Signed Support Recovery with the LASSO

Benjamin Noland

Background

Common problem: Want to estimate parameter vector $\beta^* \in \mathbb{R}^p$ in the linear model

$$y = X\beta^* + \epsilon,$$

where

- $y \in \mathbb{R}^n$ is a vector of observed responses,
- $X \in \mathbb{R}^{n \times p}$ is the design matrix, and
- $\epsilon \in \mathbb{R}^n$ is a zero-mean random vector representing the uncertainty in the model.

Background

- Problem is easy to solve in the *classical setting*: $p \leq n$. Simple linear algebra.
- Not so well understood when $p > n$. Case belongs to the active area of research known as *high-dimensional statistics*.

What to do when $p > n$?

- Assume the data is *truly low-dimensional*, i.e, that lot of the entries in β^* are actually zero.
- Define the *support* of β^* by

$$S(\beta^*) = \{i \in \{1, \dots, p\} : \beta_i^* \neq 0\},$$

  and let $k = |S(\beta^*)|$.
- Assume that $k \ll p$ (a *sparsity assumption* on β^*).
- Want to compute $S(\beta^*)$ to identify which variables are truly important.

The LASSO

A computationally tractible method for computing β^* in the high-dimensional setting is the *LASSO* (Least Absolute Shrinkage And Selection Operator):

$$\begin{array}{ll} \text{minimize} & \|y - X\beta\|_2^2 \\ \text{subject to} & \|\beta\|_1 \leq C_n \end{array},$$

where $C_n > 0$, or equivalently, as the solution to the unconstrained problem

$$\text{minimize } \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda_n\|\beta\|_1,$$

where $\lambda_n \geq 0$ is a *regularization parameter* that is in one-to-one correspondence with C_n via Lagrangian duality.

Project overview

- Restrict attention to random designs X.
- Explore the contributions of the following paper to support recovery using the LASSO:

  Wainwright, M. (2006). *Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (Lasso)*. Technical Report 709, Dept. Statistics, Univ. California, Berkeley
- See what happens when we replace the LASSO l_1-penalty term with a more general *elastic net* penalty

  $$\lambda_n \left(\frac{1}{2}(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1 \right),$$

  where $\alpha \in [0, 1]$.
- Conduct simulations and look at the results.

Signed support recovery

- Results from Wainwright paper provide necessary and sufficient conditions for the LASSO to recover the *signed support* of β^* with high probability.
- *Signed support* $\mathbb{S}_{\pm}(\beta)$ of $\beta \in \mathbb{R}^p$:

$$\mathbb{S}_{\pm}(\beta)_i = \begin{cases} +1 & \text{if } \beta_i > 0 \\ -1 & \text{if } \beta_i < 0 \\ 0 & \text{if } \beta_i = 0 \end{cases} \quad (i = 1, \ldots, p).$$

- Questions:
  - What relationships between n, p, and k yield a *unique* LASSO solution $\hat{\beta}$ satisfying $\mathbb{S}_{\pm}(\hat{\beta}) = \mathbb{S}_{\pm}(\beta^*)$?
  - For what relationships between n, p, and k does *no solution* of the LASSO yield the correct signed support?

Considered for both deterministic and random designs X.

Results from Wainwright paper

- Will quickly sketch out results from Wainwright paper for case of random design.
- Too much notation to define everything here. **See the paper!**

Sufficiency

Theorem 3 (Wainwright): Consider the linear model $y = X\beta^* + \epsilon$ with random Gaussian design $X \in \mathbb{R}^{n \times p}$ and error term $\epsilon \sim N_n(0, \sigma^2 I_n)$. Assume that the covariance matrix Σ satisfies certain regularity conditions (see paper). Consider the sequence (λ_n) of regularization parameters given by

$$\lambda_n = \lambda_n(\phi_p) = \sqrt{\frac{\phi_p \rho_u(\Sigma_{S^c|S})}{\gamma^2} \frac{2\sigma^2 \log p}{n}},$$

where $\phi_p \geq 2$. Suppose there exists $\delta > 0$ such that the sequences (n, p, k) and (λ_n) satisfy

$$\frac{n}{2k \log(p-k)} > (1+\delta)\theta_u(\Sigma)\left(1 + \frac{\sigma^2 C_{\min}}{\lambda_n^2 k}\right).$$

Sufficiency (cont.)

Then the following properties hold with probability $> 1 - c_1 \exp(-c_2 \min\{k, \log(p-k)\})$:

1. The LASSO has a unique solution $\hat{\beta} \in \mathbb{R}^p$ with $S(\hat{\beta}) \subseteq S(\beta^*)$ (i.e., its support is contained in the true support).
2. Define

$$g(\lambda_n) = c_3 \lambda_n \|\Sigma_{SS}^{-1/2}\|_\infty^2 + 20\sqrt{\frac{\sigma^2 \log k}{C_{\min} n}}.$$

   Then if $\beta_{\min} = \min_{i \in S} |\beta_i^*|$ satisfies $\beta_{\min} > g(\lambda_n)$, we have

$$\mathbb{S}_\pm(\hat{\beta}) = \mathbb{S}_\pm(\beta^*)$$

   and

$$\|\hat{\beta}_S - \beta_s^*\|_\infty \le g(\lambda_n).$$

Necessity

Theorem 4 (Wainwright): Consider the linear model $y = X\beta^* + \epsilon$ with random Gaussian design $X \in \mathbb{R}^{n \times p}$ and error term $\epsilon \sim N_n(0, \sigma^2 I_n)$. Assume that the covariance matrix Σ satisfies certain regularity conditions (see paper). Consider the sequence (λ_n) of regularization parameters from the previous theorem. Suppose there exists $\delta > 0$ such that the sequences (n, p, k) and (λ_n) satisfy

$$\frac{n}{2k \log(p-k)} < (1-\delta)\theta_l(\Sigma) \left(1 + \frac{\sigma^2 C_{\max}}{\lambda_n^2 k}\right).$$

Then with probability converging to one, no solution of the LASSO has the correct signed support.

Simulations

- Duplicated simulations from paper.
- Custom simulations where elastic net mixing parameter α was varied.

Simulations from paper

- Use LASSO solutions to estimate probability of correct signed support recovery in two cases:
  1. the design matrix X is drawn from a uniform Gaussian ensemble ($\Sigma = I_p$);
  2. the design matrix X is drawn from a non-uniform Gaussian ensemble where Σ is Toeplitz of the form

$$\Sigma = \begin{pmatrix} 1 & \mu & \mu^2 & \cdots & \mu^{p-2} & \mu^{p-1} \\ \mu & 1 & \mu & \mu^2 & \cdots & \mu^{p-2} \\ \mu^2 & \mu & 1 & \mu & \cdots & \mu^{p-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu^{p-1} & \cdots & \mu^3 & \mu^2 & \mu & 1 \end{pmatrix},$$

     where $\mu = 0.10$.

Simulations from paper (cont.)

- Consider problem sizes $p \in \{128, 256, 512\}$.
- Sparsity regimes:
  - *linear sparsity*: $k(p) = \lceil \gamma p \rceil$ for some $\gamma \in (0, 1)$;
  - *sublinear sparsity*: $k(p) = \lceil \gamma p / \log(\gamma p) \rceil$ for some $\gamma \in (0, 1)$, and
  - *fractional power sparsity*: $k(p) = \lceil \gamma p^{\delta} \rceil$ for some $\gamma, \delta \in (0, 1)$.

  where $\gamma = 0.40$ and $\delta = 0.75$.

Simulations from paper (cont.)

- We consider models fit using the family of regularization parameters λ_n given by

$$\lambda_n = \sqrt{\frac{2\sigma^2 \log k \log(p-k)}{n}},$$

where $\sigma = 0.5$ is a fixed noise level.

- For this choice of λ_n, Theorem 4 predicts failure with high probability for sequences (n, p, k) satisfying

$$\frac{n}{2k \log(p-k)} < \theta_l(\Sigma),$$

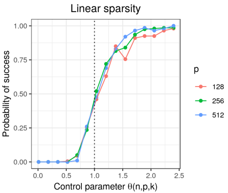
and Theorem 3 predicts success with high probability for sequences (n, p, k) such that

$$\frac{n}{2k \log(p-k)} > \theta_u(\Sigma)$$

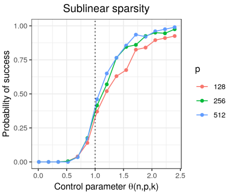
Simulations from paper (cont.)

1. For X drawn from uniform Gaussian ensemble, $\theta_l(I_p) = \theta_l(I_p) = 1$, so predict failure for $\theta < 1$, success for $\theta > 1$.
2. For X drawn from non-uniform Gaussian ensemble with Toeplitz covariance, also have $\theta_l(\Sigma) \approx 1$ and $\theta_u(\Sigma) \approx 1$.

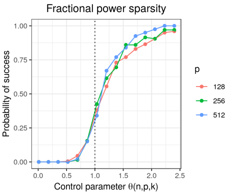
Simulations from paper (cont.)



(a) Linear



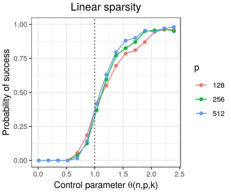
(b) Sublinear



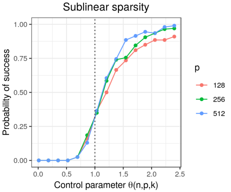
(c) Fractional power

Figure: Uniform Gaussian ensemble with LASSO

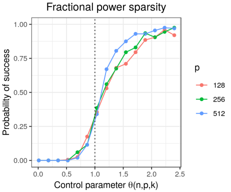
Simulations from paper (cont.)



(a) Linear



(b) Sublinear



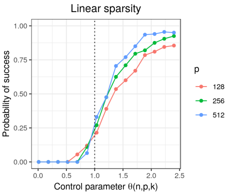
(c) Fractional power

Figure: Non-uniform Gaussian ensemble with LASSO

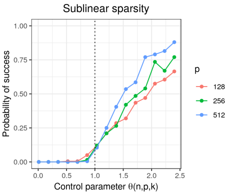
Custom simulations

- Same as for uniform Gaussian ensemble simulation from paper, but with elastic net penalty ($\alpha = 0.75$ and $\alpha = 0.50$).
- Unlike with LASSO, have no theoretical guarantees, but intuitively expect the theoretical results for LASSO to deteriorate as the contribution of the l_1-penalty term is diminished (i.e., as α gets smaller).
- This seems to be what we get!

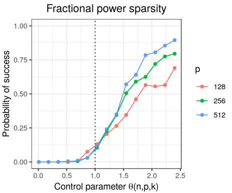
Custom simulations (cont.)



(a) Linear



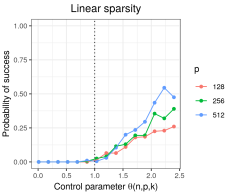
(b) Sublinear



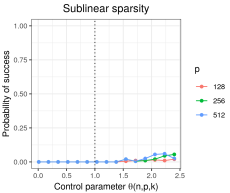
(c) Fractional power

Figure: Uniform Gaussian ensemble, $\alpha = 0.75$

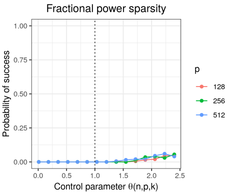
Custom simulations (cont.)



(a) Linear



(b) Sublinear



(c) Fractional power

Figure: Uniform Gaussian ensemble, $\alpha = 0.50$