STAT 499: Undergraduate Research

Week 3: Estimation in noisy settings

Facheng Yu

Compiled on: 2024-01-23, 18:59:45

3.1 From basis pursuit program

Suppose that $(y, \mathbf{X}) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$. we now consider the model $y = \mathbf{X}\theta^* + w$ with the noise vector $w \in \mathbb{R}^n$.

Extension of the basis pursuit program. The extension relaxes the constraints of the basis pursuit program, i.e., y does not have to $\mathbf{X}\theta$ for some θ. The extended program can be written as

$$\min_{\theta \in \mathbb{R}^d} \|\theta\|_1 \quad \text{such that } \frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2 \leq b^2 \tag{3.1}$$

for some noise tolerance $b > 0$.

The program above can be shown as equivalent as such a program:

$$\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2 \right\} \quad \text{such that } \|\theta\|_1 \leq R \tag{3.2}$$

for some radius $R > 0$.

Lasso program. To eliminate the constraint, one can consider the lasso program as well:

$$\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2 + \lambda_n \|\theta\|_1 \right\}. \tag{3.3}$$

Here $\lambda_n > 0$ is a regularization parameter to be chosen by the user.

Note: By Lagrangian duality theory, all three families of convex programs are equivalent, which is given in the next section.

3.2 Equivalent programs

Consider such a convex optimization:

$$p^* := \min_x f_0(x) : \quad \begin{array}{ll} f_i(x) \leq 0, & i = 1, \cdots, m, \\ h_i(x) = 0, & i = 1, \cdots, p, \end{array}$$

where the functions $f_0, f_1, \ldots, f_m$ are convex, and $h_1, \ldots, h_p$ are affine.

Let $\mathcal{D}$ be the domain of the problem and $\mathcal{X} \subseteq \mathcal{D}$ be its feasible set.

Define the Lagrangian $\mathcal{L} : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with values

$$\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x).$$

The dual function is $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with values

$$g(\lambda, \nu) := \min_x \mathcal{L}(x, \lambda, \nu).$$

The dual problem is

$$d^* = \max_{\lambda \geq 0, \nu} g(\lambda, \nu).$$

Strong duality. The convex program is said to satisfy strong duality if $p^* = d^*$.

Theorem 3.1 (Strong duality via Slater condition) *We say that the problem satisfies Slater's condition if it is strictly feasible, that is:*

$$\exists x_0 \in \mathcal{D} : f_i\left(x_0\right) < 0, \quad i = 1, \ldots, m, \quad h_i\left(x_0\right) = 0, \quad i = 1, \ldots, p.$$

If the primal problem (8.1) is convex, and satisfies the weak Slater's condition, then strong duality holds, that is, $p^* = d^*$.

We also need minimax theorem to help change the order of minimization and maximization:

Theorem 3.2 (Minimax theorem) *Let $\phi(x, y) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Define*

$$p^* = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y)$$

and

$$d^* = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y)$$

It holds that the gap $p^* - d^*$ is zero if:

- $\mathcal{X}, \mathcal{Y}$ are both convex, and one of them is compact.

- The function ϕ is convex-concave: $\phi(\cdot, y)$ is convex for every $y \in \mathcal{Y}$, and $\phi(x, \cdot)$ is concave for every $x \in \mathcal{X}$.

- The function ϕ is continuous.

In our case, the program is reduced to the situation that $m = 1$ and $p = 0$ for 3.1 and 3.2. By theorem 3.1 and theorem 3.2 we would have following statements.

Proposition 3.3 *Suppose that (3.1), (3.2), and (3.3) are convex programs. Then it holds that*

(i) For any $b > 0$, there exists $\lambda \geq 0$ such that program (3.1) and program (3.3) are equivalent;

(ii) For any $R > 0$, there exists $\lambda \geq 0$ such that program (3.2) and program (3.3) are equivalent.

Proof:
(i) The primal program is

$$\min_{\theta \in \mathbb{R}^d} \|\theta\|_1 \quad \text{such that } \frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2 - b^2 \leq 0. \tag{3.4}$$

The corresponding dual program is

$$\max_{\lambda > 0} \min_{\theta \in \mathbb{R}^d} \|\theta\|_1 + \lambda \left\{ \frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2 - b^2 \right\}. \tag{3.5}$$

By theorem 3.2 we can exchange the min and max:

$$\max_{\lambda>0}\min_{\theta\in\mathbb{R}^d}\|\theta\|_1+\lambda\left\{\frac{1}{2n}\|y-\mathbf{X}\theta\|_2^2-b^2\right\} \tag{3.6}$$

$$=\min_{\theta\in\mathbb{R}^d}\max_{\lambda>0}\|\theta\|_1+\lambda\left\{\frac{1}{2n}\|y-\mathbf{X}\theta\|_2^2-b^2\right\} \tag{3.7}$$

$$=\min_{\theta\in\mathbb{R}^d}\|\theta\|_1+\lambda^*\left\{\frac{1}{2n}\|y-\mathbf{X}\theta\|_2^2-b^2\right\}\ \text{for some } \lambda^*>0. \tag{3.8}$$

Then by theorem 3.1, we know that program (3.1) and program (3.3) are equivalent.

(ii) The proof is identical to (i). ■

3.3 Proof of the minimax theorem

Lemma 3.4 *It holds that*

$$\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}}\phi(x,y)\le\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\phi(x,y).$$

Proof: For any $x\in\mathcal{X}$, $y\in\mathcal{Y}$

$$\min_{x'\in\mathcal{X}}\phi(x',y)\le\phi(x,y)\le\max_{y'\in\mathcal{Y}}\phi(x,y').$$

Since the inequality holds for any $x\in\mathcal{X}$ and any $y\in\mathcal{Y}$, we can take max on the left and take min on the right:

$$\max_{y\in\mathcal{Y}}\min_{x'\in\mathcal{X}}\phi(x',y)\le\min_{x\in\mathcal{X}}\max_{y'\in\mathcal{Y}}\phi(x,y').$$

■

Lemma 3.5 *The following statements are equivalent:*

(1) There exists $(x^*,y^*)\in\mathcal{X}\times\mathcal{Y}$ *such that for any* $x\in\mathcal{X}$, $y\in\mathcal{Y}$,

$$\max_{y\in\mathcal{Y}}\phi(x^*,y)\le\phi(x^*,y^*)\le\min_{x\in\mathcal{X}}\phi(x,y^*).$$

(2) The minimax equation holds:

$$\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}}\phi(x,y)=\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\phi(x,y)$$

and

$$x^*=\arg\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\phi(x,y),$$

$$y^*=\arg\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}}\phi(x,y).$$

Proof: (1) $\implies$ (2): Take min and max on the left and right respectively:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\phi(x^*,y)\le\phi(x^*,y^*)\le\max_{y\in\mathcal{Y}}\min_{x\in\mathcal{X}}\phi(x,y^*).$$

Together with lemma 3.4, we have

$$\phi(x^*, y^*) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y).$$

(2) $\implies$ (1): by the definition of x^* and y^*, we have

$$\max_{y \in \mathcal{Y}} \phi(x^*, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y) = \min_{x \in \mathcal{X}} \phi(x, y^*).$$

Thus,

$$\phi(x^*, y^*) \leq \max_{y \in \mathcal{Y}} \phi(x^*, y) = \min_{x \in \mathcal{X}} \phi(x, y^*) \leq \phi(x^*, y^*),$$

which implies that $\phi(x^*, y^*) = \max_{y \in \mathcal{Y}} \phi(x^*, y) = \min_{x \in \mathcal{X}} \phi(x, y^*)$. ■

Note: Then the proof of theorem follows by proving that Lemma 3.2 (1) holds.