Supplemental Materials for:
"Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems"

Yuxin Chen * Emmanuel J. Candès *†

May 14, 2015

Abstract

This document presents the proof of the universal stability guarantees for Theorem 2 given in the paper " Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems".

1 Universal stability guarantees

In the main text, Theorem 2 has been proved for the situation where the planted solution $\boldsymbol{x}$ is fixed independent of the design vectors $\{\boldsymbol{a}_i\}$. This section proves a more universal theory: once the design vectors are selected and fixed, then with high probability TWF succeeds simultaneously for all $\boldsymbol{x} \in \mathbb{R}^n$.

Since the main text already delivers universal guarantees for the iterative refinement stage, it only remains to justify universality for truncated spectral initialization. In fact, it suffices to prove that in the absence of noise,

$$\|\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]\| \leq \delta\|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n \tag{1}$$

holds for an arbitrary small constant $\delta > 0$, as all remaining steps presented in the main text readily carry over here. To ease presentation, we shall assume $\|\boldsymbol{x}\| = 1$ from now on and denote

$$\boldsymbol{Y}_{\boldsymbol{x}} := \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{a}_i\boldsymbol{a}_i^\top \left(\boldsymbol{a}_i^\top \boldsymbol{x}\right)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \leq \alpha_y\right\}}, \tag{2}$$

which in expectation gives

$$\mathbb{E}[\boldsymbol{Y}_{\boldsymbol{x}}] = \beta_1 \boldsymbol{x}\boldsymbol{x}^\top + \beta_2 \boldsymbol{I}. \tag{3}$$

Here, $\beta_1 := \mathbb{E}\left[\xi^4 \mathbf{1}_{\{|\xi| \leq \alpha_y\}}\right] - \mathbb{E}\left[\xi^2 \mathbf{1}_{\{|\xi| \leq \alpha_y\}}\right]$ and $\beta_2 := \mathbb{E}\left[\xi^2 \mathbf{1}_{\{|\xi| \leq \alpha_y\}}\right]$ where ξ is a standard normal.

To prove universality, we discretize the unit sphere using an ϵ-net $\mathcal{N}_\epsilon$ $(0 < \epsilon < 1/2)$ of cardinality $\left(1 + \frac{2}{\epsilon}\right)^n$ so that for any unit vector $\boldsymbol{x}$, there exists an $\boldsymbol{x}_0 \in \mathcal{N}_\epsilon$ obeying $\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \epsilon$. As has been demonstrated in the main text, for any fix $\boldsymbol{x}_0 \in \mathbb{R}^n$,

$$\|\boldsymbol{Y}_{\boldsymbol{x}_0} - \mathbb{E}[\boldsymbol{Y}_{\boldsymbol{x}_0}]\| \leq \frac{\delta}{2} \tag{4}$$

holds with probability $1 - \exp(-\Omega(m))$. Taking the union bound over $\mathcal{N}_\epsilon$ yields

$$\left\|\boldsymbol{Y}_{\boldsymbol{x}_0} - \beta_1 \boldsymbol{x}_0 \boldsymbol{x}_0^\top - \beta_2 \boldsymbol{I}\right\| \leq \frac{\delta}{2}, \quad \forall \boldsymbol{x}_0 \in \mathcal{N}_\epsilon \tag{5}$$

with probability $1 - \left(1 + \frac{2}{\epsilon}\right)^n \exp(-\Omega(m))$. It then boils down to obtaining uniform control over $\|\boldsymbol{Y}_{\boldsymbol{x}} - \boldsymbol{Y}_{\boldsymbol{x}_0}\|$. To this end, we decompose $\boldsymbol{Y}_{\boldsymbol{x}} - \boldsymbol{Y}_{\boldsymbol{x}_0}$ in the following manner

$$\boldsymbol{Y}_{\boldsymbol{x}} - \boldsymbol{Y}_{\boldsymbol{x}_0} = \frac{1}{m}\left(\sum_{i\in\mathcal{I}_1} + \sum_{i\in\mathcal{I}_2} + \sum_{i\in\mathcal{I}_3} + \sum_{i\in\mathcal{I}_4}\right) \boldsymbol{a}_i\boldsymbol{a}_i^\top \cdot \left\{(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \leq \alpha_y\right\}} - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \leq \alpha_y\right\}}\right\},$$

*Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.
†Department of Mathematics, Stanford University, Stanford, CA 94305, U.S.A.

where the index sets are defined as

$$\begin{aligned}
\mathcal{I}_1 &:= \left\{ i \mid \left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0) \right| < \zeta \text{ and } |\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y - \zeta \right\}; \\
\mathcal{I}_2 &:= \left\{ i \mid \left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0) \right| < \zeta \text{ and } |\boldsymbol{a}_i^\top \boldsymbol{x}_0| \ge \alpha_y + \zeta \right\}; \\
\mathcal{I}_3 &:= \left\{ i \mid \left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0) \right| < \zeta \text{ and } |\boldsymbol{a}_i^\top \boldsymbol{x}_0| \in [\alpha_y \pm \zeta] \right\}; \\
\mathcal{I}_4 &:= \left\{ i \mid \left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0) \right| \ge \zeta \right\}.
\end{aligned}$$

Below we control each of these cases separately.

Case 1. When $i \in \mathcal{I}_1$, $\boldsymbol{a}_i^\top \boldsymbol{x}$ is absolutely controlled in the sense that

$$|\boldsymbol{a}_i^\top \boldsymbol{x}| \le |\boldsymbol{a}_i^\top \boldsymbol{x}_0| + \zeta \le \alpha_y,$$

and hence both indicator variables $\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y\right\}}$ and $\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y\right\}}$ are active. In addition, the constraint that specifies $\mathcal{I}_1$ gives

$$\left| (\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \right| = \left| \boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0) \right| \cdot \left| \boldsymbol{a}_i^\top \boldsymbol{x} + \boldsymbol{a}_i^\top \boldsymbol{x}_0 \right| \le \zeta \cdot 2\alpha_y.$$

Consequently, one can write

$$\begin{aligned}
&\left\| \frac{1}{m} \sum_{i \in \mathcal{I}_1} \boldsymbol{a}_i \boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y\right\}} - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y\right\}} \right\} \right\| = \left\| \frac{1}{m} \sum_{i \in \mathcal{I}_1} \boldsymbol{a}_i \boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \right\} \right\| \\
&\quad \le 2\zeta\alpha_y \left\| \frac{1}{m} \sum_{i \in \mathcal{I}_1} \boldsymbol{a}_i \boldsymbol{a}_i^\top \right\| \le 2\zeta\alpha_y \left\| \frac{1}{m} \boldsymbol{A}^\top \boldsymbol{A} \right\| \le 4\zeta\alpha_y, \qquad (6)
\end{aligned}$$

provided that $\|\frac{1}{m}\boldsymbol{A}^\top \boldsymbol{A}\| \le 2$ (which occurs with probability $1 - \exp(-\Omega(m))$).

Case 2. When $i \in \mathcal{I}_2$, both indicator variables are zero since

$$|\boldsymbol{a}_i^\top \boldsymbol{x}| \ge |\boldsymbol{a}_i^\top \boldsymbol{x}_0| - \zeta > \alpha_y,$$

leading to

$$\frac{1}{m} \sum_{i \in \mathcal{I}_2} \boldsymbol{a}_i \boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y\right\}} - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y\right\}} \right\} = \mathbf{0}.$$

Case 3 and Case 4. The index sets $\mathcal{I}_3$ and $\mathcal{I}_4$ represent the region where $\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y\right\}}$ and $\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y\right\}}$ might disagree. For these two cases, one can only bound

$$\left\| \frac{1}{m} \sum_{i \in \mathcal{I}_3 \cup \mathcal{I}_4} \boldsymbol{a}_i \boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top \boldsymbol{x})^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y\right\}} - (\boldsymbol{a}_i^\top \boldsymbol{x}_0)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y\right\}} \right\} \right\| \le 2\alpha_y^2 \left\| \frac{1}{m} \sum_{i \in \mathcal{I}_3 \cup \mathcal{I}_4} \boldsymbol{a}_i \boldsymbol{a}_i^\top \right\| \qquad (7)$$

using the truncation rule $|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \alpha_y$ and $|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \le \alpha_y$. Fortunately, $\mathcal{I}_3$ and $\mathcal{I}_4$ correspond to a collection of rare events. In fact, for any small constant $\zeta > 0$, the standard concentration inequality together with the union bound give

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \in [\alpha_y \pm \zeta]\right\}} \le (1+\epsilon)\, \mathbb{E}\left[\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \in [\alpha_y \pm \zeta]\right\}} \right], \quad \forall \boldsymbol{x}_0 \in \mathcal{N}_\epsilon$$

with probability at least $1 - \left(1 + \frac{2}{\epsilon}\right)^n \exp(-\Omega(m))$. When ζ is sufficiently small, this suggests

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}_0| \in [\alpha_y \pm \zeta]\right\}} \le \frac{\vartheta}{2}, \quad \forall \boldsymbol{x}_0 \in \mathcal{N}_\epsilon$$

for some small constant $\vartheta > 0$. Additionally, it follows from Lemma 6 in the main text that with high probability,

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top (\boldsymbol{x} - \boldsymbol{x}_0)| \ge \zeta\right\}} \le \frac{\vartheta}{2}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n,$$

provided that the separation ϵ of the ϵ-net is small enough. In summary, the cardinality of the index sets $\mathcal{I}_3$ and $\mathcal{I}_4$ combined together cannot exceed ϑm.

It then comes down to controlling the spectral norm of $\boldsymbol{A}_S$ for all index sets S obeying $|S| \leq \vartheta m$, where $\boldsymbol{A}_S$ denotes the submatrix of $\boldsymbol{A}$ comprising $\{\boldsymbol{a}_i\}_{i\in S}$. Standard random matrix theory (e.g. [1, Corollary 5.35]) suggests that for any given S obeying $|S| \leq \vartheta m$,

$$\|\boldsymbol{A}_S\| \leq \sqrt{\vartheta m} + \sqrt{n} + \tau$$

holds with probability exceeding $1 - 2\exp\left(-\frac{\tau^2}{2}\right)$. Note that the total number of index sets S with $|S| \leq \vartheta m$ is bounded above by [2, Example 11.1.3]

$$\binom{m}{\vartheta m} \leq e^{m\mathcal{H}(\vartheta)},$$

with $\mathcal{H}(\vartheta) := -\vartheta \log \vartheta - (1-\vartheta)\log(1-\vartheta)$ denoting the binary entropy function. Setting $\tau = 2\sqrt{\mathcal{H}(\vartheta)\, m}$ we get

$$\frac{1}{\sqrt{m}} \|\boldsymbol{A}_S\| \leq \sqrt{\vartheta} + \sqrt{\frac{n}{m}} + 2\sqrt{\mathcal{H}(\vartheta)}, \quad \forall S : |S| \leq \vartheta m.$$

with probability at least $1 - \exp(-\Omega(m))$. Thus, for any constant $\delta > 0$, one has

$$\frac{1}{\sqrt{m}} \|\boldsymbol{A}_S\| \leq \tilde{\delta}, \quad \forall S : |S| \leq \vartheta m$$

as long as m/n is sufficiently large and ϑ is sufficiently small. This combined with (7) yields

$$\left\| \frac{1}{m} \sum_{i\in\mathcal{I}_3\cup\mathcal{I}_4} \boldsymbol{a}_i\boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top\boldsymbol{x})^2 \mathbf{1}_{\{|\boldsymbol{a}_i^\top\boldsymbol{x}|\leq\alpha_y\}} - (\boldsymbol{a}_i^\top\boldsymbol{x}_0)^2 \mathbf{1}_{\{|\boldsymbol{a}_i^\top\boldsymbol{x}_0|\leq\alpha_y\}} \right\} \right\| \leq \frac{2\alpha_y^2}{m} \|\boldsymbol{A}_{\mathcal{I}_3\cup\mathcal{I}_4}\|^2 \leq 2\alpha_y^2\tilde{\delta}^2. \tag{8}$$

To finish up, we put the above cases together to deduce

$$\left\| \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{a}_i\boldsymbol{a}_i^\top \left\{ (\boldsymbol{a}_i^\top\boldsymbol{x})^2 \mathbf{1}_{\{|\boldsymbol{a}_i^\top\boldsymbol{x}|\leq\alpha_y\}} - (\boldsymbol{a}_i^\top\boldsymbol{x}_0)^2 \mathbf{1}_{\{|\boldsymbol{a}_i^\top\boldsymbol{x}_0|\leq\alpha_y\}} \right\} \right\| \leq 4\zeta\alpha_y + 2\alpha_y^2\tilde{\delta}^2, \tag{9}$$

and hence

$$\begin{aligned} \left\|\boldsymbol{Y}_{\boldsymbol{x}} - \beta_1\boldsymbol{x}\boldsymbol{x}^\top - \beta_2\boldsymbol{I}\right\| &\leq \|\boldsymbol{Y}_{\boldsymbol{x}} - \boldsymbol{Y}_{\boldsymbol{x}_0}\| + \left\|\boldsymbol{Y}_{\boldsymbol{x}_0} - \beta_1\boldsymbol{x}_0\boldsymbol{x}_0^\top - \beta_2\boldsymbol{I}\right\| + \beta_1\left\|\boldsymbol{x}_0\boldsymbol{x}_0^\top - \boldsymbol{x}\boldsymbol{x}^\top\right\| & (10) \\ &\leq 4\zeta\alpha_y + 2\alpha_y^2\tilde{\delta}^2 + \frac{\delta}{2} + 2.5\beta_1\epsilon, & (11) \end{aligned}$$

where the last inequality makes use of Lemma 2 of the main text, i.e. $\left\|\boldsymbol{x}_0\boldsymbol{x}_0^\top - \boldsymbol{x}\boldsymbol{x}^\top\right\| \leq 2.5\|\boldsymbol{x} - \boldsymbol{x}_0\|\|\boldsymbol{x}\| \leq 2.5\epsilon$. Since $\tilde{\delta}$, ζ, ϵ can all be arbitrarily small, this establishes (1) and in turn the universal guarantees.

References

[1] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing, Theory and Applications*, pages 210 – 268, 2012.

[2] T. M. Cover and J. A. Thomas. *Elements of information theory.* John Wiley & Sons, 2012.