Lecture 8

Numerical methods for solving optimization problems

Conjugate gradient method

The conjugate gradient method combines iterative methods and gradient descent methods. It is based on the idea of finding the optimal direction in the parameter space that minimizes the function. This direction is called the conjugate gradient. **Conjugate gradients are gradients of a function that are orthogonal to each other. They allow you to move along function level lines, minimizing it faster than simple gradient descent.**

Conjugate gradient method

The working principle of the conjugate gradient method can be divided into the following steps:

1. **Initialization**

First, you need to select an initial approximation for the optimal solution. Usually the initial value is chosen randomly or using some heuristic methods.

2. Gradient calculation

Next, you need to calculate the gradient of the function at the current point.

Conjugate gradient method

3. Calculating the conjugate direction

The conjugate direction is calculated based on the previous conjugate direction and the gradient of the function. This direction is chosen so that it is orthogonal to the previous conjugate direction and the gradient of the function.

4. Step calculation

The optimization step is calculated based on the conjugate direction and gradient of the function. It determines how far one must move along the conjugate direction to reach the next optimal solution point.

Conjugate gradient method

5. Update current point

The current optimal solution point is updated by moving along the conjugate direction by a certain step. This allows you to get closer to the minimum function.

6. Checking the Stop Condition

After updating the current point, the stopping condition is checked. If certain convergence criteria are achieved, the optimization process ends. Otherwise, the process repeats from step 2.

Conjugate gradient method

We construct a sequence of directions:
not just the negative gradient, but **conjugate directions**

- they are «orthogonal» in a special sense (with respect to matrix A)
- so we do not go back over previous progress

Conjugate gradient method

1. Initialization
Choose x_0, compute gradient $g_0 = \nabla f(x_0)$, initial direction $d_0 = -g_0$
At each iteration

a. **Line search** $x_{k+1} = x_k + \alpha_k d_k$ ($\alpha_k$is chosen to minimize the function along the direction)
b. **New gradient** $g_{k+1} = \nabla f(x_{k+1})$
c. **Compute** β_k
d. **New direction** $d_{k+1} = -g_{k+1} + \beta_k d_k$

3. Stopping condition $\| g_k \| < \varepsilon$

Conjugate gradient method

Properties of the conjugate gradient method

•Finite number of iterations

The conjugate gradient method guarantees convergence to the optimal solution in a finite number of iterations for quadratic functions.

•Efficiency in practice

The conjugate gradient method shows good performance in practice, especially for high-dimensional problems. It can significantly speed up the optimization process compared to other methods such as gradient descent.

Conjugate gradient method

Properties of the conjugate gradient method

- Noise resistance

The conjugate gradient method is robust to noise in the data. This means that it can continue to converge to the optimal solution even if the gradients contain some level of noise or inaccuracy.

- Suitable for large tasks

The conjugate gradient method scales well for large problems because it does not require storing the full Hessian or gradient matrix in memory. Instead, it uses only current and previous information about gradients and directions.

Conjugate gradient method

Advantages of the conjugate gradient method:

1. Efficiency.
2. Low Memory Requirement.
3. Guaranteed convergence.
4. Applicability to a large class of problems.

Conjugate gradient method

Scheme of the conjugate gradient method algorithm for $f(x) = \frac{1}{2}x^T Ax + b^T x + c$:

1. $k = 0$
2. x^0 – initial value, $g_0 = \nabla f(x^0) = Ax^0 + b, d_0 = -g_0, k = 0$ (the first step is like gradient descent, each subsequent step is constructed so as not to spoil the reduction already achieved, for this purpose, conjugate directions are used)
3. Find $x^{k+1} = x^k + \lambda_k d_k$ where

$$\lambda_k = -\frac{g_k^T d_k}{d_k^T A d_k} = -\frac{\nabla f_k^T d_k}{d_k^T A d_k}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}$$

Conjugate gradient method

β_k we find from condition $d_{k+1}Ad_k = 0$
4. $k = k + 1$, go to step 2
The criterion for stopping a one-dimensional search along each direction dk is written as $\nabla f\left(x^{k+1}\right)d_k = 0$.

Conjugate gradient method

Example. Find the local minimum of a function

$$f(x) = 2x_1^2 + x_1 x_2 + x_2^2$$

Conjugate gradient method

$$f(x) = 2x_1^2 + x_1 x_2 + x_2^2.$$

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 4x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix}.$$

$$A = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, g(x) = Ax.$$

Conjugate gradient method

The initial point $x^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Iteration 0. $g_0 = \nabla f\left(x^{(0)}\right) = Ax^{(0)}$.

$$A \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \cdot 1 + 1 \cdot 1 \\ 1 \cdot 1 + 2 \cdot 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

Therefore,

$$g_0 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

Conjugate gradient method

The first search direction $d_0 = -g_0$. So, $d_0 = -\begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} -5 \\ -3 \end{pmatrix}$.

$$\lambda_0 = -\frac{g_0^T d_0}{d_0^T A d_0}.$$

Compute $g_0^T d_0$

$$g_0^T d_0 = (5 \quad 3)\begin{pmatrix} -5 \\ -3 \end{pmatrix} = 5 \cdot (-5) + 3 \cdot (-3) = -25 - 9 = -34.$$

Compute $A d_0$

$$A d_0 = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -5 \\ -3 \end{pmatrix} = \begin{pmatrix} 4(-5) + 1(-3) \\ 1(-5) + 2(-3) \end{pmatrix} = \begin{pmatrix} -20 - 3 \\ -5 - 6 \end{pmatrix} = \begin{pmatrix} -23 \\ -11 \end{pmatrix}.$$

Conjugate gradient method

Compute $d_0^T A d_0$

$$d_0^T A d_0 = \begin{pmatrix} -5 & -3 \end{pmatrix} \begin{pmatrix} -23 \\ -11 \end{pmatrix} = (-5)\cdot(-23) + (-3)\cdot(-11) = 115 + 33 = 148.$$

So

$$d_0^T A d_0 = 148.$$

Now compute λ_0

$$\lambda_0 = -\frac{-34}{148} = \frac{34}{148} = \frac{17}{74}.$$

Thus,

$$\lambda_0 = \frac{17}{74}.$$

Conjugate gradient method

Compute the new point $x^{(1)}$

$$x^{(1)} = x^{(0)} + \lambda_0 d_0.$$

Substitute:

$$x^{(1)} = \binom{1}{1} + \frac{17}{74}\binom{-5}{-3} = \begin{pmatrix} 1 - \frac{85}{74} \\ 1 - \frac{51}{74} \end{pmatrix} = \begin{pmatrix} -\frac{11}{74} \\ \frac{23}{74} \end{pmatrix}.$$

Conjugate gradient method

Compute the gradient $g_1 = \nabla f\left(x^{(1)}\right) = Ax^{(1)}$.

Therefore,

$$g_1 = \begin{pmatrix} -\frac{21}{74} \\ \frac{35}{74} \end{pmatrix}.$$

Compute $\beta_0 = \frac{g_1^T A d_0}{d_0^T A d_0}$.

We already know

$$Ad_0 = \begin{pmatrix} -23 \\ -11 \end{pmatrix}, d_0^T A d_0 = 148.$$

Conjugate gradient method

Now compute $g_1^T A d_0$:

$$g_1^T A d_0 = \begin{pmatrix} -\frac{21}{74} & \frac{35}{74} \end{pmatrix} \begin{pmatrix} -23 \\ -11 \end{pmatrix} = \frac{49}{37}.$$

$\beta_0 = \frac{49}{5476} \approx 0.008947.$

Compute the new direction $d_1 = -g_1 + \beta_0 d_0$.

Numerically,

$$d_1 \approx \begin{pmatrix} 0.2390 \\ -0.4998 \end{pmatrix}.$$

Conjugate gradient method

Compute $\lambda_1 = -\frac{g_1^T d_1}{d_1^T A d_1}$.

$$\lambda_1 \approx -\frac{-0.3052}{0.4890} \approx 0.624.$$

Compute the new point $x^{(2)} = x^{(1)} + \lambda_1 d_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Conjugate gradient method

Iteration 2

Now we compute the gradient:

$$g_2 = \nabla f\left(x^{(2)}\right) = Ax^{(2)} = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

So

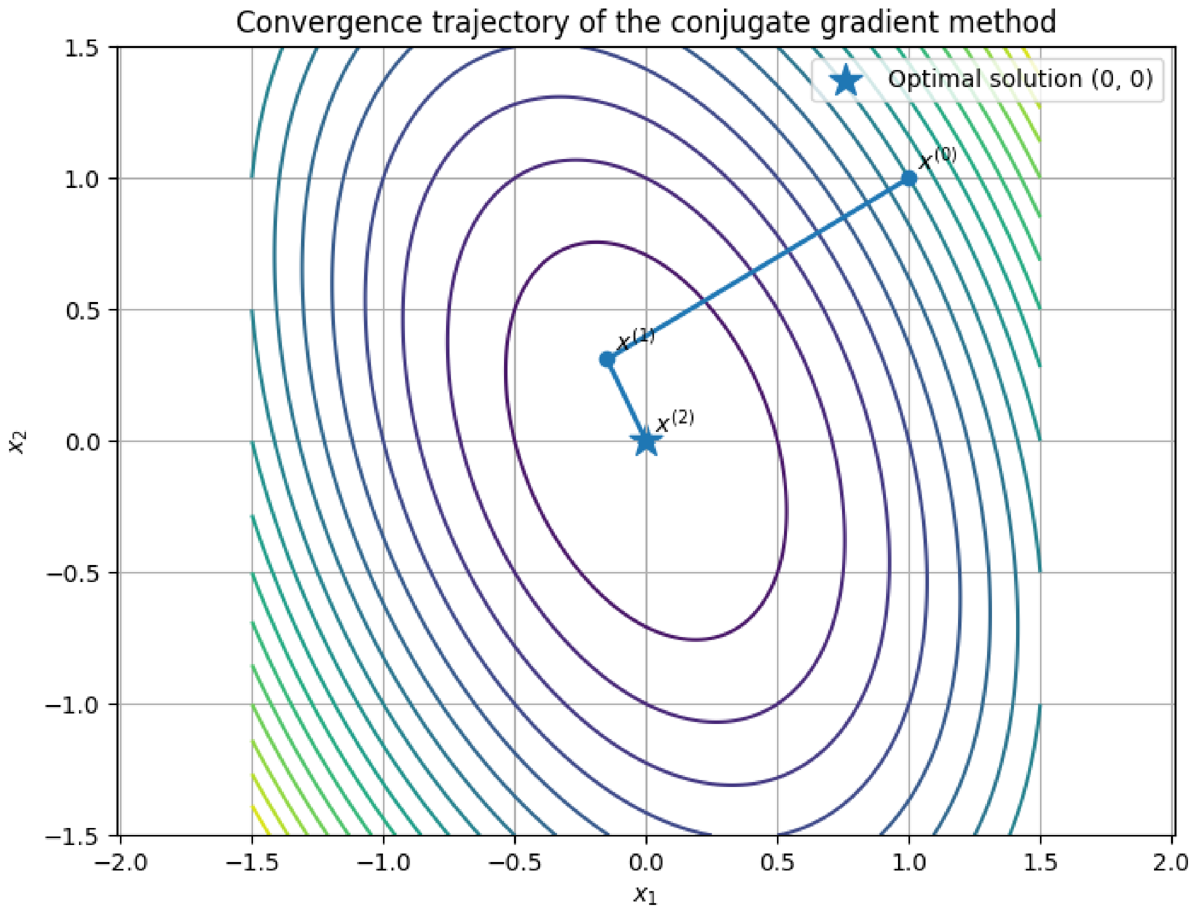
$$g_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since the gradient is zero, we have reached the minimum.

The exact minimizer is

$$x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, f(x^*) = 0.$$

Conjugate gradient method



Fletcher–Reeves Method

Let the function $f(x)$ bounded from below on the set R^n and having continuous partial derivatives at all its points be given.
It is required to find the local minimum x^* of the function $f(x)$ on the set of admissible solutions:

$$f(x^*) = \min_{x \in R^n} f(x)$$

Fletcher–Reeves Method

The strategy of the Fletcher-Reeves method is to construct a sequence of points $\{x^k\}, k = 0,1,2, \ldots$ such that $f\left(x^{k+1}\right) < f\left(x^k\right), k = 0,1,2, \ldots$
The points of the sequence $\{x^k\}$ are calculated according to the rule:

$$x^{k+1} = x^k - t_k d_k, k = 0,1,2, \ldots$$

$$d_k = \nabla f\left(x^k\right) + b_{k-1} \nabla f\left(x^{k-1}\right)$$

$$b_{k-1} = \frac{\left\|\nabla f\left(x^k\right)\right\|^2}{\|\nabla f(x^{k-1})\|^2}$$

Fletcher–Reeves Method

The step size is selected from the condition of the minimum of the function $f(x)$ over t in the direction of movement, i.e., as a result of solving the one-dimensional minimization problem:

$$f\left(x^k - t_k d_k\right) \to \min \ (t_k > 0)$$

Fletcher–Reeves Method

Fletcher-Reeves method algorithm:

1. $x^0, \varepsilon_1 > 0$ $(gradient\ accuracy), \varepsilon_2 > 0, M -$ limit number of iterations. Find the gradient $\nabla f(x)$.
2. $k = 0$
3. Find $\nabla f\left(x^k\right)$
4. Check completion criteria $\left\|\nabla f\left(x^k\right)\right\| < \varepsilon_1$: if the criterion is met, then the calculation is completed and $x^* = x^k$, if not, then go to step 5.
5. Check condition $k \geq M$ if the criterion is met, then the calculation is completed and $x^* = x^k$, if not, then for $k = 0$ go to step 6, and for $k \geq 1$ go to step 7.

Fletcher–Reeves Method

6. Find $d^0 = -\nabla f(x^0)$
7. Find

$$\beta_{k-1} = \frac{\left\| \nabla f\left(x^k\right) \right\|^2}{\left\| \nabla f\left(x^{k-1}\right) \right\|^2}$$

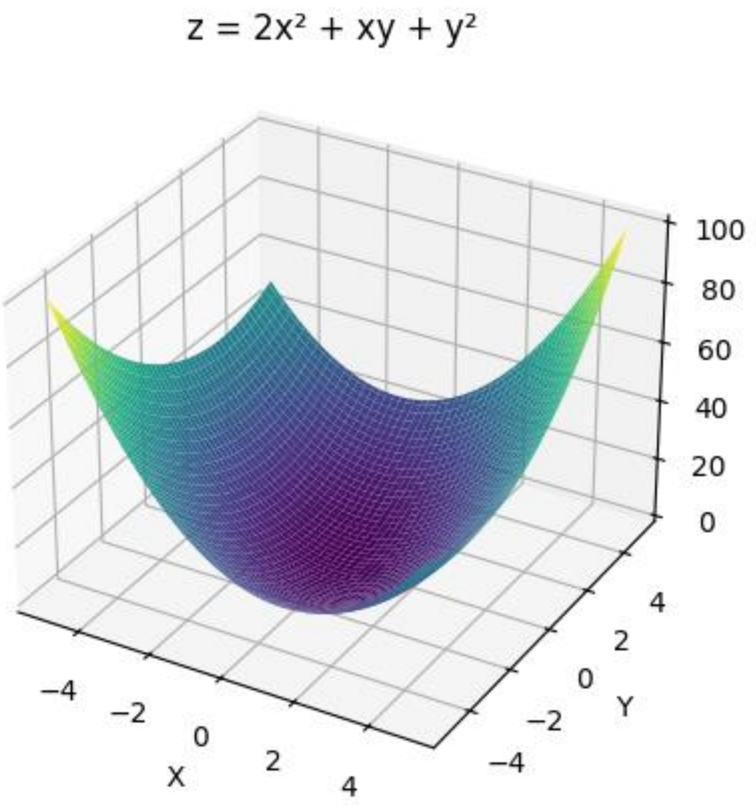
8. Find $d^k = -\nabla f\left(x^k\right) + \beta_{k-1} d^{k-1}$
9. Find t_k^* from condition $\varphi(t_k) = f\left(x^k + t_k d^k\right) \to min$
10. Find $x^{k+1} = x^k + t_k^* d^k$
11. Check if conditions are met $\left\|x^{k+1} - x^k\right\| < \varepsilon_2, \left|f\left(x^{k+1}\right) - f\left(x^k\right)\right| < \varepsilon_2$

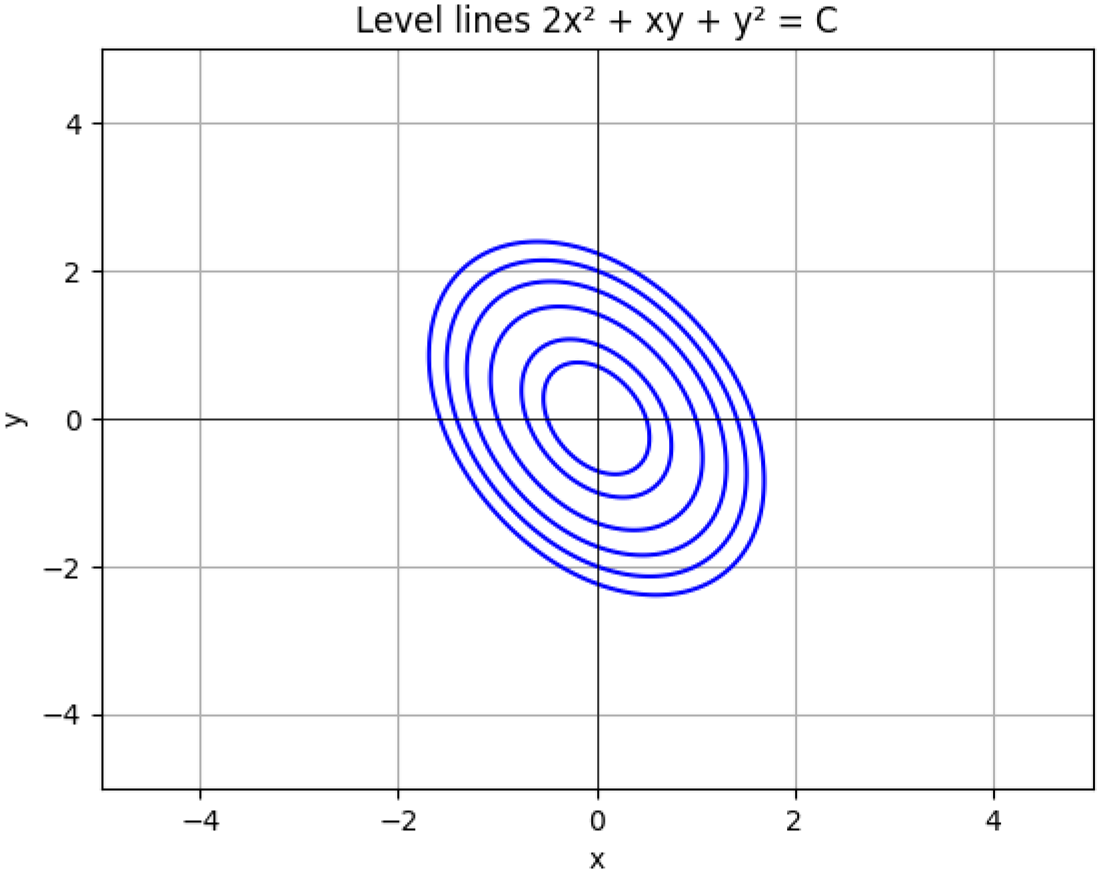
Fletcher–Reeves Method

Example. Find the local minimum of a function

$$f(x) = 2x_1^2 + x_1x_2 + x_2^2$$

Fletcher–Reeves Method





Fletcher–Reeves Method

Gradient:

$$\nabla f(x) = \begin{pmatrix} 4x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix}$$

Initialization

Choose: $x^{(0)} = (1; 1)$

Compute: $g_0 = \nabla f\left(x^{(0)}\right) = (5; 3)$

Initial direction: $d_0 = -g_0 = (-5; -3)$

Fletcher–Reeves Method

Iteration 0

Step size t_0. We minimize along direction d_0:

$$\varphi(t) = f\left(x^{(0)} + td_0\right)$$

Substitute $x_1 = 1 - 5t, x_2 = 1 - 3t$

Then:

$$\varphi(t) = 2(1 - 5t)^2 + (1 - 5t)(1 - 3t) + (1 - 3t)^2$$

Differentiate and solve:

$$\varphi'(t) = 0 \Rightarrow t_0 \approx 0.2297$$

New point $x^{(1)} = x^{(0)} + t_0 d_0 \approx (1,1) + 0.2297(-5, -3) \approx (-0.1486\ \ 0.3108)$

Fletcher–Reeves Method

Iteration 1
Gradient $g_1 = \nabla f\left(x^{(1)}\right) \approx (0.2836 \;\; 0.4730)$

Fletcher–Reeves coefficient $\beta_0 = \frac{\|g_1\|^2}{\|g_0\|^2} \approx \frac{1.268}{34} \approx 0.0373$
New direction $d_1 = -g_1 + \beta_0 d_0 \approx (0.368; -0.673)$

Step size t_1
Minimize $\varphi(t) = f\left(x^{(1)} + td_1\right)$

$$\varphi'(t) = 0 \Rightarrow t_1 \approx 0.595$$

New point

$$x^{(2)} = x^{(1)} + t_1 d_1$$
$$x^{(2)} \approx (-0.1486 \;\; 0.3108) + 0.595(0.368; -0.673)$$
$$x^{(2)} \approx (0.001; 0)$$

Fletcher–Reeves Method

Iteration 2

Gradient $g_2 = \nabla f\left(x^{(2)}\right);\ g_2 \approx (0.003; 0.006)$

Norm of gradient $\| g_2 \| = \sqrt{0.003^2 + 0.006^2} \approx 0.0067$

Stopping criterion $\| g_2 \| < \varepsilon_1$ satisfied

Final Answer

$$x^* \approx (0.001; 0)$$
$$f(x^*) \approx 2 \cdot 10^{-6}$$

Fletcher–Reeves Method