1. The directional minimization

1.1. Algorithm of minimization along a line.

Among direct minimum search methods one of the most commonly used is based on the directional minimization. Its general form is as follows:

Algorithm of the directional minimization

(1) INPUT x_1
(2) FOR $k = 1, 2, \ldots$
(3) Choose $s_k \in \mathbb{R}^n$ search direction!
(4) Find $\alpha_k \in \mathbb{R}^+$ for which

$$f(x_k + \alpha_k s_k) = \min_{\alpha \geq 0} f(x_k + \alpha s_k).$$

(5) $x_{k+1} = x_k + \alpha_k s_k$.

The following descent constraint is imposed on the process

$$f(x_1) \geq f(x_2) \geq \ \ldots \ \geq f(x_k) \geq f(x_{k+1}) \geq \ \ldots.$$

The descent constraint will hold if the directional vector s_k is chosen so that

$$g'(0) = \nabla f(x_k)^T s_k < 0.$$

Then as $g'(0) < 0$ the function g decreases at $\alpha = 0$. Consequently, there is an $\alpha_k > 0$, such that $f(x_{k+1}) = f(x_k + \alpha_k s_k) \leq f(x_k)$. By experience the function values $f(x_k)$ do not decrease fast.

1.2. Direction search methods.

There are a good deal of means to choose search directions $\{s_k\}$:
1. The fastest reduction method: $s_k = -\nabla f(x_k)$;
2. Newton-like methods: $s_k = -B_k^{-1} \nabla f(x_k)$, where B_k is a symmetric positive matrix.

In both cases, the descent property holds. In the case of the fastest reduction method we have

$$\nabla f(x_k)^T s_k = -\|\nabla f(x_k)\|^2 < 0 \quad (\nabla f(x_k) \neq 0).$$

Since matrix B_k^{-1} is positive definite, in the case of Newton-like methods we have

$$\nabla f(x_k)^T s_k = -\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) < 0 \quad (\nabla f(x_k) \neq 0).$$

1.3. **Gradient method (or the Cauchy's method, 1847).**

As we know, the gradient points in the direction of greatest increase of the function. So the most obvious choice of the direction search is $d = -\nabla f(x)$, because it is in this direction to expect the largest decrease of the function. This method is commonly referred to as the greatest decrease method. The method is summarized here below.

Gradient method algorithm:

(1) INPUT: $k = 0$ s $x_k \in \mathbb{R}^n$ point.
(2) FOR $k = 1, 2, \ldots$
(3) $\quad s_k = -\nabla f(x_k)$
(4) $\quad \alpha_k > 0 : f(x_k + \alpha_k s_k) = \min_{\alpha \geq 0} f(x_k + \alpha s_k)$
(5) $\quad x_{k+1} = x_k + \alpha_k s_k$

Stopping conditions can be $\|x_{k+1} - x_k\| < \varepsilon$, or $\|\nabla f(x_k)\| < \varepsilon$.

1.3.1. *Example.*

Exercise: Solve the following optimization problem with gradient method, with $x_1 = (0, 0)$ the guess vector and $\varepsilon = 0.005$ the tolerance level:

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \to \min!$$

Solution:
The gradient of f:

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}$$

Step 1: $x_1 = (0, 0)$

$$\nabla f(\mathbf{x}_1) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \qquad s_1 = -\nabla f(x_1) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Then $x = x_1 + \alpha s_1 = (\alpha, -\alpha)$. The univariate optimization problem and its solution:
$\varphi(\alpha) = f(\alpha, -\alpha) = \alpha^2 - 2\alpha + 5 \to \min!$. Then we get $\alpha_1 = 1$
Step 2: $x_2 = x_1 + \alpha_1 s_1 = (\alpha_1, -\alpha_1) = (1, -1)$, hence $s_2 = -\nabla f(x_2) = (1, 1)$ and $x = x_2 + \alpha s_2 = (1 + \alpha, -1 + \alpha)$. The univariate optimization problem and its solution: $\varphi(\alpha) = f(1 + \alpha, -1 + \alpha) = 5\alpha^2 - 2\alpha + 4 \to \min!$
It solution: $\alpha_2 = \frac{1}{5}$
Step 3: $x_3 = x_2 + \alpha_2 s_2 = \left(\frac{6}{5}, -\frac{4}{5}\right)$,

$$s_3 = -\nabla f(x_3) = \left(\frac{1}{5}, -\frac{1}{5}\right), x = x_3 + \alpha s_3 = \left(\frac{6}{5} + \frac{1}{5}\alpha, -\frac{4}{5} - \frac{1}{5}\alpha\right).$$

The univariate optimization problem and its solution:

$$\varphi(\alpha) = f\left(\frac{6}{5} + \frac{1}{5}\alpha, -\frac{4}{5} - \frac{1}{5}\alpha\right) = \frac{1}{25}\alpha^2 - \frac{2}{25}\alpha + \frac{19}{5} \to \min!$$

Then $\alpha_3 = 1$

Step 4: $x_4 = x_3 + \alpha_3 s_3 = \left(\frac{7}{5}, -1\right)$, $s_4 = -\nabla f(x_4) = \left(\frac{1}{5}, \frac{1}{5}\right)$, $x = x_4 + \alpha s_4 = \left(\frac{7}{5} + \frac{1}{5}\alpha, -1 + \frac{1}{5}\alpha\right)$.

The univariate optimization problem and its solution:

$$\varphi(\alpha) = f\left(\frac{7}{5} + \frac{1}{5}\alpha, -1 + \frac{1}{5}\alpha\right) = \frac{1}{5}\alpha^2 - \frac{2}{25}\alpha + \frac{94}{25} \to \min!$$

then $\alpha_4 = \frac{1}{5}$

Step 5:

$$x_5 = x_4 + \alpha_4 s_4 = \left(\frac{36}{25}, -\frac{24}{25}\right),$$
$$s_5 = -\nabla f(x_4) = \left(\frac{1}{25}, -\frac{1}{25}\right),$$
$$x = x_5 + \alpha s_5 = \left(\frac{36}{25} + \frac{1}{25}\alpha, -\frac{24}{25} - \frac{1}{25}\alpha\right)$$

The univariate optimization problem and its solution:

$$\varphi(\alpha) = f\left(\frac{36}{25} + \frac{1}{25}\alpha, -\frac{24}{25} - \frac{1}{25}\alpha\right) = \frac{1}{625}\alpha^2 - \frac{2}{625}\alpha + \frac{469}{125} \to \min!$$

Then $\alpha_5 = 1$

Step 6: $x_6 = x_5 + \alpha_5 s_5 = \left(\frac{37}{25}, -1\right) = (1.48, -1)$. The process stops because the approximation in step 5 is already close to the minimum place: $x = (1.5, -1)$.

1.4. **Conjugate gradient method (Fletcher, Reeves).**

We have seen many methods to choose the direction, for instance: $s_k = -\nabla f(x_k)$ in the gradient method and $s_k = -D_k \nabla f(x_k)$ in the DFP method which we will consider latter on. The DFP method finds the solution to the secant equation that is closest to the current estimate and satisfies the curvature condition. In the conjugate gradient method the direction to used is rather

$$s_{k+1} = -\nabla f(x_{k+1}) + \alpha_k s_k.$$

There are several proposals for α_k called the step-length. Fletcher and Reeves proposed the following step-length:

$$\alpha_k = \frac{\|\nabla f(\mathbf{x}_{k+1})\|^2}{\|\nabla f(\mathbf{x}_k)\|^2}.$$

The algorithm for Fletcher-Reeves method:

(1) Starting point, $(k = 1)$: Start at any guess vector $x_k \in^n$ and any search direction $s_k = -\nabla f(x_k)$.
(2) Intermediate steps:
- Stop if $\|\nabla f(x_k)\| < \varepsilon$, otherwise the process will continue in the next line.
- Solve the minimization problem $\min\{f(x_k + \lambda s_k) : \lambda \geq 0\}$ for λ and let the optimal solution be λ_k.
- Compute the next feasible point: $x_{k+1} = x_k + \lambda_k s_k$.

- Compute the search direction: $s_{k+1} = -\nabla f(x_{k+1}) + \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2} s_k$.
- Continue, $k := k+1$.

Theorem: Suppose that the function $f(x)$ to be minimized is quadratic, i.e. $f(x) = cx + \dfrac{1}{2}xHx$. To the solution of this minimization problem apply the Fletcher-Reeves method with an arbitrary guess vector $x_1 \in \mathbb{R}^n$ and $s_1 = -\nabla f(x_1)$ as the search direction. Implement n iterations. If $\nabla f(x_k) \neq 0$ for $k = 1, 2, \ldots, n$, then the following statements are valid:

- The vectors $s_1, s_2, \ldots, s_n$ are mutually H-conjugate.
- $s_1, s_2, \ldots, s_n$ are search descent directions.
- $\alpha_k = \dfrac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2} = \dfrac{s_k^T H \nabla f(x_{k+1})}{s_k^T H s_k}$, $k = 1, 2, \ldots, n$.

1.4.1. *Example.*

Example: Solve the following optimization problem with Fletcher-Reeves method. Let $x_1 = (0, 0)$ be the starting point and let $\varepsilon = 0.005$ be a tolerence level.

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \to \min!$$

Solution:

The gradient of the function f is

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}$$

Step 1.: The first step is same as the first step of the gradient method. That is $x_1 = (0, 0)$, $s_1 = -\nabla f(x_1) = (1, -1)$, $x = x_1 + \lambda s_1 = (\lambda, -\lambda)$

We need to determine the minimum point of the function φ:

$$\varphi(\lambda) = f(\lambda, -\lambda) = \lambda^2 - 2\lambda + 5 \to \min!$$

Hence

$$\lambda_1 = 1$$

Step 2. Therefore we get $x_2 = x_1 + \lambda_1 s_1 = (\lambda_1, -\lambda_1) = (1, -1)$, $\nabla f(x_2) = (-1, -1)$.

We depart from this step to the gradient method, for determining the direction vector is more complicated. Now calculate square of the length of the gradients of the last two point:

$$\|\nabla f(\mathbf{x}_1)\|^2 = 2, \qquad \|\nabla f(\mathbf{x}_2)\|^2 = 2$$

The new direction vectior is

$$s_2 = -\nabla f(x_2) + \frac{\|\nabla f(x_2)\|^2}{\|\nabla f(x_1)\|^2} s_1 = (2, 0)$$

$$x = x_2 + \lambda s_2 = (1 + 2\lambda, -1)$$

We need to determine teh minimum point of the function φ:

$$\varphi(\lambda) = f(1 + 2\lambda, -1) = 4\lambda^2 - 2\lambda + 4 \to \min!$$

Hence

$$\lambda_2 = \frac{1}{4}$$

Step 3. We get $x_3 = x_2 + \lambda_2 s_2 = \left(\frac{3}{2}, -1\right)$, $s_3 = -\nabla f(x_3) = (0, 0)$.
Since the gradient vector is zero at the point x_3, therefore we finish the method.
In the point $x = (1.5, -1)$ satisfies the necessary condition of minimum point, se we obtained the optimal solution.

1.5. **Newton-type optimization method with line seach.**

Now consider Newton-type optimization methods extended with line seach method
Newton-method with line search method:

(1) FOR $k = 0, 1, 2, \ldots$
(2) $\qquad Hf(x_k)\, s_k = -\nabla f(x_k)$
(3) $\qquad \alpha_k > 0 : f(x_k + \alpha_k s_k) = \min_{\alpha \geq 0} f(x_k + \alpha s_k)$
(4) $\qquad x_{k+1} = x_k + \alpha_k s_k$

Modified Newton-method with line search method:

(1) FOR $k = 0, 1, 2, \ldots$
(2) $\qquad [Hf(x_k) + E_k]\, s_k = -\nabla f(x_k)$
(3) $\qquad \alpha_k > 0 : f(x_k + \alpha_k s_k) = \min_{\alpha \geq 0} f(x_k + \alpha s_k)$
(4) $\qquad x_{k+1} = x_k + \alpha_k s_k$

One of the best known Newton-type method (quasi-Newton method) which use line search is the David-Fletcher-Powell's method.

1.5.1. *DFP (Davidon-Fletcher-Powell) method.*

Let x_1 be a guess point and D_1 a posititve definite matrix.

(1) FOR $k = 1, 2, \ldots$
(2) $\qquad d_k = -D_k \nabla f(x_k)$
(3) $\qquad \lambda_k > 0 : f(x_k + \lambda_k s_k) = \min_{\alpha \geq 0} f(x_k + \alpha s_k)$
(4) $\qquad s_k = \lambda_k d_k$
(5) $\qquad x_{k+1} = x_k + s_k$
(6) $\qquad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
(7) $\qquad D_{k+1} = D_k + \dfrac{s_k s_k^T}{s_k^T y_k} - \dfrac{D_k y_k y_k^T D_k}{y_k^T D_k y_k}$

We can speak of rank-one and rank-two formulas as in BFGS-method. Rank-two formula is defined by

$$D_{k+1} = D_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{D_k y_k y_k^T D_k}{y_k^T D_k y_k}$$

and rank-one formula by

$$D_{k+1} = D_k + \frac{(s_k - D_k y_k)(s_k - D_k y_k)^T}{(s_k - D_k y_k)^T y_k}.$$

1.5.2. *Example.*

Exercise: Solve the following optimal problem with Davidon-Fletcher-Powell method

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \to \min!$$

Use rank-one formula in the computation of matrix D_k with the starting vector x_1 and the Hessian matrix D_1:

$$x_1 = (0, 0), \qquad D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Solution:
First we compute the gradient of objective function

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}$$

Step 1: This step coincid with the gradient method.

$$\begin{aligned} x_1 &= (0, 0), \\ \nabla f(x_1) &= (-1, 1), \\ d_1 &= -D_1 \nabla f(x_1) = (1, -1), \\ x &= x_1 + \lambda d_1 = (\lambda, -\lambda) \end{aligned}$$

The univariate optimization:

$$\varphi(\lambda) = f(\lambda, -\lambda) = \lambda^2 - 2\lambda + 5 \to \min!$$

Its solution: $\lambda_1 = 1$
Step 2:

$$x_2 = x_1 + \lambda_1 d_1 = (\lambda_1, -\lambda_1) = (1, -1)$$

The computation of matrix D_2.

$$
\begin{aligned}
\nabla f(\mathbf{x}_2) &= (-1,-1),\\
s_1 &= x_2 - x_1 = \lambda_1 d_1 = (1,-1),\\
y_1 &= \nabla f(x_2) - \nabla f(x_1) = (0,-2)\\
D_1 y_1 &= (0,-2)\\
s_1 s_1^T &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\\
s_1^T y_1 &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = 2\\
(D_1 y_1)(D_1 y_1)^T &= \begin{bmatrix} 0 \\ -2 \end{bmatrix} \begin{bmatrix} 0 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}\\
(D_1 y_1)^T y_1 &= \begin{bmatrix} 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = 4\\
D_2 = D_1 + \frac{s_1 s_1^T}{s_1^T y_1} - \frac{(D_1 y_1)(D_1 y_1)^T}{(D_1 y_1)^T y_1} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\\
&= \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

The direction vector corresponding to x_2: $d_2 = -D_2 \nabla f(x_2) = (1,0)$

$$x = x_2 + \lambda d_2 = (1+\lambda, -1).$$

The univariate optimization:

$$\varphi(\lambda) = f(1+\lambda, -1) = \lambda^2 - \lambda + 4 \to \min!$$

Its solution: $\lambda_2 = \frac{1}{2}$
Step 3:

$$x_3 = x_2 + \lambda_2 d_2 = (1+\lambda_2, -1) = \left(\frac{3}{2}, -1\right)$$

Computation of matrix D_3:

$$\nabla f(x_3) = (0,0).$$

Since the gradient corresponding to x_3 is zero, the procedure stops. The minimum condition is met in the feasible point $x = (1.5, -1)$, thus the optimal solution is attained.

Since the function to minimize is is quadratic, we can continue the computations to verify the conditions of the above theorem.

$$
\begin{aligned}
s_2 &= x_3 - x_2 = \lambda_2 d_2 = \left(\frac{1}{2}, 0\right), \\
y_2 &= \nabla f(x_3) - \nabla f(x_2) = (1, 1), \\
D_2 y_2 &= (1, 0) \\
s_2 s_2^T &= \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{bmatrix} \\
s_2^T y_2 &= \begin{bmatrix} \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2} \\
(D_2 y_2)(D_2 y_2)^T &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
(D_2 y_2)^T y_2 &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \\
D_3 &= D_2 + \frac{s_2 s_2^T}{s_2^T y_2} - \frac{(D_2 y_2)(D_2 y_2)^T}{(D_2 y_2)^T y_2} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

The Hessian matrix of objective function:

$$
\mathbf{H} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}
$$

and we can check that D_3 is the inverse of the Hessian matrix.
Note that the Davidon-Fletcher-Powell method is one of the most used method.