Methods

1 General formulation

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t. } g_i(x) &\leq 0, \; i = 1, \dots, m \\
h_j(x) &= 0, \; j = 1, \dots, k
\end{aligned}
$$

Some necessary or/and sufficient conditions are known (See Optimality conditions. KKT and Convex optimization problem.

- In fact, there might be very challenging to recognize the convenient form of optimization problem.
- Analytical solution of KKT could be inviable.

1.1 Iterative methods

Typically, the methods generate an infinite sequence of approximate solutions

$$\{x_t\},$$

which for a finite number of steps (or better - time) converges to an optimal (at least one of the optimal) solution x_*.

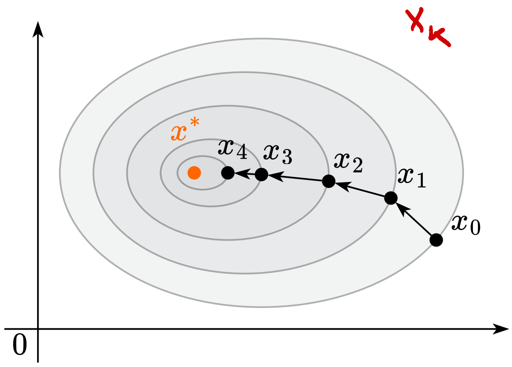
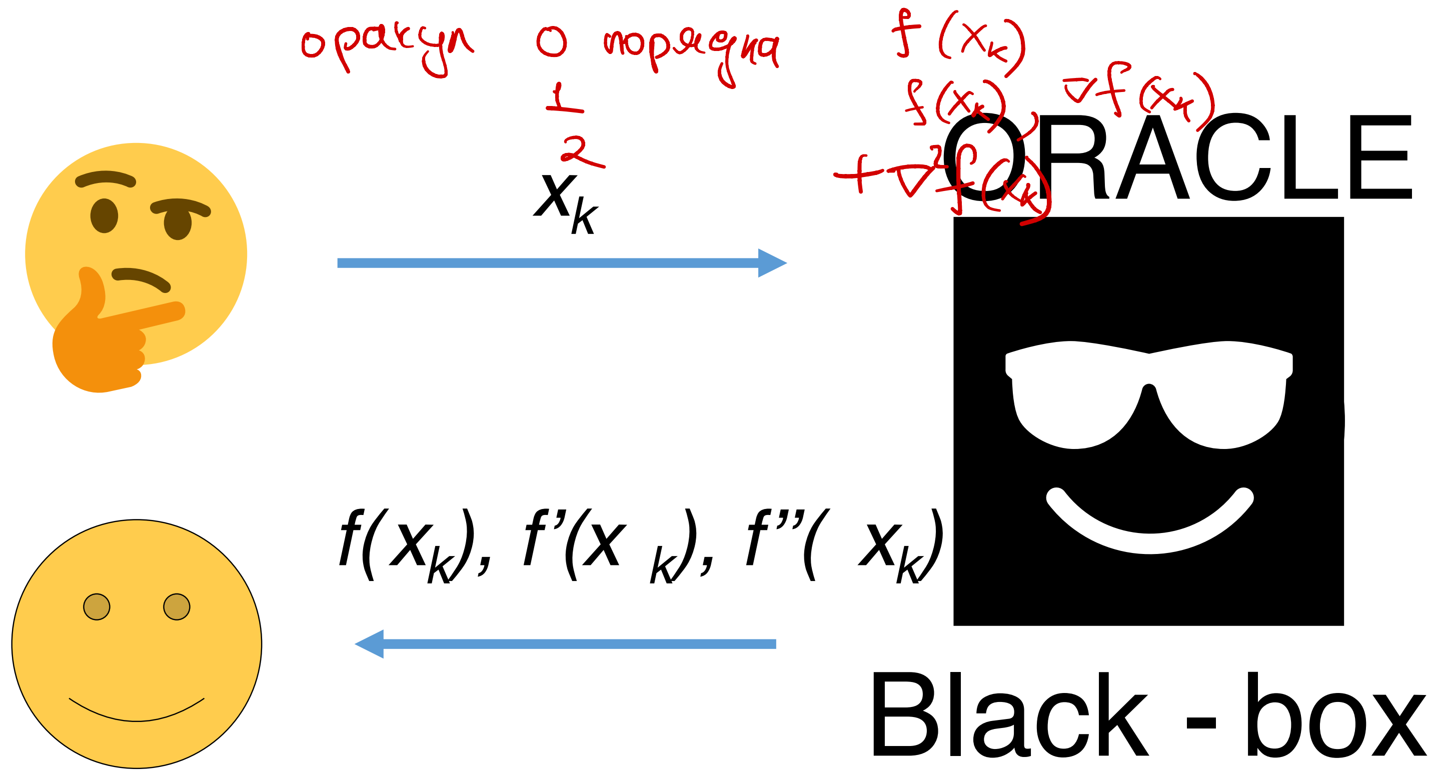


Illustration of iterative method approaches to the solution x^*

```
def GeneralScheme(x, epsilon):
    while not StopCriterion(x, epsilon):
        OracleResponse = RequestOracle(x)
        x = NextPoint(x, OracleResponse)
    return x
```

1.2 Oracle conception



Depending on the maximum order of derivative available from the oracle we call the oracles as zero order, first order, second order oravle and etc.

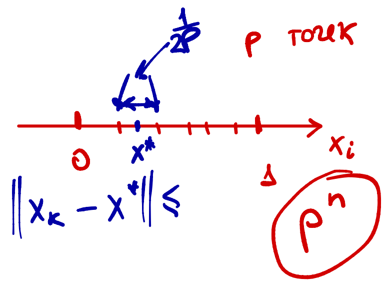
2 Unsolvability of numerical optimization problem

$\min\limits_{x \in \mathbb{C}^n} f(x)$

In general, **optimization problems are unsolvable.** ¯\\(ツ)/¯

Consider the following simple optimization problem of a function over unit cube:

$$
\begin{split}
& \min_{x \in \mathbb{R}^n} f(x) \\
\text{s.t. } & x \in \mathbb{C}^n
\end{split}
$$



We assume, that the objective function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous on $\overset{\mathbb{C}}{\cancel{\mathbb{R}}}{}^n$:

$$
|f(x) - f(y)| \leq L\|x - y\|_\infty \forall x, y \in \mathbb{C}^n,
$$

with some constant L (Lipschitz constant). Here $\mathbb{C}^n$ - the n-dimensional unit cube

$$
\mathbb{C}^n = \{x \in \mathbb{R}^n \mid 0 \leq x_i \leq 1, i = 1, \dots, n\}
$$

Our goal is to find such $\tilde{x} : |f(\tilde{x}) - f^*| \leq \varepsilon$ for some positive ε. Here f^* is the global minima of the problem. Uniform grid with p points on each dimension guarantees at least this quality:

$$
\|\tilde{x} - x_*\|_\infty \leq \frac{1}{2p},
$$

which means, that

$$
|f(\tilde{x}) - f(x_*)| \leq \frac{L}{2p} = \varepsilon
$$

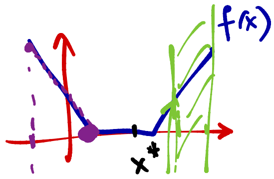
$$
p = \frac{L}{2\varepsilon}
$$

Our goal is to find the p for some ε. So, we need to sample $\left(\frac{L}{2\varepsilon}\right)^n$ points, since we need to measure function in p^n points. Doesn't look scary, but if we'll take $L = 2, n = 11, \varepsilon = 0.01$, computations on the modern personal computers will take 31,250,000 years.

~ 3 млн лет.

2.1 Stopping rules

- Argument closeness:

- Function value closeness: 

- Closeness to a critical point

$$\boxed{\|x_k - x_*\|_2 < \varepsilon}$$

$$\boxed{\|f_k - f^*\|_2 < \varepsilon}$$

$$\boxed{\|f'(x_k)\|_2 < \varepsilon}$$

But x_* and $f^* = f(x_*)$ are unknown!

Sometimes, we can use the trick:

$$\boxed{\|x_{k+1} - x_k\|} = \|x_{k+1} - x_k + x_* - x_*\| \leq \overset{\le\varepsilon}{\|x_{k+1} - x_*\|} + \overset{\le\varepsilon}{\|x_k - x_*\|} \leq 2\varepsilon$$

Note: it's better to use relative changing of these values, i.e. $\boxed{\dfrac{\|x_{k+1} - x_k\|_2}{\|x_k\|_2}}$.

> **Example**
>
> Suppose, you are trying to estimate the vector x_{true} with some approximation x_{approx}. One can choose between two relative errors:
>
> $$\frac{\|x_{approx} - x_{true}\|}{\|x_{approx}\|} \quad \frac{\|x_{approx} - x_{true}\|}{\|x_{true}\|}$$
>
> If both x_{approx} and x_{true} are close to each other, then the difference between them is small, while if your approximation is far from the truth (say, $x_{approx} = 10x_{true}$ or $x_{approx} = 0.01x_{true}$ they differ drastically).

2.2 Local nature of the methods

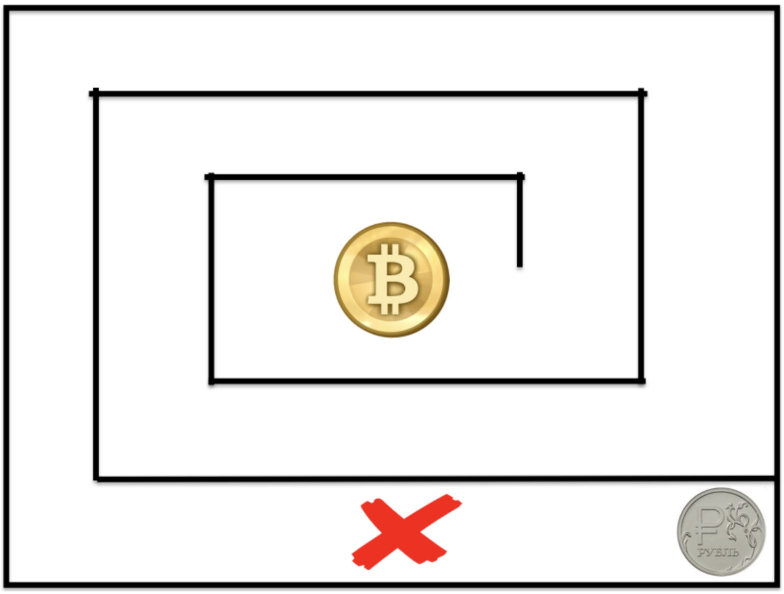
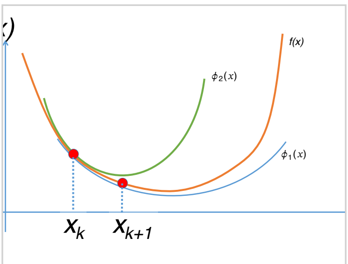
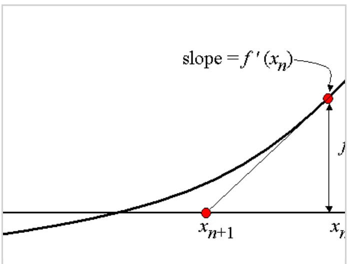


Illustration of the idea of locality in black-box optimization

3 Contents of the chapter

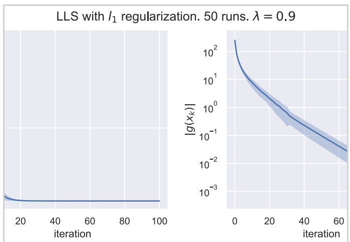


Gradient descent

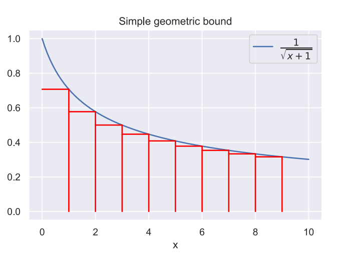


Newton method

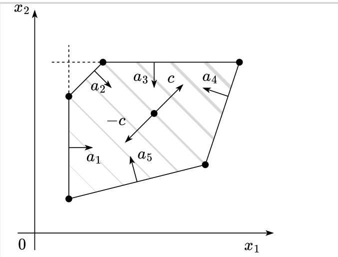
Quasi Newton methods



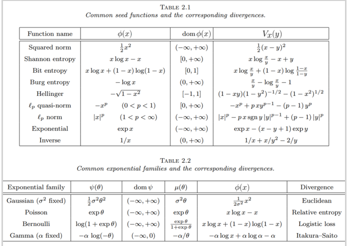
Subgradient descent



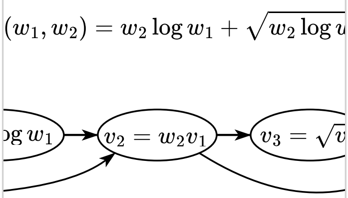
Projected subgradient descent



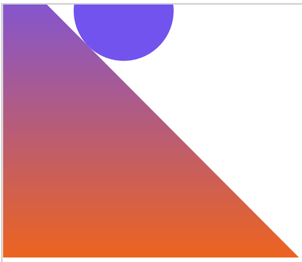
Linear Programming and simplex algorithm



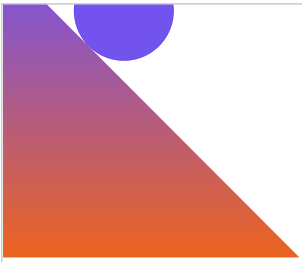
Mirror descent



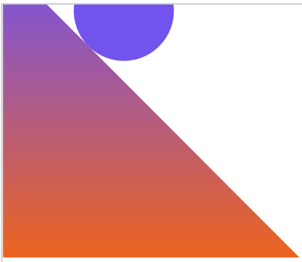
Automatic differentiation



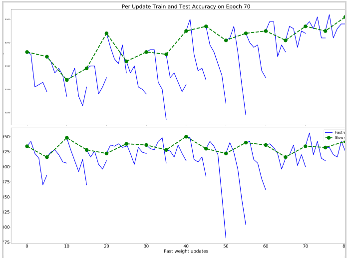
Stochastic gradient descent



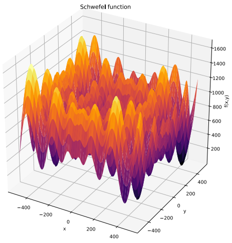
Stochastic average gradient



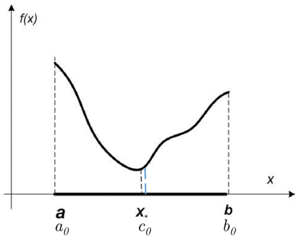
ADAM: A Method for Stochastic Optimization



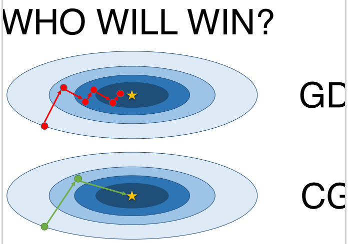
Lookahead Optimizer: k steps forward, 1 step back



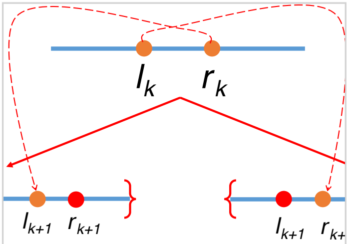
Bee algorithm



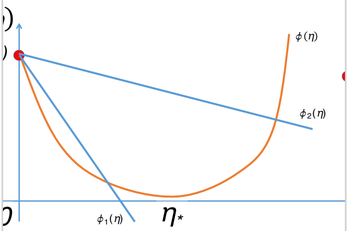
Binary search



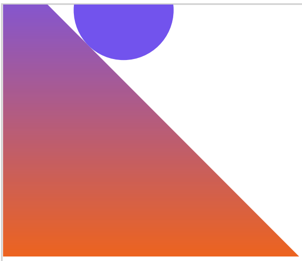
Conjugate gradients



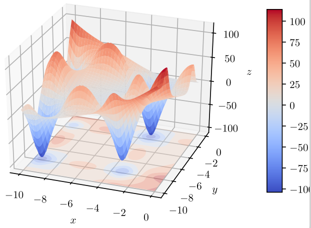
Golden search



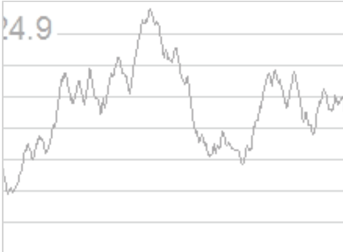
Inexact line search



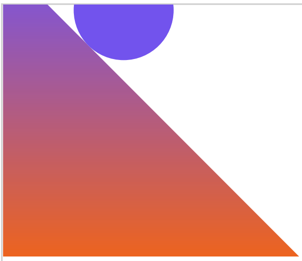
Natural gradient descent



Nelder–Mead



Simulated annealing

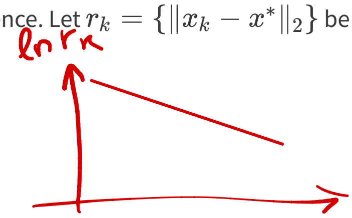


Successive parabolic interpolation

Rates of convergence

1 Speed of convergence

In order to compare perfomance of algorithms we need to define a terminology for different types of convergence. Let $r_k = \{\|x_k - x^*\|_2\}$ be a sequence in $\mathbb{R}^n$ that converges to zero.



1.1 Linear convergence

$r_{k+1} \leq q \cdot r_k$

We can define the *linear* convergence in a two different forms:

$$
\boxed{\|x_{k+1} - x^*\|_2 \leq Cq^k} \quad \text{or} \quad \|x_{k+1} - x^*\|_2 \leq q\|x_k - x^*\|_2,
$$

for all sufficiently large k. Here $q \in (0, 1)$ and $0 < C < \infty$. This means that the distance to the solution x^* decreases at each iteration by at least a constant factor bounded away from 1. Note, that sometimes this type of convergence is also called *exponential* or *geometric*. We call the q the convergence rate.

Question

Suppose, you have two sequences with linear convergence rates $q_1 = 0.1$ and $q_2 = 0.7$, which one is faster?

Example

Let us have the following sequence:

$$
\boxed{r_k = \frac{1}{3^k}}
$$

One can immediately conclude, that we have a linear convergence with parameters $q = \frac{1}{3}$ and $C = \cancel{0}\ 1$.

Question

Let us have the following sequence:

$$
\boxed{r_k = \frac{4}{3^k}}
$$

$q = \frac{1}{3}$

$C = 4$

Will this sequence be convergent? What is the convergence rate?

1.2 Sublinear convergence

If the sequence r_k converges to zero, but does not have linear convergence, the convergence is said to be sublinear. Sometimes we can considet the following class of sublinear convergence:

L

$$
\boxed{\|x_{k+1} - x^*\|_2 \leq Ck^q,}
$$

where $q < 0$ and $0 < C < \infty$. Note, that sublinear convergence means, that the sequence is converging slower, than any geometric progression.

1.3 Superlinear convergence

$q_k \to 0$

The convergence is said to be *superlinear* if:

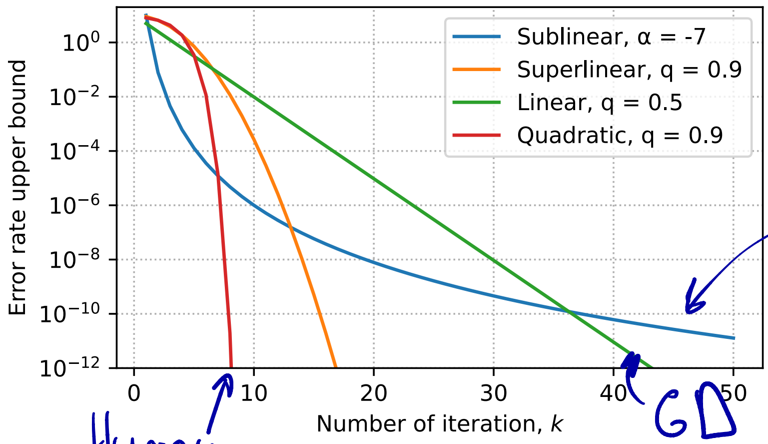
$$
\|x_{k+1} - x^*\|_2 \leq Cq^{k^2} \quad \text{or} \quad \boxed{\|x_{k+1} - x^*\|_2 \leq q_k\|x_k - x^*\|_2,}
$$

where $q \in (0, 1)$ or $0 < C_k < \infty$, $C_k \to 0$. Note, that superlinear convergence is also linear convergence (one can even say, that it is linear convergence with $q = 0$).

1.4 Quadratic convergence

$$\|x_{k+1} - x^*\|_2 \leq Cq^{2^k} \quad \text{or} \quad \|x_{k+1} - x^*\|_2 \leq C\|x_k - x^*\|_2^2,$$

where $q \in (0, 1)$ and $0 < C < \infty$.



GD для вып. функций

GD для сильновып. функций

Ньютон для сильновып. ф-ций

Difference between the convergence speed

Quasi-Newton methods for unconstrained optimization typically converge superlinearly, whereas Newton's method converges quadratically under appropriate assumptions. In contrast, steepest descent algorithms converge only at a linear rate, and when the problem is ill-conditioned the convergence constant q is close to 1.

2 How to determine convergence type

2.1 Root test

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of non-negative numbers, converging to zero, and let

предел сверху

$$q = \limsup_{k\to\infty} r_k^{1/k}$$

- If $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.
- If $q = 1$, then $\{r_k\}_{k=m}^{\infty}$ has sublinear convergence.
- The case $q > 1$ is impossible.

пример

$r_k \ \boxed{\frac{1}{k^k}} = ?$ сверхлинейная сх-ть

Тест корней

$$\lim_{k\to\infty} \left(r_k^{\frac{1}{k}} = \left(\left(\frac{1}{k}\right)^k\right)^{\frac{1}{k}}\right) = \frac{1}{k} = 0$$

$$r_k = \left(\frac{1}{100500}\right)^k$$

$$r_k^{\frac{1}{k}} = \frac{1}{100500} = q$$

2.2 Ratio test

Тест отношений

Let $\{r_k\}_{k=m}^{\infty}$ be a sequence of strictly positive numbers converging to zero. Let

$$q = \lim_{k\to\infty} \frac{r_{k+1}}{r_k}$$

- If there exists q and $0 \leq q < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with constant q.
- In particular, if $q = 0$, then $\{r_k\}_{k=m}^{\infty}$ has superlinear convergence.
- If q does not exist, but $q = \limsup_{k\to\infty} \frac{r_{k+1}}{r_k} < 1$, then $\{r_k\}_{k=m}^{\infty}$ has linear convergence with a constant not exceeding q.
- If $\liminf_{k\to\infty} \frac{r_{k+1}}{r_k} = 1$, then $\{r_k\}_{k=m}^{\infty}$ has sublinear convergence.
- The case $\liminf_{k\to\infty} \frac{r_{k+1}}{r_k} > 1$ is impossible.
- In all other cases (i.e., when $\liminf_{k\to\infty} \frac{r_{k+1}}{r_k} < 1 \leq \limsup_{k\to\infty} \frac{r_{k+1}}{r_k}$) we cannot claim anything concrete about the convergence rate $\{r_k\}_{k=m}^{\infty}$.

На дом:

Example

Let us have the following sequence:

$$r_k = \frac{1}{k}$$

root

$$\lim_{k\to\infty} \left(\frac{1}{k}\right)^{\frac{1}{k}} = \lim_{k\to\infty} \left(\frac{1}{k} + 1 - 1\right)^{\frac{1}{k}} = \lim_{k\to\infty} \left(\frac{k+1}{k} - 1\right)^{\frac{1}{k}}$$

ratio:

$$r_{k+1} = \frac{1}{k+1}$$

$$r_k = \frac{1}{k}$$

$$\frac{r_{k+1}}{r_k} = \frac{k}{k+1} = 1$$

сублин.

Determine the convergence

Example

субл.

Let us have the following sequence:

$$\frac{r_{k+1}}{r_k} = \frac{k^2}{(k+1)^2} = \left(\frac{k}{k+1}\right)^2 = 1$$

$$r_k = \frac{1}{k^2} \quad 2 \quad \frac{1}{k^{100500}}$$

Determine the convergence

Example

Let us have the following sequence:

$$r_k = \frac{1}{k^q}, q > 1$$

Determine the convergence

Try to use root test here

Let us have the following sequence:

$$r_k = \frac{1}{k^k}$$

Determine the convergence

3 References

- Code for convergence plots - [Open In Colab]
- [CMC seminars (ru)]
- Numerical Optimization by J.Nocedal and S.J.Wright