ℓ_1-regularized least squares

Amir Khalilian
New York University.

October 18, 2023

Outline

Soft-thresholding

Alternating direction method of multipliers

Notations

- $x = [x_1, \cdots, x_n]^\top$
- $\|x\|_2^2 = \sum_{i=1}^{n} x_i^2$
- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- $f(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda\|x\|_1 = \sum_{i=1}^{n} \frac{1}{2}(y_i - x_i)^2 + \lambda\,|x_i|$

The scalar problem

$$\underset{x \in \mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x \in \mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

The scalar problem

$$\underset{x\in\mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x\in\mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

- The minimizer of $f(x)$, denoted by x^*, satisfies $0 \in \partial f(x^*)$.

The scalar problem

$$\underset{x\in\mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x\in\mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

- The minimizer of $f(x)$, denoted by x^*, satisfies $0 \in \partial f(x^*)$.
- $\dfrac{\partial}{\partial x}(y-x)^2 = 2(x-y)$.

The scalar problem

$$\underset{x\in\mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x\in\mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

- The minimizer of $f(x)$, denoted by x^*, satisfies $0 \in \partial f(x^*)$.
- $\dfrac{\partial}{\partial x}(y-x)^2 = 2(x-y)$.
- $|x|$ is differentiable except at $x = 0$ i.e. $\partial|x| = \operatorname{sgn}(x)$ for $x \neq 0$.

The scalar problem

$$\underset{x\in\mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x\in\mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

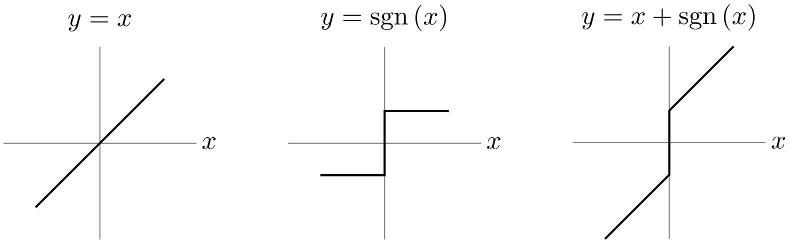
- The minimizer of $f(x)$, denoted by x^*, satisfies $0 \in \partial f(x^*)$.
- $\dfrac{\partial}{\partial x}(y-x)^2 = 2(x-y)$.
- $|x|$ is differentiable except at $x = 0$ i.e. $\partial|x| = \operatorname{sgn}(x)$ for $x \neq 0$.
- $x^* - y + \lambda \operatorname{sgn}(x^*) = 0 \Rightarrow y = x^* + \lambda \operatorname{sgn}(x^*)$

The scalar problem

let's plot $y = x^* + \lambda \operatorname{sgn}(x^*)$

The scalar problem

let's plot $y = x^* + \lambda \operatorname{sgn}(x^*)$



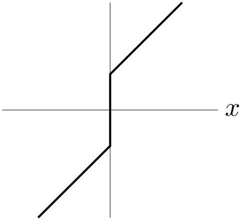
Soft-thresholding

$$\underset{x \in \mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x \in \mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y - x)^2 + \lambda\,|x|$$

Soft-thresholding

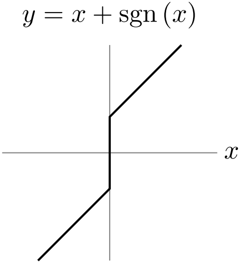
$$\underset{x \in \mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x \in \mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y-x)^2 + \lambda\,|x|$$

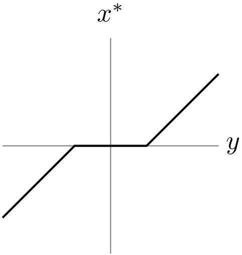
$y = x + \operatorname{sgn}(x)$



Soft-thresholding

$$\underset{x \in \mathbb{R}}{\text{minimize}}\ f(x) \equiv \underset{x \in \mathbb{R}}{\text{minimize}}\ \frac{1}{2}(y - x)^2 + \lambda\,|x|$$





- $x^* = \operatorname{sgn}(y) \times \max\{|y| - \lambda, 0\}$

Outline

Soft-thresholding

Alternating direction method of multipliers

Dual decomposition

- convex equality constrained optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to, } Ax = b$$

Dual decomposition

- convex equality constrained optimization problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to, } Ax = b \end{aligned}$$

- Lagrangian: $L(x, u) = f(x) + u^\top (Ax - b)$.

Dual decomposition

- convex equality constrained optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to, } Ax = b$$

- Lagrangian: $L(x, u) = f(x) + u^\top(Ax - b)$.
- dual function: $g(u) = \inf_x L(x, u)$.

Dual decomposition

- convex equality constrained optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to, } Ax = b$$

- Lagrangian: $L(x, u) = f(x) + u^\top (Ax - b)$.
- dual function: $g(u) = \inf_x L(x, u)$.
- dual problem: $\text{maximize } g(u)$.

Dual decomposition

- convex equality constrained optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to, } Ax = b$$

- Lagrangian: $L(x, u) = f(x) + u^\top (Ax - b)$.
- dual function: $g(u) = \inf_x L(x, u)$.
- dual problem: $\text{maximize } g(u)$.
- recover $x^* = \arg\min_x L(x, u^*)$.

Dual ascent

- apply gradient method to the dual problem: $u^{k+1} = u^k + \alpha \nabla g(u^k)$.

Dual ascent

- apply gradient method to the dual problem: $u^{k+1} = u^k + \alpha \nabla g(u^k)$.
- $\nabla g(u^k) = A\hat{x} - b$ with $\hat{x} = \arg\min_x L(x, u^k)$.

Dual ascent

- apply gradient method to the dual problem: $u^{k+1} = u^k + \alpha \nabla g(u^k)$.
- $\nabla g(u^k) = A\hat{x} - b$ with $\hat{x} = \arg\min_x L(x, u^k)$.
- algorithm:

$$x^{k+1} = \arg\min_x L(x, u^k)$$
$$u^{k+1} = u^k + \alpha A x^{k+1} - b.$$

Dual ascent

- apply gradient method to the dual problem: $u^{k+1} = u^k + \alpha \nabla g(u^k)$.
- $\nabla g(u^k) = A\hat{x} - b$ with $\hat{x} = \arg\min_x L(x, u^k)$.
- algorithm:

$$
\begin{aligned}
x^{k+1} &= \arg\min_x L(x, u^k) \\
u^{k+1} &= u^k + \alpha A x^{k+1} - b.
\end{aligned}
$$

- barely works in practice! needs lots of strong assumptions!

Dual decomposition

- what if f is separable? $f(x) = f_1(x_1) + \cdots + f_n(x_n)$

Dual decomposition

- what if f is separable? $f(x) = f_1(x_1) + \cdots + f_n(x_n)$
- algorithm:

$$x_i^{k+1} = \arg\min_{x_i} L(x_i, u^k)$$

$$u^{k+1} = u^k + \alpha \sum_{i=1}^{n} Ax_i^{k+1} - b.$$

Dual decomposition

- what if f is separable? $f(x) = f_1(x_1) + \cdots + f_n(x_n)$
- algorithm:

$$x_i^{k+1} = \arg\min_{x_i} L(x_i, u^k)$$
$$u^{k+1} = u^k + \alpha \sum_{i=1}^{n} A x_i^{k+1} - b.$$

- solves large problems! often slow, needs lots of strong assumptions!

Alternating direction method of multipliers

- constrained optimization problem

$$\text{minimize } f(x) + g(z)$$
$$\text{subject to, } Ax + Bz = c$$

Alternating direction method of multipliers

- constrained optimization problem

$$\text{minimize } f(x) + g(z)$$
$$\text{subject to, } Ax + Bz = c$$

- two sets of variables, with separable objective

Alternating direction method of multipliers

- constrained optimization problem

$$\begin{aligned} \text{minimize } & f(x) + g(z) \\ \text{subject to, } & Ax + Bz = c \end{aligned}$$

- two sets of variables, with separable objective
- augmented Lagrangian:

$$L(x, z, u) = f(x) + g(z) + u^\top (Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2$$

Alternating direction method of multipliers

- augmented Lagrangian:

$$L(x, z, u) = f(x) + g(z) + u^\top (Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2$$

Alternating direction method of multipliers

- augmented Lagrangian:

$$L(x, z, u) = f(x) + g(z) + u^\top (Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2$$

- algorithm:

$$\begin{aligned} x^{k+1} &= \arg\min_x L(x, z^k, u^k) \\ z^{k+1} &= \arg\min_z L(x^{k+1}, z, u^k) \\ u^{k+1} &= u^k + \alpha A x^{k+1} + B z^{k+1} - c. \end{aligned}$$

Alternating direction method of multipliers

- augmented Lagrangian:

$$L(x, z, u) = f(x) + g(z) + u^\top (Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2$$

- algorithm:

$$\begin{aligned} x^{k+1} &= \arg\min_x L(x, z^k, u^k) \\ z^{k+1} &= \arg\min_z L(x^{k+1}, z, u^k) \\ u^{k+1} &= u^k + \alpha A x^{k+1} + B z^{k+1} - c. \end{aligned}$$