Sharper Basis Pursuit Bounds

João F. C. Mota

August 4, 2015

In this document, we show how to compute a bound for Basis Pursuit (BP) slightly sharper than the one in [1]. Our approach is similar to the one in section 4.3 of [2].

Problem statement. Let $x^\star \in \mathbb{R}^n$. Suppose we acquire m linear measurements of $x^\star$ as $y = Ax^\star$, where $A \in \mathbb{R}^{m\times n}$, and attempt to reconstruct $x^\star$ by solving

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Ax = y\,, \end{array} \tag{1}$$

where $\|x\|_1 := \sum_{i=1}^n |x_i|$ is the ℓ_1-norm. Assuming that each entry of the matrix A is drawn i.i.d. from a zero-mean Gaussian distribution with variance $1/m$, our goal is to determine the number of measurements m that (1) requires to reconstruct $x^\star$ with high probability (over the set of matrices A).

A result by Chandrasekaran et al. Corollary 3.3 of [1] together with Proposition 3.6 and Jensen's inequality establish that if

- $y = Ax^\star$ for $A \in \mathbb{R}^{m\times n}$
- Each entry of A is drawn i.i.d. from $\mathcal{N}(0, 1/m)$
- $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and there holds $0 \notin f(x^\star)$
- $$\begin{array}{lll} \widehat{x} \in & \underset{x}{\arg\min} & f(x) \\ & \text{s.t.} & Ax = y \end{array} \tag{2}$$
- $$m \geq \mathbb{E}_g\Big[\text{dist}\big(g, \text{cone}\,\partial f(x^\star)\big)^2\Big]\,, \tag{3}$$

  where $g \sim \mathcal{N}(0, I_n)$ is a vector of independent Normal entries and $\text{dist}(x, C) = \arg\min_z\{\|z - x\|^2 : z \in C\}$ is the distance of a point x to a set C then, with probability at least $1 - \exp\big(-\frac{1}{2}(m - \sqrt{m})^2\big)$, $\widehat{x}$ is the unique solution of (2) and equals $x^\star$.

Since the distance of a point to $\text{cone}\,\partial f(x^\star) = \{td : d \in \partial f(x^\star),\, t \geq 0\}$ does not usually have a closed-form expression, we will use instead

$$m \geq \min_{t\geq 0}\; \mathbb{E}_g\Big[\text{dist}\big(g, t\,\partial f(x^\star)\big)^2\Big]\,. \tag{4}$$

It is clear that (4) implies (3). In this document, we show how to compute the right-hand side of (4) numerically for $f(x) = \|x\|_1$. Our method is a variant of the method in [2] and should give similar results.

Some notation. In our computations below, we will denote the support of $x^\star$ and its complement on $\{1,\dots,n\}$ with

$$I := \{i \,:\, x_i^\star \neq 0\} \qquad\qquad I^c := \{i \,:\, x_i^\star = 0\}\,,$$

respectively. The cardinality of I or, in other words, the sparsity of $x^\star$ will be denoted with $s := |I|$. We will also make use of the Q-function, defined as

$$Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\Big(-\frac{g^2}{2}\Big)\,\mathrm{d}g\,.$$

Using Leibniz rule, it can be checked that the derivative of the Q-function at a point x is given by

$$\frac{d}{dx}Q(x) = -\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2}\Big)\,. \tag{5}$$

Sample complexity of Basis Pursuit. The subgradient of $f(x) = \|x\|_1$ at $x^\star$ is

$$\partial\|x^\star\|_1 = \Big(\partial|x_1^\star|,\, \partial|x_2^\star|,\, \dots,\, \partial|x_n^\star|\Big)\,,$$

where

$$\partial|x_i^\star| = \begin{cases} \operatorname{sign}(x_i^\star) & , \text{ if } i \in I \\ \big[-1,1\big] & , \text{ if } i \in I^c \end{cases}$$

Recall that we are using the following notation: for a given set S and a scalar t, $t\,S := \{t\,s \,:\, s \in S\}$. We can now compute the right-hand side of (4) for $f(x) = \|x\|_1$:

$$\min_{t\ge 0}\ \mathbb{E}_g\Big[\operatorname{dist}\big(g, t\,\partial f(x^\star)\big)^2\Big] = \min_{t\ge 0}\left\{\sum_{i\in I}\mathbb{E}_g\Big[\operatorname{dist}\big(g_i\,,\, t\operatorname{sign}(x_i^\star)\big)^2\Big] + \sum_{i\in I^c}\mathbb{E}_g\Big[\operatorname{dist}\big(g_i\,,\, \big[-t,t\big]\big)^2\Big]\right\}. \tag{6}$$

The first term on the right-hand side of (6) can be computed in closed-form:

$$\begin{aligned}\sum_{i\in I}\mathbb{E}_g\Big[\operatorname{dist}\big(g_i\,,\, t\operatorname{sign}(x_i^\star)\big)^2\Big] &= \sum_{i\in I}\mathbb{E}_g\Big[\big(g_i - t\operatorname{sign}(x_i^\star)\big)^2\Big] = \sum_{i\in I}\Big(\mathbb{E}_g\big[g_i^2\big] - t\operatorname{sign}(x_i^\star)\,\mathbb{E}_g\big[g_i\big] + t^2\Big)\\ &= \sum_{i\in I}\big(1+t^2\big)\\ &= s + s\,t^2\,,\end{aligned} \tag{7}$$

where we used the linearity of the expected value and the fact that g_i is a random variable with Normal distribution.

The second term of (6) does not have a closed-form expression, but we can express it as a function of the Q-function:

$$\begin{aligned}&\sum_{i\in I^c}\mathbb{E}_g\Big[\operatorname{dist}\big(g_i\,,\, \big[-t,t\big]\big)^2\Big]\\ &\quad= \sum_{i\in I^c}\frac{2}{\sqrt{2\pi}}\int_t^{+\infty}(u-t)^2\exp\Big(-\frac{u^2}{2}\Big)\,\mathrm{d}u\\ &\quad= \sum_{i\in I^c}\frac{2}{\sqrt{2\pi}}\left[\int_t^{+\infty}u^2\exp\Big(-\frac{u^2}{2}\Big) - 2t\int_t^{+\infty}u\exp\Big(-\frac{u^2}{2}\Big) + t^2\int_t^{+\infty}\exp\Big(-\frac{u^2}{2}\Big)\right]\end{aligned} \tag{8}$$

$$= \sum_{i \in I^c} \frac{2}{\sqrt{2\pi}} \left[t \exp\Big(-\frac{t^2}{2}\Big) - 2t \int_t^{+\infty} u \exp\Big(-\frac{u^2}{2}\Big) + (1+t^2) \int_t^{+\infty} \exp\Big(-\frac{u^2}{2}\Big) \right] \tag{9}$$

$$= \sum_{i \in I^c} \frac{2}{\sqrt{2\pi}} \left[(1+t^2) \int_t^{+\infty} \exp\Big(-\frac{u^2}{2}\Big) - t \exp\Big(-\frac{t^2}{2}\Big) \right] \tag{10}$$

$$= 2(n-s)\big(1+t^2\big)Q(t) - \frac{2(n-s)}{\sqrt{2\pi}}\, t\, \exp\Big(-\frac{t^2}{2}\Big)\,. \tag{11}$$

From (8) to (9), we integrated the first term by parts:

$$\int_t^{+\infty} u^2 \exp\Big(-\frac{u^2}{2}\Big)\, \mathrm{d}u = \Big[-u \exp\Big(-\frac{u^2}{2}\Big)\Big]_t^{+\infty} + \int_t^{+\infty} \exp\Big(-\frac{u^2}{2}\Big)\, \mathrm{d}u = t\, \exp\Big(-\frac{t^2}{2}\Big) + \int_t^{+\infty} \exp\Big(-\frac{u^2}{2}\Big)\, \mathrm{d}u\,.$$

From (9) to (10), we computed the integral in the second term:

$$\int_t^{+\infty} u \exp\Big(-\frac{u^2}{2}\Big)\, \mathrm{d}u = \Big[-\exp\Big(-\frac{u^2}{2}\Big)\Big]_t^{+\infty} = \exp\Big(-\frac{t^2}{2}\Big)\,.$$

Using (7) and (11) in (6), we obtain

$$\min_{t \geq 0}\ \mathbb{E}_g\Big[\mathrm{dist}\big(g, t\,\partial f(x^\star)\big)^2\Big] = \min_{t \geq 0}\ s + s\,t^2 + 2(n-s)\big(1+t^2\big)Q(t) - \frac{2(n-s)}{\sqrt{2\pi}}\, t\, \exp\Big(-\frac{t^2}{2}\Big) =: \min_{t \geq 0}\ h(t)\,, \tag{12}$$

where

$$h(t) := s + s\,t^2 + 2(n-s)\big(1+t^2\big)Q(t) - \sqrt{\frac{2}{\pi}}(n-s)\, t\, \exp\Big(-\frac{t^2}{2}\Big)\,.$$

We now compute the first- and second-order derivatives of $h(t)$ and check that $h(t)$ is convex. There holds

$$\begin{aligned}
\frac{d}{dt}h(t) &= 2st + 4(n-s)\,t\,Q(t) + 2(n-s)\big(1+t^2\big)\frac{d}{dt}Q(t) - \sqrt{\frac{2}{\pi}}(n-s)\, \exp\Big(-\frac{t^2}{2}\Big) + \sqrt{\frac{2}{\pi}}(n-s)\, t^2\, \exp\Big(-\frac{t^2}{2}\Big) \\
&= 2st + 4(n-s)\,t\,Q(t) + 2(n-s)\big(1+t^2\big)\frac{d}{dt}Q(t) + \big(t^2-1\big)\sqrt{\frac{2}{\pi}}(n-s)\, \exp\Big(-\frac{t^2}{2}\Big) \\
&= 2st + 4(n-s)\,t\,Q(t) - \big(t^2+1\big)\sqrt{\frac{2}{\pi}}(n-s) \exp\Big(-\frac{t^2}{2}\Big) + \big(t^2-1\big)\sqrt{\frac{2}{\pi}}(n-s)\, \exp\Big(-\frac{t^2}{2}\Big) \\
&= 2st + 4(n-s)\,t\,Q(t) - 2(n-s)\sqrt{\frac{2}{\pi}} \exp\Big(-\frac{t^2}{2}\Big) \\
\frac{d^2}{dt^2}h(t) &= 2s + 4(n-s)Q(t) + 4(n-s)t\frac{d}{dt}Q(t) + 2(n-s)\sqrt{\frac{2}{\pi}}\, t\, \exp\Big(-\frac{t^2}{2}\Big) \\
&= 2s + 4(n-s)Q(t) - 2(n-s)\sqrt{\frac{2}{\pi}}\, t\, \exp\Big(-\frac{t^2}{2}\Big) + 2(n-s)\sqrt{\frac{2}{\pi}}\, t\, \exp\Big(-\frac{t^2}{2}\Big) \\
&= 2s + 4(n-s)Q(t)\,.
\end{aligned}$$

Since $Q(t) > 0$ for all t, $h(t)$ is strictly convex. To solve (12), we can apply Newton-Raphson's method to solve $\frac{dh}{dt}(t) = 0$:

$$t^{k+1} = t^k - \frac{\dot{h}\big(t^k\big)}{\ddot{h}\big(t^k\big)}\,, \tag{13}$$

where $\dot{h} := \frac{dh}{dt}$ and $\ddot{h} := \frac{dh^2}{dt^2}$. Note that we are not projecting each iterate onto $t \geq 0$. In fact, the zero $t^\star$ of $\dot{h}$ is always positive. To see why, note that $\dot{h}(0) = -2(n-s)\sqrt{2/\pi} < 0$ and $\lim_{t \to +\infty} \dot{h}(t) = +\infty$. Given that $\ddot{h}(t) \geq 0$ for all t, it means that $t^\star > 0$. Hence, the constrained problem in (12) is equivalent to the one we obtain by removing the constraint $t \geq 0$. A good starting point is $t^0 = \sqrt{2\log(n/s)}$.

Implementation issues. In some computational platforms, we have access not to the Q-function, but to the complementary error function

$$\operatorname{erfc}(x) := \frac{2}{\sqrt{\pi}} \int_x^{+\infty} \exp\left(-u^2\right) \mathrm{d}u\,.$$

By a simple change of variables, we can check that

$$\begin{aligned} Q(x) &= \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left(-\frac{g^2}{2}\right) \mathrm{d}g = \frac{1}{\sqrt{\pi}} \int_{\frac{x}{\sqrt{2}}}^{+\infty} \exp\left(-u^2\right) \mathrm{d}u \\ &= \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right). \end{aligned}$$

References

[1] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Foundations on Computational Mathematics*, vol. 12, pp. 805–849, 2012.

[2] D. Amelunxen, M. Lotz, M. Mccoy, and J. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, pp. 1–71, 2014.