Sparse solutions to linear algebraic systems

In block 3 we dealt with *overdetermined*/unsolvable systems of linear algebraic equations by reformulating them as the least squares problems. Often, for example in machine learning, statistics, or compressed sensing applications, we are interested in solving *underdetermined* systems of linear algebraic equations, which have infinitely many solutions. Out of those solutions, we are often interested in selecting the ones satisfying certain criteria, such as for example the ones having few non-zero entries. One rather efficient way of doing this is to solve the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad && \sum_{i=1}^{n} |x_i|, \\ &\text{subject to} \quad && Ax = b \end{aligned} \tag{1}$$

where the matrix A with m rows and n columns, and the vector $b \in \mathbb{R}^m$ are given.

1. Explain (it is sufficient to consider the dimension $n = 1$) why the objective function $\sum_{i=1}^{n} |x_i|$ does not satisfy the definition of a linear tranformation from $\mathbb{R}^n$ to $\mathbb{R}$, see page 82 in [Lay]. The consequence of this fact is that the optimization problem (1) is *not* a linear optimization problem; however, it can be equivalently restated as one.

2. Let us for a moment consider the situation when $n = m = 1$, $A = [2]$, $b = [\beta]$. Determine the solution to (1) as a function of $\beta \in \mathbb{R}$.

   Then, sketch the feasible set $\mathcal{F}$ and graphically determine the optimal solution to the following linear optimization problem:

   $$\begin{aligned} &\text{minimize} \quad && x^+ + x^-, \\ &\text{subject to} && \\ & && 2x^+ - 2x^- = \beta, \\ & && x^+ \geq 0, x^- \geq 0, \end{aligned} \tag{2}$$

   as a function of $\beta \in \mathbb{R}$. Hint: when sketching/solving, consider separately two cases, $\beta \geq 0$ and $\beta < 0$.

   The observation from this activity should be that if $(\bar{x}^+, \bar{x}^-)$ is the optimal solution to (2), then either $\bar{x} = \bar{x}^+$ or $\bar{x} = -\bar{x}^-$ is the optimal solution to (1). In other words, we split the optimal solution into its nonnegative and nonpositive parts; whence the notation.

For larger dimensions n, m the idea of transforming (1) into a linear optimization problem by introducing nonnegative and nonpositive parts of the solution remains valid. That is, we put

$$\tilde{x} = \begin{bmatrix} x^+ \\ x^- \end{bmatrix} \in \mathbb{R}^{2n}, \qquad \tilde{c} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{2n}, \qquad \tilde{A} = \begin{bmatrix} A & -A \end{bmatrix},$$

where the vectors $x^+ \in \mathbb{R}^n$, $x^- \in \mathbb{R}^n$. Then (1) is equivalent with a linear optimization problem:

$$\begin{aligned} &\text{minimize} \quad && \tilde{c} \cdot \tilde{x}, \\ &\text{subject to} && \\ & && \tilde{A}\tilde{x} = b, \\ & && \tilde{x} \geq 0. \end{aligned} \tag{3}$$

Throughout the following questions we will consider the particular instance of (3) where $m = 1$, $n = 5$, $A = [1, 2, 3, 4, 5]$, and $b = [10]$, from which one determines $\tilde{A}$.

3. Determine all 10 basic solutions for (3), and select out of them 5 feasible basic solutions. (These solutions correspond to the extreme points/“corners” of the 10-dimensional feasible set for $\tilde{x}$ and as such are not easy to visualize.) Evaluate the objective function in these 5 points and select the best one, which is thus a solution to (3).

4. Convert the problem (3) to a linear optimization problem in the canonical form.

5. Write down the dual problem for the linear optimization problem in the canonical form obtained in the previous question. Sketch the feasible set for the dual problem and solve it graphically. Do you obtain the same optimal value for the objective function as in question 3?

6. Apply the simplex algorithm to solve the linear optimization problem in the canonical form from question 4. [1] Relate the computed optimal solution/objective value to the ones previously obtained in questions 3 and 5. Does the strong duality hold?

Application: compressed sensing

Our objective is to represent a time-dependent signal $f(t)$, which we sample at some discrete times $t_1, t_2, \dots, t_m$, as a linear combination of a few "basis functions" $g_1, g_2, \dots, g_n$, $n \geq m$, with unknown weights x_j. That is, we would like to fulfill the equations

$$f(t_i) = \sum_{j=1}^{n} x_j g_j(t_i), \qquad \forall i = 1, \dots, m,$$

in such a way that $\sum_{j=1}^{n} |x_j|$ is as small as possible.

7. Formulate the compressed sensing problem in the form (1). That is, express the matrix A in terms of g_j and t_i, and the right hand side vector b in terms of f and t_i, $i = 1, \dots, m$, $j = 1, \dots, n$.

8. The file `data.txt` on moodle contains the times t_i (first column) and the measured signal values $f(t_i)$ (second column). Solve the instance of the compressed sensing problem given by this data and the following $n = 60$ "basis functions," which include polynomials up to degree 19 and trigonometric functions:

$$g_j(t) = \begin{cases} \cos((j-1)\arccos(t)), & j = 1, \dots, 20, \quad \text{(Chebyshev polynomials)} \\ \cos(\pi(j-20)t), & j = 21, \dots, 40, \\ \sin(\pi(j-40)t), & j = 41, \dots, 60. \end{cases}$$

   Is the obtained representation of the signal/optimal solution to the problem "compressed"/contains many zeros? Plot the data and the computed reconstruction of the signal.

   Hint: There are examples on moodle on how to read data files, solve linear probems of the type (3), and plotting the results in Matlab and Python.

[1] Look at example 7 in section 9.3 [Lay] to see how to do this.