Solving nonnegative least squares with L_1-regularization

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be   Homepage: angms.science

First draft : July 12, 2020
Last update : July 13, 2020

Problem

- Given $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$ by solving the nonnegative least squares (NNLS) with L_1-regularization

$$(\mathcal{P}) \;\; : \;\; \min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1,$$

  where $\lambda \geq 0$ is a given regularization parameter and $\|\mathbf{x}\|_1 = \sum_i |x_i|$ is the L_1-norm of $\mathbf{x}$

- This problem is related to the L_1-regularized least squares

$$(\mathcal{P}_0) \;\; : \;\; \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$

- This document: how to solve Problem $(\mathcal{P})$.

Using indicator function may makes things complicated

- If we carry the solution approach on solving $(\mathcal{P}_0)$ to $(\mathcal{P})$, things may get more complicated.
- The nonnegativity constraint $\mathbf{x} \geq 0$ can be expressed using indicator function, and Problem $(\mathcal{P})$ is equivalent to

$$(\mathcal{P}') \;\; : \;\; \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 + i_+(\mathbf{x}).$$

- This somehow makes the problem complicated: to solve $(\mathcal{P}')$, which is a problem involving indicator function, the standard approach is to make use of the proximal operator, or more precisely, the proximal gradient update, which is used to solve $(\mathcal{P}_0)$. See here for more about proximal gradient update.
- However we cannot apply proximal gradient update directly for $(\mathcal{P}')$ as the problem contains the L_1-term.
- We have to solve Problem $(\mathcal{P})$ differently.

Solving Problem $(\mathcal{P})$ component-wise

- Note that the constraint $\mathbf{x} \geq 0$ implies $\|\mathbf{x}\|_1 = \sum x_i$. This removes the absolute value sign, and is the key why $(\mathcal{P})$ has closed-form solution.
- Let the objective function in $(\mathcal{P})$ be $f(\mathbf{x})$. Let $\mathbf{a}_i$ be the ith column of $\mathbf{A}$, then $f(\mathbf{x})$ can be expressed as

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\
&= \frac{1}{2}\left\|\sum_{i=1}^{n} \mathbf{a}_i x_i - \mathbf{b}\right\|_2^2 + \lambda \sum_{i=1}^{n} x_i \quad (\because \mathbf{x} \geq 0 \therefore \|\mathbf{x}\|_1 = \sum x_i) \\
&= \frac{1}{2}(\mathbf{a}_i x_i - \mathbf{b}_{-i})^2 + \lambda x_i \qquad \mathbf{b}_{-i} = -\sum_{j \neq i} \mathbf{a}_j x_j + \mathbf{b}
\end{aligned}
$$

  we expressed $f(\mathbf{x})$ as a scalar function on x_i.
- $f(\mathbf{x})$ is minimized $\iff$ each $f(x_i)$ for all i is minimized. We now focus on minimizing $f(x_i)$.

Solving the scalar problem

- The scalar minimization subproblem is in the form

$$\underset{x\geq 0}{\operatorname{argmin}}\ f(x) = \frac{1}{2}(\mathbf{a}x - \mathbf{b})^2 + \lambda x,$$

where $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^m$ are given vectors, $\lambda \geq 0$ is given, and x is the scalar variable.

- Expand f on x, and remove terms independent to x gives

$$\underset{x\geq 0}{\operatorname{argmin}}\ \frac{\|\mathbf{a}\|_2^2}{2} x^2 + (\lambda - \langle \mathbf{a}, \mathbf{b} \rangle)x$$

which is a quadratic function in the form of $\frac{a}{2}x^2 + bx$. The minimizer of this function with respect to $x \geq 0$ is $\left[\frac{-b}{a}\right]_+ = \max\left\{0, \frac{-b}{a}\right\}$.

- The solution to the scalar minimization subproblem

$$x = \left[\frac{\langle \mathbf{a}, \mathbf{b} \rangle - \lambda}{\|\mathbf{a}\|_2^2}\right]_+ .$$

Why $\left[\frac{-b}{a}\right]_+$ solves $\underset{x\geq 0}{\operatorname{argmin}}\ \frac{a}{2}x^2 + bx$

- Simple high school trick : completing the square

$$\begin{aligned}
\frac{a}{2}x^2 + bx &= \frac{a}{2}\left(x^2 + 2\frac{b}{a}x\right) \\
&= \frac{a}{2}\left(x^2 + 2\frac{b}{a}x + \frac{b^2}{a^2} - \frac{b^2}{a^2}\right) \\
&= \frac{a}{2}\left((x + \frac{b}{a})^2 - \frac{b^2}{a^2}\right) \\
&= \frac{a}{2}\left(x + \frac{b}{a}\right)^2 + \text{constant.}
\end{aligned}$$

  This quadratic function is minimized at $x = -\frac{b}{a}$.

- If $-\frac{b}{a} \geq 0$, it satisfies the constraint $x \geq 0$, so $x = -\frac{b}{a}$ is the solution.
- If $-\frac{b}{a} < 0$, then $x = 0$ is the solution, since 0 is the closest scalar to $-\frac{b}{a}$ that satisfies the constraint $x \geq 0$.
- Combine both cases, the solution of $\underset{x\geq 0}{\operatorname{argmin}}\ \frac{a}{2}x^2 + bx$ is thus $\left[\frac{-b}{a}\right]_+$.

The overall algorithm

Algorithm 1: A simple compont-wise algorithm to solve $(\mathcal{P})$

Result: $\mathbf{x} \in \mathbb{R}^n$ that solves $(\mathcal{P})$

Pick a $i \in [n]$;

while *not converge* **do**

$\mathbf{b}_{-i} = \mathbf{b} - \mathbf{A}\mathbf{x} + \mathbf{a}_i x_i$;

$x_i = \frac{[\langle \mathbf{a}_i, \mathbf{b}_{-i} \rangle - \lambda]_+}{\|\mathbf{a}_i\|_2^2}$;

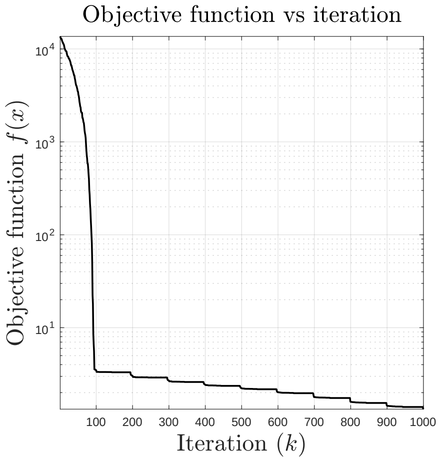
Move to other i.

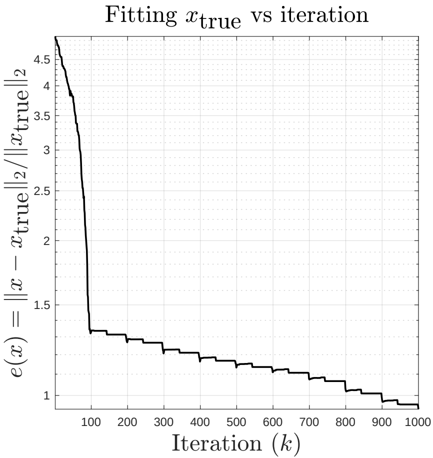
end

Note that the algorithm can be further enhenced by pre-computing $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}^\top \mathbf{b}$ to reduce the computational cost on computing the term $\langle \mathbf{a}_i, \mathbf{b}_{-i} \rangle$.

Simple example

$[m, n] = [50, 100]$, 95% sparse (only 5 non-zero in $\mathbf{x}_{\text{true}}$).
Regularization parameter $\lambda = 1$ (arbitrarily selected).





Last page - summary

- Given $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$ by solving the NNLS with L_1-regularization

$$(\mathcal{P}) \ : \ \min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1,$$

- The solution to $(\mathcal{P})$ is given component-wise as

$$x_i = \left[\frac{\langle \mathbf{a}_i, \mathbf{b}_{-i}\rangle - \lambda}{\|\mathbf{a}_i\|_2^2}\right]_+, \quad \mathbf{b}_{-i} = -\mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{a}_i x_i.$$

End of document