Optimization problems, Analytic solutions

Computational Intelligence, Lecture 4

by Sergei Savin

Spring 2025

Content

- Optimization problem
- Feasibility problem
- Norms and quadratic forms
- Problems with analytical solutions
- Weighted pseudoinverse

An optimization problem has the following form:

$$\begin{aligned} \underset{\text{decision variables}}{\text{minimize}} \quad & \text{cost function}, \\ \text{subject to} \quad & \text{constraints}. \end{aligned} \tag{1}$$

Where the solution to the optimization problem is the optimal value of the decision variables.

For example:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}), \\ \text{subject to} \quad & \begin{cases} g(\mathbf{x}) = 0, \\ h(\mathbf{x}) \leq 0. \end{cases} \end{aligned} \tag{2}$$

In this example, $\mathbf{x} \in \mathbb{R}^n$ is the decision variable, $f(\mathbf{x}) : \ \mathbb{R}^n \to \mathbb{R}$ is a cost function, $g(\mathbf{x}) = 0$ are equality constraints, and $h(\mathbf{x}) \leq 0$ are inequality constraints.

Feasibility problem

A cost function is always scalar. A special case of a cost function is a constant:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} \quad 0, \\ & \text{subject to} \quad \begin{cases} g(\mathbf{x}) = 0, \\ h(\mathbf{x}) \leq 0. \end{cases} \end{aligned} \tag{3}$$

In this case any $\mathbf{x}$ that satisfies constraints would be a solution to the problem. It is called a *feasibility problem*. We solved this type of problems to find out if there exist any $\mathbf{x}$ that satisfies constraints.

Often an optimization problem would not feature constraints:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \tag{4}$$

We can call it *unconstrained optimization*.

Note that the decision variable $\mathbf{x}$ can belong to a set $\mathbf{x} \in \mathcal{X}$ or the cost function may have a domain $f: \mathcal{D} \to \mathbb{R}$; in these cases, the set of allowed values of $\mathbf{x}$, as well as the domain of the function represent *implicit* constraints.

For example, the problem:

$$\underset{x}{\text{minimize}} \quad \ln x$$

has an implicit constraint $x \geq 0$.

Equatily constraints, Lagrangian

Consider optimization problems with equality constraints:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}), \\ \text{subject to} \quad & g(\mathbf{x}) = 0 \end{aligned} \tag{5}$$

We solve it by constructing its *Lagrangian*:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^\top g(\mathbf{x}) \tag{6}$$

Extremum of the Lagrangian corresponds to the solution of the original problem:

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 0 \tag{7}$$

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} = 0 \tag{8}$$

Some types of optimization problems admit an analytic solution. For example:
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. (form 1)

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{x}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{9}$$

All solutions to $\mathbf{A}\mathbf{x} = \mathbf{b}$ are written as $\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + \mathbf{N}\mathbf{z}$, where $\mathbf{N} = \text{null}(\mathbf{A})$, and $\mathbf{A}^{+}\mathbf{b} \in \text{row}(\mathbf{A})$ as we proved previously.

Since the null space solution $\mathbf{N}\mathbf{z}$ and the row space paricular solution $\mathbf{A}^{+}\mathbf{b}$ are orthogonal, the minimum norm solution corresponds to $\mathbf{z} = \mathbf{0}$, hence:

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} \tag{10}$$

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

This illustrates an important fact: the 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.

Least squares with constraints

Problem 4. (form 2)

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{x}^\top \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{11}$$

Let us find Lagrangian $L = \frac{1}{2}\mathbf{x}^\top \mathbf{x} + \lambda^\top(\mathbf{A}\mathbf{x} - \mathbf{b})$. It achieves extremum at solution of the original problem:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} + \mathbf{A}^\top \lambda = 0 \tag{12}$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{A}\mathbf{x} - \mathbf{b} = 0 \tag{13}$$

We can express $\mathbf{x} = -\mathbf{A}^\top \lambda$ and substitute it $\mathbf{A}\mathbf{A}^\top \lambda = -\mathbf{b}$. If $\det(\mathbf{A}\mathbf{A}^\top) \neq 0$, then:

$$\lambda = -(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{b} \tag{14}$$

$$\mathbf{x} = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{b} \tag{15}$$

Least squares with constraints

We can prove that $\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1} = \mathbf{A}^+$.

We can re-write the expression $\mathbf{A}\mathbf{A}^\top\lambda = -\mathbf{b}$ using SVD decomposition $\mathbf{A} = \mathbf{C}\Sigma\mathbf{R}^\top$:

$$\mathbf{C}\Sigma\mathbf{R}^\top\mathbf{R}\Sigma\mathbf{C}^\top\lambda = -\mathbf{b} \tag{16}$$

$$\mathbf{C}\Sigma\Sigma\mathbf{C}^\top\lambda = -\mathbf{b} \tag{17}$$

$$\mathbf{C}^\top\lambda = -\Sigma^{-2}\mathbf{C}^\top\mathbf{b} \tag{18}$$

Next, we do the same with $\mathbf{x} = -\mathbf{A}^\top\lambda$:

$$\mathbf{x} = -\mathbf{R}\Sigma\mathbf{C}^\top\lambda \tag{19}$$

$$\mathbf{x} = \mathbf{R}\Sigma\Sigma^{-2}\mathbf{C}^\top\mathbf{b} \tag{20}$$

$$\mathbf{x} = \mathbf{R}\Sigma^{-1}\mathbf{C}^\top\mathbf{b} \tag{21}$$

But $\mathbf{R}\Sigma^{-1}\mathbf{C}^\top = \mathbf{A}^+$, so $\mathbf{x} = \mathbf{A}^+\mathbf{b}$. More in Appendix.

Problem 5.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{22}$$

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:

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}\mathbf{D}\mathbf{A}^{+}\mathbf{b} \tag{23}$$

$$\mathbf{x}^{*} = (\mathbf{I} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}\mathbf{D})\mathbf{A}^{+}\mathbf{b} \tag{24}$$

Problem 6.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x} + \mathbf{f}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{25}$$

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

$$\mathbf{z}^{*} = -(\mathbf{DN})^{+}(\mathbf{DA}^{+}\mathbf{b} + \mathbf{f}) \tag{26}$$

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} - \mathbf{N}(\mathbf{DN})^{+}(\mathbf{DA}^{+}\mathbf{b} + \mathbf{f}) \tag{27}$$

A symmetric positive-definite matrix $\mathbf{M}$ has the following properties:

$$\mathbf{M} = \mathbf{M}^\top \tag{28}$$

$$\mathbf{x}^\top \mathbf{M} \mathbf{x} > 0, \quad \forall \mathbf{x} \neq 0 \tag{29}$$

$$\det(\mathbf{M}) \neq 0 \tag{30}$$

Positive-definite matrices (with non-repeating eigenvalues) have orthogonal eigenvectors, making their eigenbasis orthonormal:

$$\mathbf{M}\mathbf{V} = \mathbf{V}\Lambda \tag{31}$$

$$\mathbf{M} = \mathbf{V}\Lambda\mathbf{V}^\top \tag{32}$$

Matrix square root

For a symmetric positive-definite matrix $\mathbf{M}$ we can define a square root:

$$\mathbf{M} = \sqrt{\mathbf{M}}\sqrt{\mathbf{M}} \tag{33}$$

$$\mathbf{M} = \mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}} \tag{34}$$

We can find a square root via eigendecomposition:

$$\mathbf{M} = \mathbf{V}\Lambda\mathbf{V}^\top \tag{35}$$

$$\Lambda = \mathbf{F}\mathbf{F} \tag{36}$$

$$\mathbf{M} = \mathbf{V}\mathbf{F}\mathbf{F}\mathbf{V}^\top \tag{37}$$

$$\mathbf{M} = \mathbf{V}\mathbf{F}\mathbf{V}^\top\mathbf{V}\mathbf{F}\mathbf{V}^\top \tag{38}$$

$$\sqrt{\mathbf{M}} = \mathbf{V}\mathbf{F}\mathbf{V}^\top \tag{39}$$

$$\mathbf{M} = \sqrt{\mathbf{M}}\sqrt{\mathbf{M}} \tag{40}$$

Problem 7.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{c}^\top \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{41}$$

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.

Consider a weighted pseudoinverse problem:

$$\text{minimize} \quad ||\mathbf{A}\mathbf{x} - \mathbf{b}||_{\mathbf{W}} \tag{42}$$

where $||\mathbf{x}||_{\mathbf{W}} = \sqrt{\mathbf{x}^\top \mathbf{W} \mathbf{x}}$ and $\mathbf{W} > 0$. We can re-write the problem as:

$$\text{minimize} \quad (\mathbf{A}\mathbf{x} - \mathbf{b})^\top \mathbf{W}^{\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} (\mathbf{A}\mathbf{x} - \mathbf{b}) \tag{43}$$

But this is the same as solving least-squares problem for equality $\mathbf{W}^{\frac{1}{2}} \mathbf{A}\mathbf{x} = \mathbf{W}^{\frac{1}{2}} \mathbf{b}$, which is does via Moore-Penrose pseudoinverse:

$$\mathbf{x} = (\mathbf{W}^{\frac{1}{2}} \mathbf{A})^{+} \mathbf{W}^{\frac{1}{2}} \mathbf{b} \tag{44}$$

Consider a weighted pseudoinverse problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{W} \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{45}$$

We can use Lagrange multipliers to rewrite the problem as minimization of the function $L(\mathbf{x}, \lambda) = \mathbf{x}^\top \mathbf{W} \mathbf{x} + \lambda^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$; optimality conditions imply that $\frac{\partial L}{\partial \mathbf{x}} = 0$ and $\frac{\partial L}{\partial \lambda} = \mathbf{A}\mathbf{x} - \mathbf{b} = 0$, so:

$$2\mathbf{x}^\top \mathbf{W} + \lambda^\top \mathbf{A} = 0 \tag{46}$$

This implies $\mathbf{x} = \frac{1}{2}\mathbf{W}^{-1}\mathbf{A}^\top \lambda$, and since $\mathbf{A}\mathbf{x} - \mathbf{b} = 0$, we get:

$$\frac{1}{2}\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top \lambda = \mathbf{b} \tag{47}$$

$$\lambda = 2(\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top)^{+}\mathbf{b} \tag{48}$$

$$\mathbf{x} = \mathbf{W}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top)^{+}\mathbf{b} \tag{49}$$

Example, 1

A drone flies along a line defined by a point $\mathbf{c}$ and a direction $\mathbf{v}$. Find a point at which it will pass closest to a ground station located at $\mathbf{a}$.

Solution. We can describe any point on the line as $\mathbf{x} = \mathbf{v}\alpha + \mathbf{c}$. So, the problem can be written as:

$$\begin{aligned} \underset{\mathbf{x},\alpha}{\text{minimize}} \quad & \frac{1}{2}(\mathbf{x} - \mathbf{a})^\top (\mathbf{x} - \mathbf{a}), \\ \text{subject to} \quad & \mathbf{x} = \mathbf{v}\alpha + \mathbf{c} \end{aligned} \tag{50}$$

Lagrangian of the problem is:

$$L = \frac{1}{2}(\mathbf{x} - \mathbf{a})^\top (\mathbf{x} - \mathbf{a}) + \lambda^\top (\mathbf{v}\alpha + \mathbf{c} - \mathbf{x}) \tag{51}$$

$$L = \frac{1}{2}(\mathbf{x} - \mathbf{a})^\top (\mathbf{x} - \mathbf{a}) + \lambda^\top (\mathbf{v}\xi + \mathbf{c} - \mathbf{x}) \tag{52}$$

Partial dervatives of the Lagrangian:

$$\frac{\partial L}{\partial \mathbf{x}} = (\mathbf{x} - \mathbf{a}) - \lambda = 0 \tag{53}$$

$$\frac{\partial L}{\partial \alpha} = \lambda^\top \mathbf{v} = 0 \tag{54}$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{v}\xi + \mathbf{c} - \mathbf{x} = 0 \tag{55}$$

This can be expressed as:

$$\begin{bmatrix} \mathbf{I} & 0 & -\mathbf{I} \\ 0 & 0 & \mathbf{v}^\top \\ -\mathbf{I} & \mathbf{v} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \xi \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ 0 \\ -\mathbf{c} \end{bmatrix} \tag{56}$$

which can be solved via inverse.

Using equations $\mathbf{x} - \mathbf{a} = \lambda$ and $\lambda^\top \mathbf{v} = 0$ and $\mathbf{v}\xi + \mathbf{c} = \mathbf{x}$ we can write:

$$\mathbf{v}^\top(\mathbf{x} - \mathbf{a}) = 0 \tag{57}$$

$$\mathbf{v}^\top\mathbf{v}\xi + \mathbf{v}^\top\mathbf{c} - \mathbf{v}^\top\mathbf{a} = 0 \tag{58}$$

Since $\mathbf{v}^\top\mathbf{v} = 1$, so $\xi = \mathbf{v}^\top(\mathbf{a} - \mathbf{c})$. With that:

$$\mathbf{x} = \mathbf{v}\mathbf{v}^\top\mathbf{a} + (\mathbf{I} - \mathbf{v}\mathbf{v}^\top)\mathbf{c} \tag{59}$$

$$\lambda = (\mathbf{I} - \mathbf{v}\mathbf{v}^\top)(\mathbf{c} - \mathbf{a}) \tag{60}$$

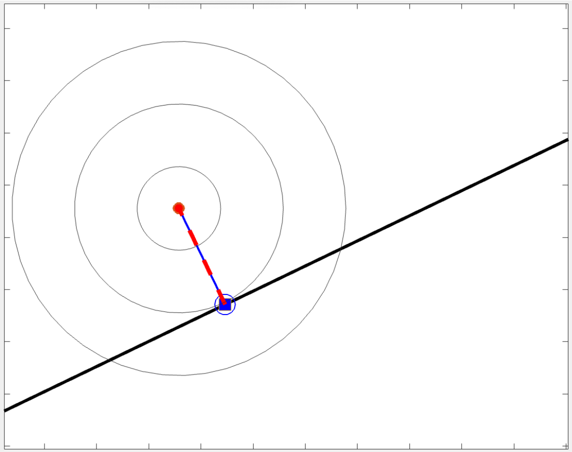

Figure 1: An illustration of the problem

Lecture slides are available via Github, links are on Moodle:

github.com/SergeiSa/Computational-Intelligence-2025

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{x}^\top \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{61}$$

All solutions to $\mathbf{A}\mathbf{x} = \mathbf{b}$ are written as $\mathbf{x} = \mathbf{A}^+\mathbf{b} + \mathbf{N}\mathbf{z}$, where $\mathbf{N} = \text{null}(\mathbf{A})$, and $\mathbf{A}^+\mathbf{b} \in \text{row}(\mathbf{A})$ as we proved previously. The cost function is:

$$f_c = \frac{1}{2}(\mathbf{A}^+\mathbf{b} + \mathbf{N}\mathbf{z})^\top(\mathbf{A}^+\mathbf{b} + \mathbf{N}\mathbf{z}) \tag{62}$$

We find extremum:

$$\frac{\partial f_c}{\partial \mathbf{z}} = \mathbf{N}^\top\mathbf{A}^+\mathbf{b} + \mathbf{N}^\top\mathbf{N}\mathbf{z} = 0 \tag{63}$$

$$\mathbf{z} = -\mathbf{N}^\top\mathbf{A}^+\mathbf{b} \tag{64}$$

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} - \mathbf{N}\mathbf{N}^\top\mathbf{A}^+\mathbf{b} \tag{65}$$

Columns of $\mathbf{A}^+$ lie in the row space of $\mathbf{A}$, so $\mathbf{N}^\top\mathbf{A}^+ = 0$:

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} \tag{66}$$