Optimization problems, Analytic solutions

Computational Intelligence, Lecture 3

by Sergei Savin

Spring 2024

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.

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 $x \geq 0$.

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 $\mathbf{x} = -\mathbf{A}^{+}\mathbf{b}$.

Note that the following problems will always have the same solutions:

- minimize $||\mathbf{A}\mathbf{x} + \mathbf{b}||$;
- minimize $(\mathbf{A}\mathbf{x} + \mathbf{b})^\top(\mathbf{A}\mathbf{x} + \mathbf{b})$;

This is because square root is a monotonic function.

This **does not** imply equivalence of the following problems:

- minimize $\sum ||\mathbf{A}_i\mathbf{x} + \mathbf{b}_i||$;
- minimize $\sum(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)^\top(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i)$;

Problem 4.

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

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.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{D}\mathbf{x}||, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{6}$$

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:

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

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

Problem 6.

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

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:

$$\mathbf{z}^{*} = -(\mathbf{D}\mathbf{N})^{+}(\mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{f}) \tag{10}$$

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} - \mathbf{N}(\mathbf{D}\mathbf{N})^{+}(\mathbf{D}\mathbf{A}^{+}\mathbf{b} + \mathbf{f}) \tag{11}$$

Problem 7.

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

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.

Weighted pseudoinverse, unconstrained type

Consider a weighted pseudoinverse problem:

$$\text{minimize} \quad ||\mathbf{A}\mathbf{x} - \mathbf{b}||_{\mathbf{W}} \tag{13}$$

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

$$\text{minimize} \quad (\mathbf{A}\mathbf{x} - \mathbf{b})^\top \mathbf{W}^{\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} (\mathbf{A}\mathbf{x} - \mathbf{b}) \tag{14}$$

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:

$$\mathbf{x} = (\mathbf{W}^{\frac{1}{2}} \mathbf{A})^+ \mathbf{W}^{\frac{1}{2}} \mathbf{b} \tag{15}$$

Consider a weighted pseudoinverse problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{x}^\top \mathbf{W} \mathbf{x}, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{16}$$

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:

$$2\mathbf{x}^\top \mathbf{W} + \lambda^\top \mathbf{A} = 0 \tag{17}$$

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:

$$\frac{1}{2}\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top \lambda = \mathbf{b} \tag{18}$$

$$\lambda = 2(\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top)^{+}\mathbf{b} \tag{19}$$

$$\mathbf{x} = \mathbf{W}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^\top)^{+}\mathbf{b} \tag{20}$$

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

github.com/SergeiSa/Computational-Intelligence-2024