

Least Squares and Quadratic Programming

Computational Intelligence, Lecture 4

by Sergei Savin

Spring 2022

- Problems with analytical solutions
- Problems with inequality constraints
- Quadratic programming
- Homework

PROBLEMS WITH ANALYTICAL SOLUTIONS

Problem 1. minimize $\|\mathbf{x}\|$. Solution $\mathbf{x} = \mathbf{0}$.

Problem 2. minimize $\|\mathbf{Ax}\|$. Solution $\mathbf{x} = \mathbf{0}$.

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

We observe that minimum of $\|\mathbf{Ax} + \mathbf{b}\|$ coincides with the minimum of $(\mathbf{Ax} + \mathbf{b})^2 = (\mathbf{Ax} + \mathbf{b})^\top (\mathbf{Ax} + \mathbf{b}) = \mathbf{b}^\top \mathbf{b} + \mathbf{b}^\top \mathbf{Ax} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{b} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax}$, whose minimum coincides with the minimum of $\mathbf{b}^\top \mathbf{Ax} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{b} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax}$. Since $\mathbf{b}^\top \mathbf{Ax}$ is a scalar, it is equal to its own transpose $(\mathbf{b}^\top \mathbf{Ax})^\top = \mathbf{x}^\top \mathbf{A}^\top \mathbf{b}$. Taking derivative with respect to \mathbf{x} we find gradient of the obtained *cost function*, and set it to zero, since that is the condition for the extreme point:

$$\frac{\partial}{\partial \mathbf{x}} (2\mathbf{b}^\top \mathbf{Ax} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax}) = 2\mathbf{b}^\top \mathbf{A} + 2\mathbf{x}^\top \mathbf{A}^\top \mathbf{A} = \mathbf{0} \quad (1)$$

Thus we get expression: $2\mathbf{x}^\top \mathbf{A}^\top \mathbf{A} = -2\mathbf{b}^\top \mathbf{A}$, which we transpose and find value of \mathbf{x} :

$$\mathbf{x} = -(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \quad (2)$$

The shorthand for this formula is $\mathbf{x} = -\mathbf{A}^+ \mathbf{b}$, and it is called Moore-Penrose pseudoinverse.

Remember that a projector is often defined as $\mathbf{P} = \mathbf{A} \mathbf{A}^+$. Now we can provide an explicit formula for a projector:

$$\mathbf{P} = \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \quad (3)$$

However, neither (2) not (3) should be directly used in as algorithms, since usually there are better ways to compute those quantities (via SVD decomposition, for example).

Problem 4.

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && ||\mathbf{x}||, \\ & \text{subject to} && \mathbf{Ax} = \mathbf{c}. \end{aligned} \tag{4}$$

All solutions to $\mathbf{Ax} = \mathbf{c}$ are written as $\mathbf{x} = \mathbf{A}^+\mathbf{c} + \mathbf{Nz}$, 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{Nz} and row space particular solution $\mathbf{A}^+\mathbf{c}$ are orthogonal, 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, 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.

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

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:

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

Problem 6.

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \|\mathbf{D}\mathbf{x} + \mathbf{f}\|, \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{7}$$

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:

$$\mathbf{z}^* = -(\mathbf{D}\mathbf{N})^+(\mathbf{D}\mathbf{A}^+\mathbf{b} + \mathbf{f}) \tag{8}$$

$$\mathbf{x}^* = \mathbf{A}^+\mathbf{b} - \mathbf{N}(\mathbf{D}\mathbf{N})^+(\mathbf{D}\mathbf{A}^+\mathbf{b} + \mathbf{f}) \tag{9}$$

Problem 7.

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{c}^\top \mathbf{x}, \\ & \text{subject to} && \mathbf{A} \mathbf{x} = \mathbf{b}. \end{aligned} \tag{10}$$

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.

Problem 9.

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & \|\mathbf{D}\mathbf{x} + \mathbf{f}\|, \\ \text{subject to} & \mathbf{x} \leq \mathbf{b}.\end{array}\tag{11}$$

Problem 10.

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & \|\mathbf{x}\|, \\ \text{subject to} & \mathbf{A}\mathbf{x} \leq \mathbf{b}.\end{array}\tag{12}$$

Problem 11.

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & \|\mathbf{D}\mathbf{x} + \mathbf{f}\|, \\ \text{subject to} & \begin{cases} \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ \mathbf{C}\mathbf{x} = \mathbf{d}. \end{cases}\end{array}\tag{13}$$

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:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x}, \\ & \text{subject to} && \begin{cases} \mathbf{A} \mathbf{x} \leq \mathbf{b}, \\ \mathbf{C} \mathbf{x} = \mathbf{d}. \end{cases} \end{aligned} \tag{14}$$

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

- Solve all examples of problems with and without inequalities using `quadprog` function in the language of your choice.

Lecture slides are available via Moodle.

You can help improve these slides at:

github.com/SergeiSa/Computational-Intelligence-Slides-Spring-2022



Check Moodle for additional links, videos, textbook suggestions.