From Least Squares to Quadratic Programming, Convexity

Computational Intelligence, Lecture 3

by Sergei Savin

Spring 2023

Problem 1. minimize $||\mathbf{x}||$.

Problem 2. minimize $||\mathbf{A}\mathbf{x}||$.

Problem 3. minimize $||\mathbf{A}\mathbf{x} + \mathbf{b}||$.

We know solution of minimize $||\mathbf{A}\mathbf{x} - \mathbf{b}||$, which is $\mathbf{x} = \mathbf{A}^{+}\mathbf{b}$.
Therefore the problem 3 has a solution $\mathbf{x} = -\mathbf{A}^{+}\mathbf{b}$.

Problem 4.

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{x}||, \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{c}.
\end{aligned}
\tag{1}
$$

All solutions to $\mathbf{A}\mathbf{x} = \mathbf{c}$ are written as $\mathbf{x} = \mathbf{A}^{+}\mathbf{c} + \mathbf{N}\mathbf{z}$, where $\mathbf{N} = \text{null}(\mathbf{A})$, and $\mathbf{A}^{+}\mathbf{c} \in \text{row}(\mathbf{A})$ as we proved previously. Since null space solution $\mathbf{N}\mathbf{z}$ and row space paricular solution $\mathbf{A}^{+}\mathbf{c}$ are orthagonal, the minimum norm solution corresponds to $\mathbf{z} = \mathbf{0}$, hence $\mathbf{x} = \mathbf{A}^{+}\mathbf{c}$.

Thus, the solution is $\mathbf{x} = \mathbf{A}^{+}\mathbf{c}$. Notice that solutions for the problem 4 and problem 3 are written identically (sans the sign), even though problem 3 asks us to minimize residual of the linear system, while problem 4 - find minimum norm solution.

This illustrates an important fact that solution to the least squares problem, formulated either as "minimization of a residual" or as a "minimum norm solution" are given by the same formula, which we call Moore-Penrose pseudoinverse.

Problem 5.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{2}$$

One way to think about it is to first find all solution to the constraint equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ and then find optimal one among them. As we know, all solutions are given as: $\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + \mathbf{N}\mathbf{z}$, where $\mathbf{N} = \text{null}(\mathbf{A})$. Then our cost function becomes: $||\mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{D}\mathbf{N}\mathbf{z}||$, which is equivalent to the problem 3. Thus, we can write solution as: $\mathbf{z}^{*} = -(\mathbf{D}\mathbf{N})^{+}\mathbf{D}\mathbf{A}^{+}\mathbf{b}$. In terms of $\mathbf{x}$ solution is:

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}\mathbf{D}\mathbf{A}^{+}\mathbf{b} \tag{3}$$

$$\mathbf{x}^{*} = (\mathbf{I} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}\mathbf{D})\mathbf{A}^{+}\mathbf{b} \tag{4}$$

Problem 6.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x} + \mathbf{f}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{5}$$

After the same initial step, we arrive at the cost function $||\mathbf{D}\mathbf{N}\mathbf{z} + \mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{f}||$. It is only different in the constant term, and the solution is found as follows:

$$\mathbf{z}^{*} = -(\mathbf{D}\mathbf{N})^{+}(\mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{f}) \tag{6}$$

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}(\mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{f}) \tag{7}$$

Problem 7.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{c}^\top \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{8}$$

where $\mathbf{H}$ is positive-definite.

Assume that we found a decomposition $\mathbf{H} = \mathbf{D}^\top \mathbf{D}$. We can also find such $\mathbf{f}$ that $2\mathbf{f}^\top \mathbf{D} = \mathbf{c}^\top$. Then our cost function becomes $\mathbf{x}^\top \mathbf{D}^\top \mathbf{D} \mathbf{x} + 2\mathbf{f}^\top \mathbf{D} \mathbf{x}$, which as we saw before has coinciding minimum with the cost function $||\mathbf{D}\mathbf{x} + \mathbf{f}||$.

Therefore the problem has the same solution as Problem 5, after the mentioned above change in constants.

Problems with inequality constraints

Problem 9.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x} + \mathbf{f}||, \\ \text{subject to} \quad & \mathbf{x} \leq \mathbf{b}. \end{aligned} \tag{9}$$

Problem 10.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{x}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b}. \end{aligned} \tag{10}$$

Problem 11.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x} + \mathbf{f}||, \\ \text{subject to} \quad & \begin{cases} \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ \mathbf{C}\mathbf{x} = \mathbf{d}. \end{cases} \end{aligned} \tag{11}$$

Mentioned problems can be described together as quadratic programs. The name is due to the cost function being quadratic (or equivalent). They are allowed to have linear equality or inequality constraints.

General form of a quadratic program is given below:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x}, \\ \text{subject to} \quad & \begin{cases} \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ \mathbf{C}\mathbf{x} = \mathbf{d}. \end{cases} \end{aligned} \tag{12}$$

where $\mathbf{H}$ is positive-definite and $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ describe a *convex region*.

Domain, Convexity

Domain

Problem 1. Find minimum of the function $f = x^2 + 2y^2$ if $x \in \mathbb{R}$ and $y \in \mathbb{R}$.

Problem 2. Find minimum of the function $f = x^2 + 2y^2$ if $x \in [1\ 2]$ and $y \in [2\ 5]$.

Note that solutions of problems 1 and 2 are different, and this is only due to the difference of the allowed values that the *decision variables* x and y can assume.

Definition 1

Space of all allowed values that decision variables can assume is called the *domain* of optimization problem.

Problem 3. Find minimum of the function $f = -x^2$ if $x \in [-3\ 2]$.

Problem 4. Find minimum of the function $f = -x^2$ if $x \in \mathbb{R}$.

Problem 5. Find minimum of the function $f = -x^2$ if $x \in (-\infty\ 2]$.

The major difference between domains of the problems 2, 3 vs problems 1, 4 and 5 is that the later are *not bound* (i.e., you can construct a sequence of the values in the domain that would approach infinity).

We can see that in the case of problems 3-5, bounded domain allows the problem to have a solution.

Problem 6. Find maximum of the function $f = x^2$ if $1 \le x < 2$.

Problem 7. Find minimum of the function $f = x^2$ if $1 \le x < 2$.

This time, it is the fact that one of the *boundaries* of the domain was not included in the domain that has lead the problem 6 to have no solution, while problem 7 had one. For the problem 6 we can pick a value arbitrary close to $x = 2$, approaching it from the left, but for any such value, there always will be other values of the decision variable closer to $x = 2$ and hence producing larger values of f.

Definition 2

Domain is *convex* iff for any two points in the domain, the line segment connecting them is also in the domain.

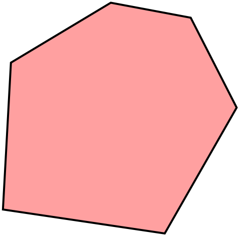

Figure 1: Convex domain

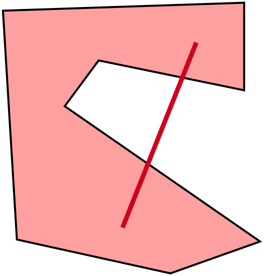

Figure 2: Non-convex domain

In the proofs it is convenient to remember that for any two points $\mathbf{x}_1$ and $\mathbf{x}_2$, all points in the line segment connecting them are given as $\mathbf{x}_l = \alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$, where $\alpha \in [0 \ 1]$. This is called *convex combination*.

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \mathbb{R}^n$ is convex.

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{x} \leq \mathbf{h}\}$ is convex.

Proof: Note that $\alpha \mathbf{x}_1 \leq \alpha \mathbf{h}$ and $(1-\alpha)\mathbf{x}_2 \leq (1-\alpha)\mathbf{h}$, hence, $\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \leq \alpha \mathbf{h} + (1-\alpha)\mathbf{h} = \mathbf{h}$. $\square$

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^2 : \ x_1^2 + x_2^2 \leq h^2\}$ is convex.

Proof: This is the same as $||\mathbf{x}|| \leq h$. Note that $||\alpha \mathbf{x}_1|| \leq \alpha h$ and $||(1-\alpha)\mathbf{x}_2|| \leq (1-\alpha)h$, also

$$||\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2|| \leq ||\alpha \mathbf{x}_1|| + ||(1-\alpha)\mathbf{x}_2||$$

$$||\alpha \mathbf{x}_1|| + ||(1-\alpha)\mathbf{x}_2|| \leq \alpha h + (1-\alpha)h = h$$

So the convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ is still in the domain. $\square$

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{x}^\top \mathbf{H} \mathbf{x} \leq 1\}$, where $\mathbf{H} \succ 0$ is positive-definite symmetric matrix is convex.

For any positive-definite symmetric $\mathbf{H}$ it is true that $\mathbf{H} = \mathbf{D}^\top \mathbf{D}$, where $\mathbf{D} = \sqrt{\mathbf{H}}$ is called a matrix square root and it is full rank. With that $\mathbf{x}^\top \mathbf{H} \mathbf{x} \leq 1$ becomes $\mathbf{x}^\top \mathbf{D}^\top \mathbf{D} \mathbf{x} \leq 1$. Defining $\mathbf{y} = \mathbf{D}\mathbf{x}$ we get $\mathcal{X} = \{\mathbf{D}^{-1}\mathbf{y} : \ \mathbf{y}^\top \mathbf{y} \leq 1\}$. This is a linearly deformed previously covered domain, and as such it is also convex.

Examples of non-convex domains

$x \in \mathcal{X}$, $\mathcal{X} = [-1\ 2] \cup [3\ 7]$ is not convex..

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^2 :\ x_1^2 + x_2^2 \geq h^2\}$ is not convex. Prove it.

$\mathbf{x} \in \mathcal{X}$, $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n :\ \mathbf{x}^\top \mathbf{H} \mathbf{x} \geq 1\}$, where $\mathbf{H}$ is positive-definite symmetric matrix is not convex. Prove it.

These proves simply require one counter-example to show that the defining property of convex domains does not hold.

Definition 3

Function $f(\mathbf{x})$ defined on a domain $\mathcal{D}$, for which it holds that $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, $f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$ is called *a convex function.*

Figure 3: Convex function

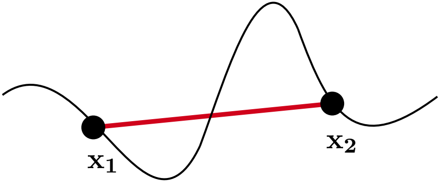

Figure 4: Non-convex function

Here are some single-variable convex functions:

- $f(x) = 1$
- $f(x) = x$, $f(x) = x + 1$, $f(x) = 6x + 3$
- $f(x) = x^2$, $f(x) = (x - 5)^2$, $f(x) = (x + 1)^2 - 10$
- $f(x) = x^3$, if $x > 0$

Here are some multi-variable convex functions:

- $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$
- $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{H}\mathbf{x}$, $\mathbf{H} \succ 0$

Definition 3

If the domain of the optimization problem is convex and the cost function is convex, it is called a *convex optimization problem.*

Additionally, we will always assume that the domain of the convex optimization problem contains its boundary. Also, without the loss of generality, we will consider only minimization problems.

There are a few important properties of convex optimization problems (with our additional assumption):

- If the domain is non-empty, there is a solution.
- The problem has no local minima. We can find a path from any point to the solution, along which the cost function will not increase.

Lecture slides are available via Github, links are on Moodle

You can help improve these slides at:
github.com/SergeiSa/Computational-Intelligence-Slides-Spring-2023