Bounds on l_2-error for hard sparse models

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics
Lancaster University

Contents

- Consider the vector matrix pair $(y, X) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$ linked by the linear model
$$y = X\theta^* + \varepsilon$$
where $\varepsilon \in \mathbb{R}^n$ is the noise vector.
- Aim is to ensure $\|\hat{\theta} - \theta^*\|_2^2$ is small.
- The latter requires conditions on the random design matrix X.
- Further, the choice of (n, d, s) is important in ensuring that $\|\hat{\theta} - \theta^*\|_2^2$ is small.
- In the presence of noise, a natural extension to the basis pursuit program was accomplished by minimizing a weighted combination of the data-fidelity term $\|y - X\beta\|_2^2$ with the L_1-norm penalty:

$$\hat{\theta} \in \arg\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2n} \|y - X\theta\|_2^2 + \lambda_n \|\theta\|_1 \right\}, \quad \text{(Lagrangian Lasso)} \quad (1)$$

or equivalently, the constrained forms of Lasso

$$\min_{\theta\in\mathbb{R}^d}\left\{\frac{1}{2n}\|y-X\theta\|_2^2\right\},\quad \text{s.t.}\quad \|\theta\|_1\leq R,\quad \text{(Constrained Lasso)} \tag{2}$$

$$\min_{\theta\in\mathbb{R}^d}\|\theta\|_1\quad \text{s.t.}\quad \frac{1}{2n}\|y-X\theta\|_2^2\leq b^2,\quad \text{(Relaxed BP)} \tag{3}$$

- Let $\Delta := \hat{\theta}-\theta^*$. The goal is thus to bound $\|\Delta\|_2^2$.
- In low dimensions, a bound on $\|X\Delta\|_2^2$ would provide guarantees for $\|\Delta\|_2^2$.
- This, however, no longer holds true in high dimensions, as X has a non-trivial nullspace.
- Perfect recovery not feasible in noisy settings.
- We thus focus on bounding the L_2 error $\|\hat{\theta}-\theta^*\|_2$.

- In noisy settings, the required condition is slightly stronger than the Restricted Nullspace Property - namely that the restricted eigenvalues of the matrix $\frac{X'X}{n}$ are lower bounded over a cone.
- In particular, for a constant $\alpha \geq 1$, let us define the set

$$\mathbb{C}_\alpha(S) := \{\Delta \in \mathbb{R}^d \mid \|\Delta_{S^c}\|_1 \leq \alpha \|\Delta_S\|_1\}$$

We then say, the matrix X satisfies the Restricted Eigenvalue Condition over S with parameters (κ, α), if

$$\frac{1}{n}\|X\Delta\|_2^2 \geq \kappa \|\Delta\|_2^2, \quad \forall \Delta \in \mathbb{C}_\alpha(S)$$

Holder's inequality

If $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$ and $\mathbb{E}\left(|Y|^p\right) \leq \infty$ and $\mathbb{E}\left(|Z|^q\right) \leq \infty$ then

$$\mathbb{E}\left(|YZ|\right) \leq \left[\mathbb{E}\left(|Y|^p\right)\right]^{1/p} \left[\mathbb{E}\left(|Z|^q\right)\right]^{1/q}$$

[See White (2014)]

A special case of the Holder's inequality is the Cauchy-Schwarz inequality as follows:

Cauchy-Schwarz inequality

If $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, such that $p = q = 2$ and $\mathbb{E}\left(|Y|^2\right) \leq \infty$ and $\mathbb{E}\left(|Z|^2\right) \leq \infty$ then

$$\mathbb{E}\left(|YZ|\right) \leq \left[\mathbb{E}\left(|Y|^2\right)\right]^{1/2} \left[\mathbb{E}\left(|Z|^2\right)\right]^{1/2}$$

[See White (2014)]

Example

For $u \in \mathbb{R}^s$, prove

$$\|u\|_1 \leq \sqrt{s}\|u\|_2$$

Solution:

$$\begin{aligned}
\|u\|_1 &= \sum_{i=1}^{s} |u_i| \\
&= \sum_{i=1}^{s} |u_i.1| \\
&\leq \left(\sum_{i=1}^{s} |u_i|^2\right)^{1/2} \left(\sum_{i=1}^{s} 1\right)^{1/2} \\
&= \sqrt{s}\|u\|_2
\end{aligned}$$

Following Wainwright (2019), we now state a result that provides a bound on the error $\|\hat{\theta} - \theta^*\|_2$ in the case of a hard sparse vector θ^*.
Let us impose the following conditions:

A1. The vector θ^* is supported on a subset $S \subseteq \{1, 2, \cdots, d\}$ with $|S| = s$.

A2. The design matrix satisfies the Restricted Eigenvalue condition over S with parameters $(\kappa, 3)$.

Theorem (L_2-error between $\hat{\theta}$(1) and θ^*)

Any solution to the Lagrangian Lasso (1) with regularization parameter lower bounded as $\lambda_n \geq 2\|\frac{X'\varepsilon}{n}\|_\infty$, satisfies the bound

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{3}{\kappa}\sqrt{s}\lambda_n$$

Proof: Recall from earlier that $\|\hat{\theta}\|_1 \leq \|\theta^*\|_1$. Thus, from (1), we get

$$\frac{1}{2n}\|y - X\hat{\theta}\|_2^2 + \lambda_n\|\hat{\theta}\|_1 \leq \frac{1}{2n}\|y - X\theta^*\|_2^2 + \lambda_n\|\theta^*\|_1$$

Note from earlier that $\Delta := \hat{\theta} - \theta^*$, and furthermore, for two column vectors u and v, we can expand

$$\|u - v\|_2^2 = \|u\|_2^2 - 2u'v + \|v\|_2^2$$

Thus,

$$\frac{1}{2n}\|y\|_2^2 - \frac{1}{n}(X\hat{\theta})'y + \frac{1}{2n}\|X\hat{\theta}\|_2^2 + \lambda_n\|\hat{\theta}\|_1 \leq$$
$$\frac{1}{2n}\|y\|_2^2 - \frac{1}{n}(X\theta^*)'y + \frac{1}{2n}\|X\theta^*\|_2^2 + \lambda_n\|\theta^*\|_1$$

which simplifies to

$$\frac{1}{n}\|X\Delta\|_2^2 \leq \frac{2(X'\varepsilon)'\Delta}{n} + 2\lambda_n\{\|\theta^*\|_1 - \|\hat{\theta}\|_1\} \tag{4}$$

Expressing θ^* under s-sparse condition, we get

$$\begin{aligned}
\|\theta^*\|_1 - \|\hat{\theta}\|_1 &= \|\theta^*_S + \theta^*_{S^c}\|_1 - \|\theta^* + \Delta\|_1 \\
&= \|\theta^*_S + \theta^*_{S^c}\|_1 - \|\theta^*_S + \theta^*_{S^c} + \Delta_S + \Delta_{S^c}\|_1 \\
&= \|\theta^*_S\|_1 - \|\theta^*_S + \Delta_S + \Delta_{S^c}\|_1 \\
&= \|\theta^*_S\|_1 - \|\theta^*_S + \Delta_S\|_1 - \|\Delta_{S^c}\|_1
\end{aligned}$$

Substituting the above expression in (4), we obtain

$$\frac{1}{n}\|X\Delta\| \leq \frac{2(X'\varepsilon)'\Delta}{n} + 2\lambda_n\{\|\theta^*_S\|_1 - \|\theta^*_S + \Delta_S\|_1 - \|\Delta_{S^c}\|_1\} \tag{5}$$

Using Holder's inequality we have

$$\frac{2(X'\varepsilon)'\Delta}{n} \leq \left\|\frac{2(X'\varepsilon)'}{n}\right\|_\infty \|\Delta\|_1$$

which upon substitution in (5) yields

$$\frac{1}{n}\|X\Delta\| \leq \left\|\frac{2(X'\varepsilon)'}{n}\right\|_\infty \|\Delta\|_1 + 2\lambda_n\{\|\theta^*_S\|_1 - \|\theta^*_S + \Delta_S\|_1 - \|\Delta_{S^c}\|_1\} \quad (6)$$

Furthermore, from the triangle inequality we have

$$\|\theta^*_S + \Delta_S\|_1 \leq \|\theta^*_S\|_1 + \|\Delta_S\|_1$$

thus,

$$\|\theta^*_S\|_1 - \|\theta^*_S + \Delta_S\|_1 - \|\Delta_{S^c}\|_1 \leq \|\theta^*_S\|_1 - \|\theta^*_S\|_1 + \|\Delta_S\|_1 - \|\Delta_{S^c}\|_1 \quad (7)$$

where plugging in (7) into (6) yields

$$\frac{1}{n}\|X\Delta\| \leq 2\left\|\frac{(X'\varepsilon)'}{n}\right\|_\infty \|\Delta\|_1 + 2\lambda_n\{\|\Delta_S\|_1 - \|\Delta_{S^c}\|_1\}. \quad (8)$$

Recall the condition $\lambda_n \geq 2\left\|\frac{X'\varepsilon}{n}\right\|_\infty$. Substituting the choice of λ_n into equation (8) yields

$$\frac{1}{n}\|X\Delta\| \leq \lambda_n\{3\|\Delta_S\|_1 - \|\Delta_{S^c}\|_1\} \tag{9}$$

Inequality (9) shows that $\Delta \in \mathbb{C}_3(S)$ - i.e. since

$$0 \leq \frac{1}{n}\|X\Delta\| \leq \lambda_n\{3\|\Delta_S\|_1 - \|\Delta_{S^c}\|_1\}$$

then $\|\Delta_{S^c}\|_1 \leq 3\|\Delta_S\|_1$, so that the Restricted Eigenvalue condition may be applied. In other words,

$$\kappa\|\Delta\|_2^2 \leq 3\lambda_n\sqrt{s}\|\Delta\|_2,$$

which implies the Theorem.

Theorem (L_2-error between $\hat{\theta}$(2) and θ^*)

Any solution of the Constrained Lasso (2) with $R = \|\theta^*\|_1$ satisfies the bound

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{4}{\kappa}\sqrt{s}\left\|\frac{X'\varepsilon}{n}\right\|_\infty$$

Proof: Given the choice of $t = \|\theta^*\|_1$, the target vector θ^* is feasible. Since $\hat{\theta}$ is optimal, we have the inequality

$$\frac{1}{2n}\|y - X\hat{\theta}\|_2^2 \leq \frac{1}{2n}\|y - X\theta^*\|_2^2.$$

As before, define the vector $\Delta := \hat{\theta} - \theta^*$. As in the proof of the previous Theorem, we can easily show

$$\frac{1}{n}\|X\Delta\|_2^2 \leq \frac{2(X'\varepsilon)'\Delta}{n} \tag{10}$$

Once again, we know from Holder's inequality that

$$\frac{2(X'\varepsilon)'\Delta}{n} \leq 2\left\|\frac{(X'\varepsilon)'}{n}\right\|_{\infty}\|\Delta\|_1$$

which when plugged into (10), yields

$$\frac{1}{n}\|X\Delta\|_2^2 \leq 2\left\|\frac{(X'\varepsilon)'}{n}\right\|_{\infty}\|\Delta\|_1 \tag{11}$$

On the other hand, following the proof of the analysis of basis pursuit in earlier sessions we obtain

$$\|\Delta_{S^c}\|_1 \leq \|\Delta_S\|_1$$

under our constraint on θ and the error $\Delta \in \mathbb{C}_1(S)$, whence

$$\|\Delta\|_1 = \|\Delta_S\|_1 + \|\Delta_{S^c}\|_1 \leq 2\|\Delta_S\|_1 \leq 2\sqrt{s}\|\Delta\|_2$$

Because $\mathbb{C}_1(S) \subset \mathbb{C}_3(S)$, we may apply the Restricted Eigenvalue condition to the left hand side of (11) and obtain

$$\frac{\|X\Delta\|_2^2}{n} \geq \kappa \|\Delta\|_2^2$$

Putting together all the pieces yields the bound of the Theorem.

Theorem (L_2-error between $\hat{\theta}$(3) and θ^*)

Any solution to the Relaxed Basis Pursuit program (3) with $b^2 \geq \frac{\|\varepsilon\|_2^2}{2n}$*, satisfies the bound*

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{4}{\kappa}\sqrt{s}\left\|\frac{X'\varepsilon}{n}\right\| + \frac{2}{\sqrt{\kappa}}\sqrt{b^2 - \frac{\|\varepsilon\|_2^2}{2n}}$$

Proof: Note that

$$\frac{1}{2n}\|y - X\theta^*\|_2^2 = \frac{\|\varepsilon\|_2^2}{2n} \leq b^2$$

where the inequality on the right hand side of the above equation is the consequence of the assumed choice of b. Thus, the target vector θ^* is feasible, and since $\hat{\theta}$ is optimal, we have $\|\hat{\theta}\|_1 \leq \|\theta^*\|_1$. As previously reasoned, the error vector $\Delta := \hat{\theta} - \theta^*$ must then belong to the cone $\mathbb{C}_1(S)$. Now by the feasibility of $\hat{\theta}$, we have

$$\frac{1}{2n}\|y - X\hat{\theta}\|_2^2 \leq b^2 = \frac{1}{2n}\|y - X\theta^*\|_2^2 + \left(b^2 - \frac{\|\varepsilon\|_2^2}{2n}\right)$$

where rearranging as in the proofs of previous Theorems yields the modified inequality

$$\frac{\|X\Delta\|_2^2}{n} \leq 2\frac{(X'\varepsilon)'\Delta}{n} + 2\left(b^2 - \frac{\|\varepsilon\|_2^2}{2n}\right)$$

As in the previous Theorem, we may apply to Holder's inequality on the right hand side of the equation to yield

$$2\frac{\varepsilon' X\Delta}{n} \leq 2\left\|\frac{\varepsilon' X}{n}\right\|_\infty \|\Delta\|_1$$

we obtain

$$\frac{\|X\Delta\|_2^2}{n} \leq 2\left\|\frac{\varepsilon' X}{n}\right\|_\infty \|\Delta\|_1 + 2\left(b^2 - \frac{\|\varepsilon\|_2^2}{2n}\right)$$

Moreover, we know that

$$\|\Delta\|_1 = \|\Delta_S\|_1 + \|\Delta_{S^c}\|_1 \leq 2\|\Delta_S\|_1 \leq 2\sqrt{s}\|\Delta\|_2$$

Therefore,

$$\frac{\|X\Delta\|_2^2}{n} \leq 4\sqrt{s}\left\|\frac{\varepsilon' X}{n}\right\|_\infty \|\Delta\|_2 + 2\left(b^2 - \frac{\|\varepsilon\|_2^2}{2n}\right)$$

Applying the Restricted Eigenvalue condition as in the previous Theorem to the left hand side leads to

$$\kappa\|\Delta\|_2^2 \leq 4\sqrt{s}\left\|\frac{\varepsilon' X}{n}\right\|_\infty \|\Delta\|_2 + 2\left(b^2 - \frac{\|\varepsilon\|_2^2}{2n}\right)$$

which implies that

$$\|\Delta\|_2 \leq \frac{8}{\kappa}\sqrt{s}\left\|\frac{X'\varepsilon}{n}\right\|_\infty + \frac{2}{\sqrt{\kappa}}\sqrt{b^2 - \frac{\|\varepsilon\|_2^2}{2n}}$$

as claimed.

References

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.