] = 0$$

so

$$z^T A[x^*(S) - x^*(T)] = 0$$

$\square$

Conjugate directions

Theorem

Let $z_1, \ldots, z_m$, $z_i \in \mathbb{R}^n$, $z_i \neq 0$, $m \leq n$ be m mutually conjugate directions with respect to the poisitive definite matrix A, then the minimum of the quadratic function

$$f(x) = a + b^T x + \frac{1}{2} x^T A x$$

over the affine set generated by the point $x_0 \in \mathbb{R}^n$ and the vectors $z_1, \ldots, z_m$ will be found by searching along each of the conjugate directions once only

Proof: The minimum will be a point $x_0 + \alpha_1^* z_1 + \ldots + \alpha_m^* z_m$, such that the α_i^* minimize

$$f\left(x_0 + \sum_{j=1}^{m} \alpha_j z_j\right) = f(x_0) + \sum_{j=1}^{m} \left[\alpha_j z_j^T (b + Ax_0) + \frac{1}{2}\alpha_j^2 z_j^T A z_j\right]$$

Since in the above expression there are no $\alpha_j \alpha_k$ terms with $j \neq k$, the optimal α_j are found minimizing each summand:

$$\min_{\alpha_j} \left[f(x_0) + \alpha_j z_j^T (b + Ax_0) + \frac{1}{2}\alpha_j^2 z_j^T A z_j\right] = \min_{\alpha_j} f(x_0 + \alpha_j z_j), \quad j = 1, \ldots, m$$

$\square$

Conjugate directions

Example. Consider the quadratic function

$$f(x,y) = 2x^2 + 6y^2 + 2xy + 2x + 3y + 3$$

that can also be written as

$$f(x,y) = \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 12 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + 3$$

We choose $z_1 = (1,0)^T$. A conjugate direction to z_1 with respect to Q is $z_2 = v(-1/2, 1)^T$. Let us find the minimum of f generated by the point $x_0 = (0,0)^T$ and the vectors z_1, z_2.
Staring with the z_1 direction, we want to minimize

$$F_1(\theta) = f(x_0 + \theta z_1) = f(\theta,\, 0)$$

The minima is achieved for $\theta_1^* = -1/2$.
Proceeding now with the z_2 direction, we need to minimize

$$F_2(\theta) = f(x_0 + \theta z_2) = f(-\theta/2,\, \theta)$$

and, using the same method, we get $\theta_2^* = -2/11$.
So, the minimum of f is then given by

$$x^* = x_0 + \theta_1^* z_1 + \theta_2^* z_2 = \begin{pmatrix} -\frac{9}{22} \\ -\frac{2}{11} \end{pmatrix}$$

Exercise 5: Compute θ_1^* and θ_2^* using the quadratic method for the 1-dimensional minimization

Conjugate gradient methods

Conjugate gradient methods generate a sequence

$$x^k = x^{k-1} + \alpha_k z^k i, \quad k = 1, 2, \dots$$

Suposse that the direction z^k is given, and define

$$F(\alpha_k) = f(x^{k-1} + \alpha_k z^k)$$

The value of α_k is chosen such that

$$\frac{dF(\alpha_k^*)}{d\alpha_k} = (z^k)^T \nabla f(x^{k-1} + \alpha_k^* z^k) = (z^k)^T \nabla f(x^k) = 0$$

Assume that f is the quadratic function

$$f(x) = a + b^T x + \frac{1}{2} x^T Q x$$

with Q an $n \times n$ symmetric positive definite matrix. Then, the gradient of f at two points are related by

$$\nabla f(x^k) = \nabla f(x^{k-1}) + Q(x^k - x^{k-1}) \quad \Leftrightarrow \quad \left(b + Qx^k = b + Qx^{k-1} + Q(x^k - x^{k-1})\right)$$

Conjugate gradient methods (cont.)

If $x^k = x^{k-1} + \alpha_k z^k$ we can obtain an explicit formula for α_k^* from the condition $dF(\alpha_k^*)/d\alpha_k = 0$:

$$\begin{aligned}(z^k)^T \nabla f(x^k) &= (z^k)^T \left(\nabla f(x^{k-1}) + Q(x^k - x^{k-1}) \right) \\ &= (z^k)^T \left(\nabla f(x^{k-1}) + \alpha_k^* Q z^k \right) = 0\end{aligned}$$

$$\Rightarrow \quad \alpha_k^* = -\frac{(z^k)^T \nabla f(x^{k-1})}{(z^k)^T Q z^k}$$

Since

$$\begin{aligned} f(x^k) &= f(x^{k-1}) + (x^k - x^{k-1})^T \nabla f(x^{k-1}) + \frac{1}{2}(x^k - x^{k-1})^T Q (x^k - x^{k-1}) \\ &= f(x^{k-1}) + \alpha_k^* (z^k)^T \nabla f(x^{k-1}) + \frac{1}{2}(\alpha_k^*)^2 (z^k)^T Q z^k \end{aligned}$$

and using the value obtained for α_k^* we get

$$f(x^{k-1}) - f(x^k) = \frac{\left[(z^k)^T \nabla f(x^{k-1})\right]^2}{2(z^k)^T Q z^k} > 0$$

so we have a descent method

Conjugate gradient methods. The algorithm

- We would like to have an algorithm that converges rapidly or, even better, that terminates in a finite number of steps when applied to minimizing a quadratic function.
- We have already seen that if the search directions z^k are mutually conjugate with respect to Q for $k = 1, \ldots, n$, then the point x^n will be the exact minimum of the quadratic function.
- The choice of the conjugate directions can be done in the following way:
  1. We start at a point $x^0 \in \mathbb{R}^n$ and choose
  $$z^1 = -\nabla f(x^0)$$
  2. The next point is
  $$x^1 = x^0 + \alpha_1^* z^1$$
  where α_1^* has been computed with the formula given above
  3. We evaluate $\nabla f(x^1)$ and set
  $$z^2 = -\nabla f(x^1) + \beta_{11} z^1,$$
  where β_{11} is such that z^1 and z^2 will be Q-conjugate, this is
  $$(z^1)^T Q z^2 = (z^1)^T Q(-\nabla f(x^1) + \beta_{11} z^1) = 0,$$
  from which it follows
  $$\beta_{11} = \frac{(z^1)^T Q \nabla f(x^1)}{(z^1)^T Q z^1}.$$

Conjugate gradient methods. The algorithm (cont.)

4. Once z^2 is known, we determine $x^2 = x^1 + \alpha_2^* z^2$, with α_2^* computed with the formula given above. We evaluate $\nabla f(x^2)$ and the new direction will be
$$z^3 = -\nabla f(x^2) + \beta_{21} z^1 + \beta_{22} z^2,$$
with β_{21} and β_{22} such that $(z^1)^T Q z^3 = (z^2)^T Q z^3 = 0$.
5. In general, we get
$$z^{k+1} = -\nabla f(x^k) + \sum_{j=1}^{k} \beta_{kj} z^j, \quad k = 0, \dots, n-1.$$

If the function f is not quadratic, the computation of β_{ij} is long.

We shall show how these directions can be generated more easily.

The following result will be useful in the sequel

Theorem

Let $f(x) = a + b^T x + \frac{1}{2} x^T Q x$ *and* $x^0 \in \mathbb{R}^n$ *be given, and assume that the* m *nonzero vectors* $z^1, \ldots, z^m$, $z^j \in \mathbb{R}^n$, $m \leq n$, *are mutually conjugate w.r.t* Q.

Starting at x^0, *we move to* $x^1, \ldots, x^m$ *along* $z^1, \ldots, z^m$, *respectively, such that*

$$(z^j)^T \nabla f(x^j) = 0, \quad j = 1, \ldots, m$$

then

$$(z^j)^T \nabla f(x^m) = 0, \quad j = 1, \ldots, m$$

Conjugate gradient methods. Proof of the Theorem

Proof: For $j = m$ the result is obvious

Since, as we have already seen, $\nabla f(x^k) = \nabla f(x^{k-1}) + Q(x^k - x^{k-1})$, it follows that the gradient of f at any two points are related by

$$\begin{aligned} \nabla f(x^m) &= \nabla f(x^{m-1}) + Q(x^m - x^{m-1}) \\ &= \nabla f(x^{m-2}) + Q(x^{m-1} - x^{m-2}) + Q(x^m - x^{m-1}) \\ &= \nabla f(x^{m-2}) + Q(x^m - x^{m-2}), \end{aligned}$$

so

$$\nabla f(x^m) = \nabla f(x^j) + Q(x^m - x^j), \quad j = 1, \ldots, m-1. \tag{1}$$

From $x^j = x^{j-1} + \alpha_j^* z^j$, for $j = 1, \ldots, m$, it follows that

$$x^m = x^{m-1} + \alpha_m^* z^m = x^{m-2} + \alpha_{m-1}^* z^{m-1} + \alpha_m^* z^m = \ldots$$

so

$$x^m - x^j = \sum_{i=j+1}^{m} \alpha_i^* z^i, \quad j = 0, \ldots, m-1$$

Conjugate gradient methods. Proof of the Theorem (cont.)

In this way, we can write

$$\nabla f(x^m) = \nabla f(x^j) + \sum_{i=j+1}^{m} \alpha_i^* Q z^i, \quad j = 1, \ldots, m-1,$$

from which it follows that

$$(z^j)^T \nabla f(x^m) = (z^j)^T \nabla f(x^j) + \sum_{i=j+1}^{m} \alpha_i^* (z^j)^T Q z^i = 0, \quad j = 1, \ldots, m-1.$$

since the first term of the right-hand side vanishes, according to the hypothesis, and the second by conjugacy.

□

Conjugate gradient methods. Corollary

Corollary

If in the above theorem $m = n$, then $\nabla f(x^n) = 0$ and x^n is the unconstrained minimum of f.

Proof: Since the z^j are linearly independent, from

$$\sum_{j=1}^{n} (z^j)^T \nabla f(x^n) = \sum_{j=1}^{n} \nabla f(x^n)^T z^j = 0,$$

it follows that $\nabla f(x^n) = 0$. $\square$

Conjugate gradient methods. Computation of the β_{ij} coefficients

For the computation of the constants β_{ij} we will use the following remark:

- Let $\gamma^i = \nabla f(x^i) - \nabla f(x^{i-1}), \quad i = 1, ..., n$

  If f is a quadratic function, then $\gamma^i = Q(x^i - x^{i-1})$.

  Choose x^i and z^i such that $x^i - x^{i-1} = \alpha_i^* z^i$.

  According to the definition of γ^i

  $$\gamma^i = Q(x^i - x^{i-1}) = \alpha_i^* Q z^i \quad \Rightarrow \quad (\gamma^i)^T = \alpha_i^* (z^i)^T Q, \quad i = 1, ..., n$$

  so

  $$(\gamma^i)^T z^j = \alpha_i^* (z^i)^T Q z^j, \quad i = 1, ..., n, \quad j = 1, ..., n.$$

  If $z^1, ..., z^k$, $k \leq n$ are chosen to be mutually conjugate w.r.t. Q, we get

  $$(\gamma^i)^T z^j = 0, \quad i = 1, ..., k, \quad j = 1, ..., k, \quad i \neq j.$$

Let us use these computations to obtain another expression of β_{11}.

Computation of the β_{ij} coefficients

From

$$(\gamma^1)^T z^2 = (\gamma^1)^T[-\nabla f(x^1)+\beta_{11}\nabla f(x^0)] = -(\nabla f(x^1)-\nabla f(x^0))^T(\nabla f(x^1)+\beta_{11}\nabla f(x^0))$$

we get

$$\beta_{11} = \frac{(\nabla f(x^1) - \nabla f(x^0))^T \nabla f(x^1)}{(\nabla f(x^1) - \nabla f(x^0))^T(-\nabla f(x^0))}.$$

On the other hand, since $z^1 = -\nabla f(x^0)$, and recalling that the value of α_k^* was chosen such that

$$\frac{dF(\alpha_k^*)}{d\alpha_k} = \frac{df(x^{k-1} + \alpha_k^* z^k)}{d\alpha_k} = (z^k)^T \nabla f(x^{k-1} + \alpha_k^* z^k) = (z^k)^T \nabla f(x^k) = 0$$

it follows that

$$(z^1)^T \nabla f(x^1) = -(\nabla f(x^0))^T \nabla f(x^1) = 0,$$

so

$$\beta_{11} = \frac{(\nabla f(x^1))^T \nabla f(x^1)}{(\nabla f(x^0))^T \nabla f(x^0)}.$$

Computation of the β_{ij} coefficients

The point x^2 is reached by minimizing along the conjugate directions z^1 and z^2. According to the last Theorem

$$(z^1)^T \nabla f(x^2) = (z^2)^T \nabla f(x^2) = 0.$$

Substituting $z^1 = -\nabla f(x^0)$ and $z^2 = -\nabla f(x^1) + \beta_{11} z^1$ in these equalities, we get

$$(\nabla f(x^0))^T \nabla f(x^2) = 0, \quad (\nabla f(x^1))^T \nabla f(x^2) = 0.$$

From $(\gamma^1)^T z^3 = 0$, $(\gamma^2)^T z^3 = 0$ and the above equalities, it follows that

$$\begin{aligned} \beta_{21} &= 0, \\ \beta_{22} &= \frac{(\nabla f(x^2))^T \nabla f(x^2)}{(\nabla f(x^1))^T \nabla f(x^1)}. \end{aligned}$$

In a similar way, we can also establish that

$$\begin{aligned} \beta_{kj} &= 0, \quad \text{for} \quad k \neq j \\ \beta_{kk} &= \frac{(\nabla f(x^k))^T \nabla f(x^k)}{(\nabla f(x^{k-1}))^T \nabla f(x^{k-1})}, \quad k = 1, \dots, n \end{aligned}$$

thus

$$z^{k+1} = -\nabla f(x^k) + \frac{(\nabla f(x^k))^T \nabla f(x^k)}{(\nabla f(x^{k-1}))^T \nabla f(x^{k-1})} z^k. \tag{2}$$

The conjugate gradient algorithm

1. Choose a starting point $x^0 \in \mathbb{R}^n$.
2. Evaluate $\nabla f(x^0)$ and then set $z^1 = -\nabla f(x^0)$.
3. Move to $x^1, x^2, ..., x^n$ by minimizing $f(x)$ along the directions $z^1, ..., z^n$ computed according to

$$z^{k+1} = -\nabla f(x^k) + \frac{(\nabla f(x^k))^T \nabla f(x^k)}{(\nabla f(x^{k-1}))^T \nabla f(x^{k-1})} z^k$$

4. After these n minimizations, restart the procedure by letting x^n and $-\nabla f(x^n)$ be the new x^0 and z^1.
5. Repite the above two steps until

$$(\nabla f(x^k))^T \nabla f(x^k) \leq \epsilon,$$

where ϵ is some predetermined small number.

The conjugate gradient algorithm. Example

Example: Consider
$$f(x) = \frac{3}{2}x^2 + \frac{1}{2}y^2 - xy - 2x,$$
so
$$a = 0, \quad b = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}.$$
We take
$$x^0 = \begin{pmatrix} -2 \\ 4 \end{pmatrix}, \quad \nabla f(x^0) = \begin{pmatrix} -12 \\ 6 \end{pmatrix}, \quad z^1 = \begin{pmatrix} 12 \\ -6 \end{pmatrix}.$$
Minimizing $f(x^0 + \alpha_1 z^1)$ with respect to α_1 we get $\alpha_1^* = 5/17$, so
$$x^1 = \begin{pmatrix} 26/17 \\ 38/17 \end{pmatrix}, \quad \nabla f(x^1) = \begin{pmatrix} 6/17 \\ 12/17 \end{pmatrix}$$
So, we have
$$z^2 = -\nabla f(x^1) + \frac{(\nabla f(x^1))^T \nabla f(x^1)}{(\nabla f(x^0))^T \nabla f(x^0)} z^1 = -\begin{pmatrix} 6/17 \\ 12/17 \end{pmatrix} + \frac{(6/17)^2 + (12/17)^2}{(-12)^2 + 6^2} \begin{pmatrix} 12 \\ -6 \end{pmatrix}$$
$$= -\begin{pmatrix} 90/289 \\ 210/289 \end{pmatrix}.$$
Minimizing $f(x^1 + \alpha_2 z^2)$ with respect to α_2 we get $\alpha_2^* = 17/10$. Consequently $x^2 = (1, 1)^T$, which is the global minimum of f.

Powell's method

We start presenting Powell's method as an empirical technique, later we will justify its underlying principles. The method does not require the computation of derivatives.

The basic version of the method is as follows:

1. Each stage of the procedure consists of $n+1$ successive 1-dimensional line searches
2. The first n searches are done along n linearly independent directions
3. The $(n+1)$th search is done along the direction connecting the obtained best point (obtained at the end of the n preceeding 1-dimensional line searches) with the starting point of that stage.
4. After these searches, one of the first n directions is replaced by the $(n+1)$th and a new stage begins.

Powell's method

The kth stage of the method is given by the following steps:

1. Let $x_B^{k-1} = t_0^k \in \mathbb{R}^n$ be the starting point of the kth stage and $\Delta_1^k, \ldots, \Delta_n^k$, n linearly independent directions.
2. Determine θ_j^*, per $j = 1, \ldots, n$ such that

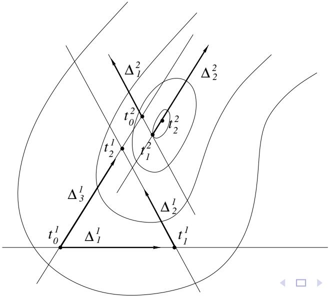
$$f(t_{j-1}^k + \theta_j^* \Delta_j^k) = \min_{\theta_j} f(t_{j-1}^k + \theta_j \Delta_j^k),$$

   and define

$$t_j^k = t_{j-1}^k + \theta_j^* \Delta_j^k.$$

3. The new search directions are

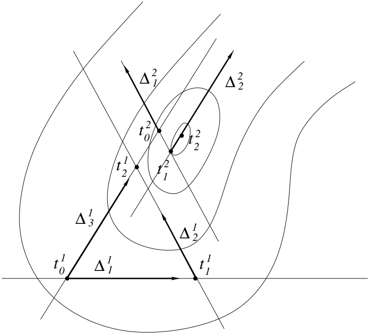
$$\begin{aligned} \Delta_j^{k+1} &= \Delta_{j+1}^k, \quad j = 1, \ldots, n-1, \\ \Delta_n^{k+1} &= \Delta_{n+1}^k = t_n^k - t_0^k. \end{aligned}$$



Powell's method (cont.)

4. Find θ^*_{n+1} such that

$$f(t^k_n + \theta^*_{n+1}(t^k_n - t^k_0)) = \min_{\theta_{n+1}} f(t^k_n + \theta_{n+1}(t^k_n - t^k_0)),$$



5. Take as new initial point

$$x^k_B = t^k_n + \theta^*_{n+1}(t^k_n - t^k_0).$$

6. If $\|x^{k-1}_B - x^k_B\| < \epsilon$ ($\epsilon > 0$ fixed) stop, otherwise proceed to stage $k+1$

Powell's method. Example: first step

Let

$$f(x,y) = \frac{3}{2}x^2 + \frac{1}{2}y^2 - xy - 2x,$$

which has a minimum at (1, 1).

1. We start with

$$x_B^0 = t_0^1 = \begin{pmatrix} -2 \\ 4 \end{pmatrix}, \quad \Delta_1^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \Delta_1^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

2. The first minimization is in the Δ_1^1 direction

$$\min_{\theta_1} f(t_0^1 + \theta_1 \Delta_1^1) = \min_{\theta_1} \left\{ \frac{3}{2}(-2+\theta_1)^2 + \frac{1}{2}4^2 - (-2+\theta_1)4 - 2(-2+\theta_1) \right\}$$

$$\Rightarrow \quad \theta_1^* = 4, \quad \Rightarrow \quad t_1^1 = (2,\ 4)^T$$

3. Now we minimize in the Δ_2^1 direction

$$\min_{\theta_2} f(t_1^1 + \theta_2 \Delta_2^1) = \min_{\theta_2} \left\{ \frac{3}{2}2^2 + \frac{1}{2}(4+\theta_2)^2 - 2(4+\theta_2) - 4 \right\}$$

$$\Rightarrow \quad \theta_2^* = -2, \quad \Rightarrow \quad t_2^1 = (2,\ 2)^T$$

4. Consequently, the new direction is

$$\Delta_3^1 = \begin{pmatrix} 2-(-2) \\ 2-4 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

Powell's method. Example: first step

5. Next we minimize along the new direction Δ_3^1

$$\min_{\theta_3} f(t_2^1 + \theta_3 \Delta_3^1) =$$

$$= \min_{\theta_3} \left\{ \frac{3}{2}(2+4\theta_3)^2 + \frac{1}{2}(2-2\theta_3)^2 - (2+4\theta_3)(2-2\theta_3) - 2(2+4\theta_3) \right\}$$

$$\Rightarrow \quad \theta_3^* = -2/17, \quad \Rightarrow \quad x_B^1 = t_0^2 = \begin{pmatrix} 2-8/17 \\ 2+4/17 \end{pmatrix} \begin{pmatrix} 26/17 \\ 38/17 \end{pmatrix}$$

This concludes the first iteration of the algorithm. The first two search directions of the second iteration are

$$\Delta_1^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \Delta_2^2 = \begin{pmatrix} 4 \\ -2 \end{pmatrix}.$$

Powell's method. Example: second step

1. The first minimization is in the Δ_1^2 direction

$$\min_{\theta_1} f(t_0^2+\theta_1\Delta_1^2) = \min_{\theta_1}\left\{\frac{3}{2}\left(\frac{26}{17}\right)^2+\frac{1}{2}\left(\frac{38}{17}+\theta_1\right)^2-\frac{26}{17}\left(\frac{38}{17}+\theta_1\right)-\frac{52}{17}\right\}$$

$$\Rightarrow \quad \theta_1^* = -12/17, \quad \Rightarrow \quad t_1^2 = (26/17,\ 26/17),$$

2. The second minimization is in the Δ_2^2 direction

$$\min_{\theta_2} f(t_1^2+\theta_2\Delta_2^2) =$$

$$= \min_{\theta_2}\left\{\frac{3}{2}\left(\frac{26}{17}+4\theta_2\right)^2+\frac{1}{2}\left(\frac{26}{17}-2\theta_2\right)^2-\left(\frac{26}{17}+4\theta_2\right)\left(\frac{26}{17}-2\theta_2\right)-2\left(\frac{26}{17}+4\theta_2\right)\right\}$$

$$\Rightarrow \quad \theta_2^* = -18/289, \quad \Rightarrow \quad t_2^2 = (370/289,\ 478/289)$$

3. The new direction is

$$\Delta_3^2 = \begin{pmatrix} -72/289 \\ -168/289 \end{pmatrix}$$

4. Finally, when we compute $\min_{\theta_3} f(t_2^2+\theta_3\Delta_3^2) = \ldots$ we get

$$\theta_3^* = 9/8, \quad x_B^2 = (1,\ 1),$$

That is, the exact minimum of the quadratic function is found in two iterations

Powell's method. Example 2

In the above example the directions $\Delta_{1,2}^k$ $(k = 1, 2)$ were linearly independent. This condition is important, as is shown in the next example.

Let

$$f(x, y, z) = (x - y + z)^2 + (-x + y + z)^2 + (x + y - z)^2,$$

that has a minimum at $(x^*, y^*, z^*) = (0,\ 0,\ 0)$. Start Powell's method with

$$x_B^0 = \begin{pmatrix} 1/2 \\ 1 \\ 1/2 \end{pmatrix}, \quad \Delta_1^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \Delta_2^1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \Delta_3^1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The results of the first three steps are

j	Δ_j^1	t_j^1	$f(t_j^1)$
0	–	$(1/2,\ 1,\ 1/2)$	2
1	Δ_1^1	$(1/2,\ 1,\ 1/2)$	2
2	Δ_3^1	$(1/2,\ 1/3,\ 1/2)$	$2/3$
3	Δ_3^1	$(1/2,\ 1/3,\ 5/18)$	$42/81$

The new direction is

$$t_3^1 - t_0^1 = \begin{pmatrix} 1/2 \\ 1/3 \\ 5/18 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 1 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2/3 \\ -2/9 \end{pmatrix}$$

The new search directions are

$$\Delta_1^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \Delta_2^2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \Delta_3^2 = \begin{pmatrix} 0 \\ -2/3 \\ -2/9 \end{pmatrix}.$$

Thus, the first component of all forthcoming points reached will remain equal to $1/2$, and the true optimum at $(x^*, y^*, z^*) = (0,\ 0,\ 0)$ can never be reached.

Convergence of Powell's method

Let us show how the properties about conjugate directions can be used to prove quadratic termination of Powell's method under suitable assumptions.

Assume that:

- The function f is quadratic and Q is a positive define matrix
- The initial point is $x_B^0 \in \mathbb{R}^n$.
- The initial directions $\Delta_1^1, \ldots, \Delta_n^1$ are linearly independent

After the steps of the first stage, we have:

- A new direction $\Delta_n^2 = t_n^1 - t_0^1 = z^1$,
- A new starting point $x_B^1 = t_0^2$.

Let us see that if a point in $\mathbb{R}^n$ is optimal in n linearly independent directions, then it must be the global optimum of the quadratic function.

- Assume that $t_n^1 \neq t_0^1$, that is: $\Delta_n^2 \neq 0$.
- The point $x_B^1 = t_n^1 + \theta_{n+1}^*(t_n^1 - t_0^1)$, is a minimum of f in the $\Delta_n^2 = t_n^1 - t_0^1$ direction.
- If $\Delta_1^2, \ldots, \Delta_n^2$ are linearly independent, then after the second step of the method we arrive at a point t_n^2 that is also a minimum of f in the Δ_n^2 direction and will be contained in a parallel afine set
- Because of the properties of the conjugate directions, the direction $z^2 = t_n^2 - t_0^2$ is conjugate to z^1 with respect to Q.

Convergence of Powell's method

- After k steps of the procedure, we have generated k non-zero directions $z^1, \ldots, z^k$ mutually conjugate w.r.t. Q
- If the directions $\Delta_1^k, \ldots, \Delta_{n-k}^k$, $z^1, \ldots, z^k$ are linearly independent, then $z^{k+1} = t_n^{k+1} - t_0^{k+1}$ will be conjugate to $z^1, \ldots, z^k$.
- After completing n stages all the search directions are mutually conjugate w.r.t. Q and, according to a preceeding Theorem, the minimum of f over $\mathbb{R}^n$ has been reached.

Avoiding linearly dependent search directions

We can modify Powell's method to avoid linearly dependent search directions. The new method does not possess the quadratic termination property, but has a satisfactory performance.

Let

- $x_B^{k-1} = t_0^k$ be the starting point of the kth stage
- $\Delta_1^k, \ldots, \Delta_n^k, n$ linearly independent directions
- t_j^k, $j = 1, ..., n$ the minima of f along the directions $\Delta_1^k, \ldots, \Delta_n^k$

We want to find the index m such that

$$f(t_{m-1}^k) - f(t_m^k) = \max_{j=1,\ldots,n} \{f(t_{j-1}^k) - f(t_j^k)\}.$$

Set $\Delta_{n+1}^k = t_n^k - t_0^k$

If $\|t_n^k - t_0^k\| < \epsilon$ stop, otherwise find α_{n+1}^* such that

$$f(t_0^k + \alpha_{n+1}^* \Delta_{n+1}^k) = \min_{\alpha_{n+1}} f(t_0^k + \alpha_{n+1} \Delta_{n+1}^k),$$

and let $t_0^{k+1} = x_B^k = t_0^k + \alpha_{n+1}^*$.

Avoiding linearly dependent search directions

If $\|x_B^k - x_B^{k-1}\| < \epsilon$, stop (convergence).
Otherwise, if

$$|\alpha_{n+1}^*| < \left(\frac{f(t_0^k) - f(t_0^{k+1})}{f(t_{m+1}^k) - f(t_m^k)} \right)^{1/2}, \tag{3}$$

set $\Delta_j^{k-1} = \Delta_j^k$, $j = 1, ..., n$.

In other words, the search directions of the $(k+1)$th stage in the same as in the kth stage. If (3) does not hold, set

$$\begin{aligned} \Delta_j^{k-1} &= \Delta_j^k, \quad j = 1, ..., m-1 \\ \Delta_j^{k-1} &= \Delta_{j+1}^k, \quad j = m, ..., n \end{aligned}$$

and proceed to stage $k+1$

Avoiding linearly dependent search directions. Example

Consider again

$$f(x,y,z) = (x-y+z)^2 + (-x+y+z)^2 + (x+y-z)^2$$

The first steps are the same as before. We can see that the largest function decrease is obtained by going from t_1^1 to t_2^1, hence $m = 2$.
We have

$$\Delta_4^1 = (0,\ -2/3,\ -2/9)^T$$

We find that $\alpha_4^* = 9/8$ minimizes

$$f(1/2,\ 1-(2/3)\alpha_4,\ 1/2-(2/9)\alpha_4) \quad \Rightarrow$$

$$t_0^2 = (1/2,\ 1,\ 1/2)^T + (9/8)(0,\ -2/3,\ -2/9)^T = (1/2,\ 1/4,\ 1/4)^T \quad \Rightarrow f(t_0^2) = 1/2$$

Now

$$\left(\frac{f(t_0^1)-f(t_0^2)}{f(t_1^1)-f(t_2^1)}\right)^{1/2} = \left(\frac{2-1/2}{2-2/3}\right)^{1/2} = (9/8)^{1/2}$$

Since $\alpha_4^* > (9/8)^{1/2}$, we see that (3) does not hold. Accordingly, the new directions will be the independents vectors

$$\begin{aligned}
\Delta_1^2 &= (1,\ 0,\ 0)\\
\Delta_2^2 &= (0,\ 0,\ 1)\\
\Delta_3^2 &= (0,\ -(2/3),\ -(2/9))
\end{aligned}$$