New Restricted Isometry Results for Noisy Low-rank Recovery

Karthik Mohan and Maryam Fazel

***Abstract*—The problem of recovering a low-rank matrix consistent with noisy linear measurements is a fundamental problem with applications in machine learning, statistics, and control. Reweighted trace minimization, which extends and improves upon the popular nuclear norm heuristic, has been used as an iterative heuristic for this problem. In this paper, we present theoretical guarantees for the reweighted trace heuristic. We quantify its improvement over nuclear norm minimization by proving tighter bounds on the recovery error for low-rank matrices with noisy measurements. Our analysis is based on the Restricted Isometry Property (RIP) and extends some recent results from Compressed Sensing. As a second contribution, we improve the existing RIP recovery results for the nuclear norm heuristic, and show that recovery happens under a weaker assumption on the RIP constants.**

I. Introduction

The noisy affine rank minimization problem aims to find the lowest rank matrix consistent with noisy linear measurements,

$$\begin{array}{ll}\text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \|\mathcal{A}(X) - b\|_2 \leq \epsilon,\end{array} \tag{1}$$

where $X \in \mathbb{R}^{m\times n}$ is the variable, $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map, $b = \mathcal{A}(X_0) + e$ denotes the noisy measurements with $\|e\|_2 \leq \epsilon$. $X_0 \in \mathbb{R}^{m\times n}$ is the matrix we aim to recover. We denote its singular value decomposition by $X_0 = U\Sigma V^T = X_{0,r} + (X_0 - X_{0,r})$, where $X_{0,r}$ is formed by truncating the SVD after r terms. This problem (and its variations) have many applications including collaborative filtering, quantum tomography, system identification, and Euclidean embedding (see e.g. [19], [6] and references therein). When X is a diagonal matrix, problem (1) reduces to the classical problem of compressed sensing, where the goal is to recover a *sparse* vector. Many approaches for problem (1) have been proposed using this analogy, including the nuclear norm heuristic [10] (analogous to ℓ_1 minimization), the reweighted trace heuristic [11] (analogous to reweighted ℓ_1), and SVT [1], as well as alternative methods not based on norm minimization such as ADMiRA [14], and SVP [15].

The *Nuclear Norm Heuristic* (NNH) has been particularly popular and has been extensively studied from both theoretical and algorithmic perspectives. This heuristic replaces rank in the objective of (1) with the nuclear norm (also known as the Schatten 1-norm or trace norm) of the matrix, denoted by $\|X\|_* = \sum_i^{\min\{m,n\}} \sigma_i(X)$ where $\sigma_i(X)$ are the singular values. NNH solves the convex problem

$$\begin{array}{ll}\text{minimize} & \|X\|_* \\ \text{subject to} & \|\mathcal{A}(X) - b\|_2 \leq \epsilon.\end{array} \tag{2}$$

Electrical Engineering Department, University of Washington, Seattle. Email: karna@u.washington.edu, mfazel@u.washington.edu.

Research funded in part by NSF CAREER grant ECCS-0847077.

This heuristic is improved upon by the *Reweighted Trace Heuristic* (RTH) [11], discussed in section III, which uses a weighted objective with iterative weight updates. RTH can also be interpreted as locally minimizing a smooth concave function, logarithm of the determinant of the matrix, instead of its rank. Both RTH and its vector analog, reweighted ℓ_1 minimization [8], empirically show better recovery properties than NNH or ℓ_1 minimization (see [8], [12], [17]). Recently, analytical results for the reweighted ℓ_1 heuristic were given by Needell in [18]. However, no theoretical guarantees on the performance of the RTH have been available.

In this paper, we present the first theoretical guarantees for the RTH, in the case where the matrix variable is positive semidefinite and the map $\mathcal{A}$ satisfies the *Restricted Isometry Property* (RIP). Extending the approach in [18], we quantify the improvement of RTH over nuclear norm minimization by proving tighter bounds on the recovery error for low-rank matrices with noisy measurements. As another contribution, we extend recent RIP results for ℓ_1 minimization [2], [4], [3] to the NNH, and show that recovery happens under a *weaker* assumption on the RIP constants. The weaker condition is useful in our analysis of the RTH, and may also be of independent interest.

Recent results [6] show random maps satisfying RIP guarantee recovery with the least possible number of measurements, $\mathcal{O}(\max\{m,n\}r)$. These maps also yield very tight (in an oracle sense) error bounds for noisy recovery. Furthermore, even in applications where the RIP is not satisfied such as Matrix Completion, restricted versions of it (e.g.,[15]) have proven useful. These results encourage the use of RIP as an analytical tool in certain contexts such as analysis of noisy recovery. It may also lead to more applications that use random ensembles.

The paper is organized as follows. The improved RIP conditions are derived in Section II, and are used in the analysis of the RTH given in Section III. In Section IV, we put our contributions in perspective and discuss some possible extensions of this work.

II. RIP result for the Nuclear Norm Heuristic

In this section, we give an RIP-based recovery result for the Nuclear Norm Heuristic (NNH), extending the improvements from compressed sensing ([2],[4]). We begin with a few definitions.

Definition II.1. *The r-restricted isometry constant δ_r of a linear operator, $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is the smallest constant for which*

$$(1 - \delta_r)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \delta_r)\|X\|_F^2 \tag{3}$$

holds for all matrices X of rank at most r.

U.S. Government Work Not Protected by U.S. Copyright

Definition II.2. *Let $t = \min(m,n)$. If $r + r' \leq t$, the r,r'-restricted orthogonality constant $\theta_{r,r'}$ of a linear operator $\mathcal{A} : R^{m\times n} \rightarrow \mathbb{R}^p$ is the smallest constant for which*

$$|\langle \mathcal{A}(X), \mathcal{A}(X')\rangle| \leq \theta_{r,r'}\|X\|_F\|X'\|_F \quad (4)$$

holds for all matrices X of rank at most r and all matrices X' of rank at most r', where X, X' are such that if the SVD of X is $X = [U_r U_{m-r}]\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_r^T\\ V_{n-r}^T\end{bmatrix}$ then $X' = U_{m-r}Z_1 + Z_2V_{n-r}^T$ for some $Z_1 \in \mathbb{R}^{(m-r)\times n}, Z_2 \in \mathbb{R}^{m\times(n-r)}$.

We note that in the above definition of $\theta_{r,r'}$, X, X' satisfy $\langle X, X'\rangle = 0$. It can be shown that (similar to the vector case in [7])

$$\theta_{r,r'} \leq \delta_{r+r'}, \theta_{r_1,r_2} \leq \theta_{r_3,r_4} \forall r_1 \leq r_2, r_3 \leq r_4 \text{ s.t. } r_1 + r_2 \leq t, r_3 + r_4 \leq t. \quad (5)$$

Our recovery result for the nuclear norm heuristic is as follows:

Theorem II.3. *Let $\mathcal{A}$ have a $(2r+\alpha r)$-restricted isometry constant $\delta_{2r+\alpha r}$ and a $(2r+\alpha r, \beta r)$-restricted orthogonality constant $\theta_{2r+\alpha r,\beta r}$, where $2\alpha \leq \beta \leq 4\alpha$. Let X^* be the solution obtained through nuclear norm norm minimization and X_0 be as defined before. If $\delta_{2r+\alpha r} + \frac{1}{\sqrt{\beta}}\theta_{2r+\alpha r,\beta r} < 1$, then $H = X^* - X_0$ satisfies*

$$\|H\|_F \leq C\epsilon + \frac{B}{\sqrt{r}}\|X_0 - X_{0,r}\|_*,$$

where the constants B, C are a function of the isometry and orthoganality constants.

The following nuclear norm inequality proves useful in getting sharper bounds and can be shown through the fact that the nuclear norm is the dual norm of the spectral norm.

Lemma II.4. *For any block matrix, $X = \begin{bmatrix}X_1 & X_2\\ X_3 & X_4\end{bmatrix}$, $\|X\|_* \geq \|X_1\|_* + \|X_4\|_*$.*

Now, we come to the proof of the main theorem.

Theorem II.3 Proof: The proof goes along similar lines to that in [5],[4] where the idea of support splitting is used to show that a sparse vector can be recovered through (ℓ_1) minimization if the matrix A defining the constraint set satisfies RIP. Let SVD of $X_0 = \begin{bmatrix}U_r & U_{m-r}\end{bmatrix}\begin{bmatrix}\Sigma_r & 0\\ 0 & \bar{\Sigma}\end{bmatrix}\begin{bmatrix}V_r^T\\ V_{n-r}^T\end{bmatrix} = X_{0,r} + (X_0 - X_{0,r})$, with $X_{0,r} = U_r\Sigma_r V_r^T$. Define $\mathcal{P}_U = U_rU_r^T, \mathcal{P}_{U^\perp} = U_{m-r}U_{m-r}^T, \mathcal{P}_V = V_rV_r^T, \mathcal{P}_{V^\perp} = V_{n-r}V_{n-r}^T$. Let $H = \mathcal{P}_T(H) + \mathcal{P}_{T^\perp}(H)$, with

$$\begin{aligned}\mathcal{P}_T(H) &= \mathcal{P}_UH\mathcal{P}_V + \mathcal{P}_UH\mathcal{P}_{V^\perp} + \mathcal{P}_{U^\perp}H\mathcal{P}_V\\ \mathcal{P}_{T^\perp}(H) &= \mathcal{P}_{U^\perp}H\mathcal{P}_{V^\perp} \quad (6)\end{aligned}$$

We note that T denotes the set $\{Z : Z = \mathcal{P}_UH\mathcal{P}_V + \mathcal{P}_UH\mathcal{P}_{V^\perp} + \mathcal{P}_{U^\perp}H\mathcal{P}_V$ for some $H \in \mathbb{R}^{m\times n}\}$. And $T^\perp$ denotes the set $\{Z : Z = \mathcal{P}_{U^\perp}H\mathcal{P}_{V^\perp}$ for some $H \in \mathbb{R}^{m\times n}\}$, i.e. the set of all matrices whose row and column spaces are orthogonal to the row and column space of $X_{0,r}$. Also note that $\text{rank}(\mathcal{P}_T(H)) \leq 2r$ and that $X_{0,r}^TP_{T^\perp}(H) = X_{0,r}P_{T^\perp}(H)^T = 0$.

We define $U^THV = \begin{bmatrix}U_r^THV_r & U_r^THV_{n-r}\\ U_{m-r}^THV_r & U_{m-r}^THV_{n-r}\end{bmatrix} = \begin{bmatrix}\bar{H}_1 & \bar{H}_2\\ \bar{H}_3 & \bar{H}_4\end{bmatrix}$. Therefore, $\|X_0\|_* \geq \|X_0 + H\|_* = \|U^T(X_0 + H)V\|_* = \left\|\begin{bmatrix}\Sigma_r + \bar{H}_1 & \bar{H}_2\\ \bar{H}_3 & \bar{\Sigma} + \bar{H}_4\end{bmatrix}\right\|_* \geq \|\Sigma_r + \bar{H}_1\|_* + \|\bar{\Sigma} + \bar{H}_4\|_* \geq \|X_{0,r}\|_* - \|\mathcal{P}_UH\mathcal{P}_V\|_* - \|X_0 - X_{0,r}\|_* + \|\mathcal{P}_{U^\perp}H\mathcal{P}_{V^\perp}\|_*$, where the second inequality follows from Lemma II.4. Thus,

$$\|\mathcal{P}_{T^\perp}(H)\|_* \leq \|\mathcal{P}_UH\mathcal{P}_V\|_* + 2\|X_0 - X_{0,r}\|_* \quad (7)$$

Let $\mathcal{P}_{T^\perp}(H) = [U_*U_{11}U_{12}U_{21}U_{22}..]\,\Sigma_{H^\perp}[V_*V_{11}V_{12}V_{21}V_{22}..]^T$ be the SVD of $\mathcal{P}_{T^\perp}(H)$, with the singular values in $\Sigma_{H^\perp}$ decreasing from top to the bottom. $\Sigma_{H^\perp}$ is made up of diagonal blocks $\Sigma_*, \Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}, \ldots$ with Σ_* and Σ_{i1} of size αr and Σ_{i2} of size $(\beta - \alpha)r$ $(2\alpha \leq \beta \leq 4\alpha)$ $\forall i \geq 1$. Denote, $H_* = U_*\Sigma_*V_*^T$ and $H_{i1} = U_{i1}\Sigma_{i1}V_{i1}^T, H_{i1} = U_{i2}\Sigma_{i2}V_{i2}^T$, $H_i = H_{i1} + H_{i2}$ $\forall i \geq 1$.

At this point, we note that our goal is to have a good bound on $\|H\|_F^2 = \|\mathcal{P}_T(H) + H_*\|_F^2 + \sum_{i\geq 1}\|H_i\|_F^2$. Thus two key steps follow. The first is to get a good inequality between $\sum_{i\geq 1}\|H_i\|_F$ and $\|\mathcal{P}_T(H) + H_*\|_F$. We combine (7) and the shifting inequality(see e.g. [4]) towards this end. The second step is to get a bound on $\|\mathcal{P}_T(H) + H_*\|_F$ using RIP and restricted orthogonality.

We note that the shifting inequality can be used to get tighter ℓ_2, ℓ_1 inequalities between pairs of vectors. This idea is easily extended to matrices by applying the shifting inequality to the singular values as below.

$$\begin{aligned}\|H_1\|_F &= \|H_{11} + H_{12}\|_F \leq \frac{\|H_*\|_* + \|H_{11}\|_*}{\sqrt{\beta r}}\\ \|H_i\|_F &= \|H_{i1} + H_{i2}\|_F\\ &\leq \frac{\|H_{(i-1)2}\|_* + \|H_{i1}\|_*}{\sqrt{\beta r}}, \quad \forall i \geq 2 \quad (8)\end{aligned}$$

Using the inequalities in (8), (7) and Lemma II.4 it is easy to see that,

$$\begin{aligned}\sum_{i\geq 1}\|H_i\|_F &\leq \frac{1}{\sqrt{\beta r}}\left(\|H_*\|_* + \sum_{i\geq 1}\|H_i\|_*\right)\\ &\leq \sqrt{\frac{1}{\beta}}\|\mathcal{P}_T(H)\|_F + \frac{2}{\sqrt{\beta r}}\|X_0 - X_{0,r}\|_* \quad (9)\end{aligned}$$

We upper and lower bound $|\langle \mathcal{A}(H), \mathcal{A}(\mathcal{P}_T(H) + H_*)\rangle|$ to derive a bound for $\mathcal{P}_T(H) + H_*$ and thus a bound for H. Denoting, $S = |\langle \mathcal{A}(H), \mathcal{A}(\mathcal{P}_T(H) + H_*)\rangle|$ it can be shown(analogous to the derivation in Section 3.2 [4]) that,

$$\begin{aligned} S &\geq \|\mathcal{P}_T(H)+H_*\|_F((1-\delta_{2r+\alpha r} \\ &\quad -\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r})\|\mathcal{P}_T(H)+H_*\|_F - \\ &\quad \theta_{2r+\alpha r,\beta r}\frac{2}{\sqrt{\beta r}}\|X_0-X_{0,r}\|_*) \end{aligned} \tag{10}$$

and that,

$$S \leq 2\epsilon\sqrt{1+\delta_{2r+\alpha r}}\|\mathcal{P}_T(H)+H_*\|_F \tag{11}$$

Combining (10) and (11), we get,

$$\begin{aligned} \|\mathcal{P}_T(H)+H_*\|_F &\leq \frac{2\epsilon\sqrt{\delta_{2r+\alpha r}+1}}{1-\delta_{2r+\alpha r}-\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}} \\ &\quad +\frac{\theta_{2r+\alpha r,\beta r}\frac{2}{\sqrt{\beta r}}\|X_0-X_{0r}\|_*}{1-\delta_{2r+\alpha r}-\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}} \end{aligned} \tag{12}$$

Combining, (12) and (9), we get a bound on $\|H\|_F$,

$$\begin{aligned} \|H\|_F^2 &= \|\mathcal{P}_T(H)+H_*\|_F^2+\sum_{i\geq 1}\|H_i\|_F^2 \\ &\leq \|\mathcal{P}_T(H)+H_*\|_F^2+\left(\sum_{i\geq 1}\|H_i\|_F\right)^2 \\ &\leq \left(C\epsilon+\frac{B}{\sqrt{r}}\|X_0-X_{0r}\|_*\right)^2 \end{aligned} \tag{13}$$

where, $C = \frac{2\sqrt{1+\delta_{2r+\alpha r}}\sqrt{1+\frac{1}{\beta}}}{1-\delta_{2r+\alpha r}-\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}}$, $B = \frac{2}{\sqrt{\beta}}\left(\frac{\sqrt{1+\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}}{1-\delta_{2r+\alpha r}-\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}}+1\right)$. Note that for (12) to hold, we need that the denominator to be positive, i.e.

$$\delta_{2r+\alpha r}+\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}<1 \tag{14}$$

■

Let $\beta = 4\alpha$. Assume that $r = 10k+\omega$, where $0\leq\omega\leq 9$ and $k\in Z_+$. Then, $(2r+\alpha r)+(4\alpha)r < \zeta r$ $(\zeta\in Z_+)$ if $\alpha r < \frac{\zeta-2}{5}r = 2k(\zeta-2)+\omega\frac{\zeta-2}{5}$. Choose α such that,

$$\alpha r = \begin{cases} 2k(\zeta-2) \text{ if } 0\leq\omega\leq 4 \\ 2k(\zeta-2)+\zeta-2 \text{ otherwise} \end{cases} \tag{15}$$

Then, $\alpha r, \beta r = 4\alpha r$ are integers. Using (5), we get that $\delta_{2r+\alpha r}+\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r} < \delta_{\zeta r}(1+\frac{1}{\sqrt{\beta}}) < 1$ if $\delta_{\zeta r} < \frac{\sqrt{\beta}}{1+\sqrt{\beta}}$.

In particular if $\omega = 0$, we get that for $\zeta = 3,4,5$, $\delta_{3r} < 2\sqrt{5}-4 = 0.4721$, $\delta_{4r} < \frac{8-\sqrt{40}}{3} = 0.558$ and $\delta_{5r} < \frac{12-\sqrt{60}}{7} = 0.607$ are sufficient for (14) to hold and thus for (13) to hold. Even if $\omega\neq 0$, for reasonably large k, the upper bound on the RIP constants$(\delta_{3r},\delta_{4r},\delta_{5r})$ that are sufficient for recovery very quickly tend to the upper bounds above. It can also be shown by extending the analysis in [2] that $\delta_{2r} < 0.307$ is sufficient for (14) to hold and thus for (13) to hold. Also observe that with zero measurement error,$(\epsilon = 0)$ and $\mathrm{rank}(X_0)\leq r$, the nuclear norm minimization exactly recovers X_0(if at least one of the above conditions on δ_{2r}, δ_{3r}, δ_{4r}, δ_{5r} are satisfied).

Our $\delta_{2r} < 0.307$ result compares well with the recent SVP result [15], where if $\delta_{2r} < \frac{1}{3}$, the SVP algorithm guarantees recovery. The SVP algorithm, though efficient requires apriori knowledge of the rank of X_0. Our results also improves greatly on the previous result of [19] and [9], where recovery is shown using nuclear norm minimization if $\delta_{5r} < 0.21$, the result of $\delta_{3r} < \frac{\sqrt{3}}{\sqrt{3}+4} = 0.302$ in [13] and also on the RIP result of $\delta_{4r} < \sqrt{2}-1$ given in [6] for recovery using NNH. We also note Theorem 2.3 in [6] mentions that if $p > O(nr)$, then recovery can be guaranteed with high probability if the map $\mathcal{A}$ is chosen from certain random distributions. Our δ_{2r} result reduces the constant in $O(nr)$ by a factor of around 2 as compared to the δ_{4r} result given in [6].

III. RIP result for Reweighted Trace Heuristic

In this section we use the guarantee result in the previous section to give a first guarantee result for the *Reweighted Trace Heuristic*(RTH). The RTH iteratively minimizes the linearization to a concave surrogate for rank(X), the surrogate being $\log\det(X+\gamma I)$, where $\gamma > 0$. The $(k+1)^{th}$ iteration of the RTH [11] when X is restricted to be positive semidefinite is given by:

$$\begin{aligned} X^{k+1} = &\quad \mathrm{argmin}_X \,\mathrm{Tr}\,(X^k+\gamma^{k+1}I)^{-1}X \\ \text{subject to} &\quad X\geq 0, \|\mathcal{A}(X)-b\|_2\leq\epsilon \end{aligned} \tag{16}$$

where, $\gamma^{k+1} > 0$ is a constant to ensure invertibility, $\mathcal{A}: \mathbb{R}^{m\times m}\rightarrow\mathbb{R}^p$ and $b = \mathcal{A}(X_0)+e$, $\|e\|_2\leq\epsilon$. Interestingly, our analysis shows that γ^k plays an important role in bounding the error of recovery. We make an additional assumption that X_0 be of rank at most r with SVD, $X_0 = U\Sigma V^T = [U_r U_{m-r}]\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_r^T\\ V_{m-r}^T\end{bmatrix}$.

Let $W^k = X^k+\gamma^{k+1}I$ and let, $X^{k+1} = X_0+H^{k+1}$. Let the smallest non-zero singular value of X_0 be μ. Also, we assume that $\|X^k-X_0\|_F\leq M^k$.

We then have the following theorem that gives conditions for the reweighted trace heuristic to have a better recovery error bound than nuclear norm minimization.

Theorem III.1. *Let $\mathcal{A}$ have the RIP constant δ_{3r}, obeying $\delta_{3r} < 2\sqrt{5}-4$. Let X^1 be the solution obtained through nuclear norm minimization (2).Then $\|X^k-X_0\|_F\leq E(k) = 2\epsilon\sqrt{1+\delta_{3r}}\frac{\sqrt{1+\frac{10}{8}C_{1,k}^2}}{1-\delta_{3r}(1+C_{1,k}\sqrt{\frac{10}{8}})}$ $\forall k\geq 2$, where $C_{1,k}$ is a constant that depends on $E(k-1),\gamma^k,\mu$. If $C_{1,k} < 1$ $\forall k\geq 2$, then the sequence, $\{E(k)\}$ converges to a limit $E < E(1)$. In particular, if $\mu > 3E(1)$ and $\gamma^k = \frac{E(k-1)(\mu+E(k-1))}{\mu-3E(k-1)}$ $\forall k\geq 2$ then the sequence, $\{E(k)\}$ converges to a limit $E < E(1)$.*

Before we proceed with the proof, we list some useful inequalities for eigenvalues and singular values.

Lemma III.2. *[16] Let $A,B\in\mathbb{R}^{n\times n}$ be hermitian matrices. Let $\lambda_1(C)\geq\lambda_2(C)\ldots\lambda_{n-1}(C)\geq\lambda_n(C)$ denote the ordered*

eigen values of any matrix C. Let the singular values be ordered similarly and be denoted by $\sigma_i(C)$. Then,

$$\begin{aligned}\lambda_i(A)+\lambda_n(B)\le\lambda_i(A+B)\le\lambda_i(A)+\lambda_1(B)\\ \sigma_i(A+B)\le\sigma_i(A)+\sigma_1(B)\\ \sigma_i(A)\sigma_n(B)\le\sigma_i(AB)\le\sigma_i(A)\sigma_1(B)\end{aligned}\quad(17)$$

If both A,B are positive semidefinite, then the following inequality also holds.

$$\lambda_i(A)\lambda_n(B)\le\lambda_i(AB)\le\lambda_i(A)\lambda_1(B)\quad(18)$$

Theorem III.1 proof: The proof is inspired by the analysis in [18] and is shown in two parts. In the first part, we assume a bound on the error, $\|H^k\|_F\le M^k$ from the previous iteration and derive a bound on the error $\|H^{k+1}\|_F$ for the next iteration of the RTH. In the second part of the proof, we use the recursive error bound expression derived in first part to show that the error bounds converge to a limiting error bound of RTH under the assumptions of the theorem.

Proof Part 1 To simplify the notation, we drop the superscripts on $\gamma^{k+1},H^{k+1},M^k,X^k,W^k$ and refer to them as γ,H,M,X,W respectively in this part of the proof. It follows from (16) that $\operatorname{Tr}W^{-1}X_0\ge\operatorname{Tr}W^{-1}(X_0+H)=\operatorname{Tr}W^{-1}X_0+\operatorname{Tr}(W^k)^{-1}H$. Thus,

$$\operatorname{Tr}W^{-1}H\le 0\quad(19)$$

We can decompose the matrix U^TX^kU as $U^TX^kU=\begin{bmatrix}\bar X_1&\bar X_2\\ \bar X_3&\bar X_4\end{bmatrix}$ and U^TWU as $\bar W=\begin{bmatrix}\bar W_1&\bar W_2\\ \bar W_3&\bar W_4\end{bmatrix}=U^TWU=\begin{bmatrix}\bar X_1+\gamma I&\bar X_2\\ \bar X_3&\bar X_4+\gamma I\end{bmatrix}$. Note that $U^TW^{-1}U=(\bar W)^{-1}$ is given by $\bar W^{-1}=\begin{bmatrix}\bar W_1^{-1}&\bar W_2^{-1}\\ \bar W_3^{-1}&\bar W_4^{-1}\end{bmatrix}$ where,

$$\begin{aligned}\bar W_1^{-1}&=(\bar X_1+\gamma I-\bar X_2(\bar X_4+\gamma I)^{-1}\bar X_3)^{-1}\\ \bar W_4^{-1}&=(\bar X_4+\gamma I-\bar X_3(\bar X_1+\gamma I)^{-1}\bar X_2)^{-1}\\ \bar W_2^{-1}&=-(\bar X_1+\gamma I)^{-1}\bar X_2\times\bar W_4^{-1}\\ \bar W_3^{-1}&=-(\bar X_4+\gamma I)^{-1}\bar X_3\times\bar W_1^{-1}\end{aligned}\quad(20)$$

are obtained through the formula for block matrix inversion. Multiplying on left side of U^TX^kU by U and right hand side by U^T, we get four terms which sum to X^k. This results in an additive decomposition: $X=X_1+X_2+X_3+X_4$. A similar additive decomposition for W^{-1},H gives $W^{-1}=W_1^{-1}+W_2^{-1}+W_3^{-1}+W_4^{-1},H=H_1+H_2+H_3+H_4$.

(19) can be decomposed as:

$$\operatorname{Tr}W^{-1}H_4\le-\operatorname{Tr}W^{-1}(H_1+H_2+H_3)\quad(21)$$

It holds that $\operatorname{Tr}W^{-1}H_4=\operatorname{Tr}W_4^{-1}H_4$, since the other terms in the additive decomposition of $(W^k)^{-1}$ cancel out. Let $\bar H=U^THU=\begin{bmatrix}\bar H_1&\bar H_2\\ \bar H_3&\bar H_4\end{bmatrix}$.

Note that both, W_4^{-1} and H_4 are positive semidefinite ($X_0+H=X^{k+1}\succeq 0\implies\bar H_4\succeq 0\implies H_4\succeq 0$). Also, $\operatorname{Tr}W_4^{-1}H_4=\operatorname{Tr}U^TW_4^{-1}UU^TH_4U=\operatorname{Tr}\bar W_4^{-1}\bar H_4\ge\sigma_{min}(\bar W_4^{-1})\operatorname{Tr}(\bar H_4)=\sigma_{min}(\bar W_4^{-1})\operatorname{Tr}(H_4)$, where the last inequality follows from (17). $\sigma_{min}(\bar W_4^{-1})\ge\frac{1}{M+\gamma}$ and thus,

$$\operatorname{Tr}W_4^{-1}H_4\ge\frac{1}{M+\gamma}\operatorname{Tr}(H_4)=\frac{1}{M+\gamma}\|H_4\|_*\quad(22)$$

We note at this point that $\mathcal{P}_T(H)=H_1+H_2+H_3$ and $\mathcal{P}_{T^\perp}(H)=H_4$ (where $\mathcal{P}_T$,$\mathcal{P}_{T^\perp}$ are as defined in (6)). We now upper bound $-\operatorname{Tr}W^{-1}\mathcal{P}_T(H)$ in terms of $\|\mathcal{P}_T(H)\|_*$. Note that, $-\operatorname{Tr}W^{-1}\mathcal{P}_T(H)=-\operatorname{Tr}(W_1{}^{-1}+W_2{}^{-1}+W_3{}^{-1})\mathcal{P}_T(H)$. We can bound the above quantity by observing that $|\operatorname{Tr}A'B|\le\|A\|_2\|B\|_*$ for any two A,B. Thus,

$$\begin{aligned}|\operatorname{Tr}W^{-1}\mathcal{P}_T(H)|&\le\|(W_1^{-1}+W_2^{-1}+W_3^{-1})\|_2\|\mathcal{P}_T(H)\|_*\\ &=\|U^T(W_1^{-1}+W_2^{-1}+W_3^{-1})U\|_2\|\mathcal{P}_T(H)\|_*\\ &\le(\|\bar W_1^{-1}\|_2+\|\bar W_3^{-1}\|_2)\|\mathcal{P}_T(H)\|_*\end{aligned}$$

where the last inequality uses the fact that $\|\bar W_3^{-1}\|_2=\max(\|\bar W_3^{-1}\|_2,\|\bar W_2^{-1}\|_2)$. We now upper bound each of $\|\bar W_1^{-1}\|_2$, $\|\bar W_3^{-1}\|_2$.Define $G(M,\gamma)=\frac{M+\gamma}{\gamma(\mu-M)+\gamma^2-M^2}$. Then,

$$\|\bar W_3^{-1}\|_2\le\|\bar W_1^{-1}\|_2\frac{M}{\gamma}\le\frac{G(M,\gamma)M}{M+\gamma}.$$

Thus,$\|\bar W_1^{-1}\|_2+\|\bar W_3^{-1}\|_2\le G(M,\gamma)$. The above inequalities can be checked using the definitions in (20) and the inequalities in (17),(18). One key step is bounding $\sigma_{min}(\bar X_1+\gamma I)=\lambda_{min}(\bar X_1+\gamma I)=\lambda_{min}(\bar X_1)+\gamma$. Since, $\|X-X_0\|_2\le M$, we have that $\|\bar X_1-\Sigma\|_2\le M$. Since, $\bar X_1-\Sigma$ is symmetric, $\|\bar X_1-\Sigma\|_2\ge|\lambda_i(\bar X_1-\Sigma)|$ and thus $\lambda_{min}(\bar X_1-\Sigma)\ge-M$. From (17), we have that $\lambda_{min}(\bar X_1-\Sigma)\le\lambda_{min}(\bar X_1)-\lambda_{min}(\Sigma)$. Hence, $\lambda_{min}(\bar X_1+\gamma I)\ge\mu-M+\gamma$. The inequalities now follow by using the fact that $\bar W^{-1}\succ 0$ and through successive applications of the inequalities in (17). Therefore,

$$-\operatorname{Tr}W^{-1}\mathcal{P}_T(H)\le G(M,\gamma)\|\mathcal{P}_T(H)\|_*\quad(23)$$

Combining (21),(22), and (23), we have that,

$$\|\mathcal{P}_{T^\perp}(H)\|_*\le(M+\gamma)G(M,\gamma)\|\mathcal{P}_T(H)\|_*\quad(24)$$

Thus we have bounded the $\|\mathcal{P}_{T^\perp}(H)\|_*$ in terms of $\|\mathcal{P}_T(H)\|_*$. We can now proceed using a similar analysis as in section II to obtain a bound for $\|H\|_F$. We get that,

$$\|H\|_F\le 2\epsilon\frac{\sqrt{1+\frac{1}{\beta}C_1(M,\gamma)^2}\sqrt{1+\delta_{2r+\alpha r}}}{1-\delta_{2r+\alpha r}-C_1(M,\gamma)\sqrt{\frac{1}{\beta}}\theta_{2r+\alpha r,\beta r}}\quad(25)$$

where, $C_1(M,\gamma)=(M+\gamma)G(M,\gamma)$ and α,β are as defined earlier.

Proof Part 2 Note that, the recovery error using nuclear norm minimization (13) can be obtained by setting $C_1(M,\gamma)=1$ in (25). To simplify our analysis, we let $\beta=4\alpha$ with α chosen as in (15). Then,$\delta_{2r+\alpha r}<\delta_{3r},\theta_{2r+\alpha r,\beta r}<\delta_{3r}$. Therefore, a weaker upper bound can be obtained from (25) as

$$\|H\|_F\le D(M,\gamma)\epsilon,\quad(26)$$

where, $D(M,\gamma) = 2\epsilon\sqrt{1+\delta_{3r}}\frac{\sqrt{1+\frac{1}{\beta}C_1(M,\gamma)^2}}{1-\delta_{3r}(1+C_1(M,\gamma)\sqrt{\frac{1}{\beta}})}$. Let $E(k+1) = D(E(k),\gamma^{k+1}) \quad \forall k \geq 1$. Also denote $C_{1,k+1} = C_1(E(k),\gamma^{k+1}) \quad \forall k \geq 1$. Thus,$E(k+1)$ denotes an upper bound on the error at the end of iteration $k+1$. Since the weight is chosen to be identity in the first iteration,we have that $E(1) = D_1\epsilon$, where $D_1 = \frac{2\sqrt{1+\delta_{3r}}\sqrt{1+\frac{1}{\beta}}}{1-\delta_{3r}(1+\sqrt{\frac{1}{\beta}})}$.

Note that $E(2) \leq E(1)$ iff $C_{1,2} \leq 1$. This gives us a bound on μ (the minimum non-zero singular value in X_0):

$$\mu \geq 3E(1) + \frac{2}{\gamma}E(1)^2 \tag{27}$$

Assuming that $E(k) \leq E(k-1)$ for all previous k,it holds true that $C_{1,k+1} \leq C_{1,k}$ and thus $E(k+1) \leq E(k)$ for all k. Assuming $\delta_{3r} < 2\sqrt{5}-4$ (see results from section II), $E(k)$ is always positive. Thus,$E(k)$ converges to a limit E. The limit can be obtained by solving the equation, $E = D(E,\gamma)$. Given a μ, the optimal, $\gamma^2(\mu) = \frac{E(1)(\mu+E(1))}{(\mu-3E(1))}$ minimizes $E(2)$. We also require that $\mu \geq 3E(1)+\frac{2}{\gamma^2}E(1)^2$, i.e. $\gamma^2 > \frac{2E(1)^2}{\mu-3E(1)}$ for $E(2) < E(1)$. This is ensured if $\mu \geq 3E(1)$ (since $\gamma^2 > 0$). Hence, if γ^k is chosen so that $E(k)$ is minimized, i.e. $\gamma^k = \frac{E(k-1)(\mu+E(k-1))}{(\mu-3E(k-1))}$, then the limiting error bound, $E < E(1)$. ■

IV. Discussion and Future Work

We gave an RIP-basesd deterministic recovery result for nuclear norm minimization with RIP constants that are better than in existing results([9],[13],[19]). We then used this result to give a guarantee of recovery for RTH. To understand how RTH compares with the NNH, we vary δ_{3r}, ν(where we let $\mu \geq \nu E(1)$) and using the recursive expressions for $E(k)$, we compute how $E(k)/E(1)$ varies for $k = 2,3,4,5$. We choose γ^k optimally at each iteration as defined in the previous section. The results in Table I show that $E(k)/E(1)$ decreases and converges as k increases which is consistent with the statements in Theorem III.1. A surprising phenomenon is that as δ_{3r} increases from 0.2 to 0.45, $E(k)/E(1)$ reduces drastically. This can be explained by the fact that as δ_{3r} increases and approaches 0.472, $E(1)$ becomes very large but since $C_{1,k} < 1$, $E(k)$ doesn't grow as large and hence the small ratio. The second to last column shows the rapid growth of $E(1)$ with increase in δ_{3k}. The last column shows that if $\mu > 3E(1)$ and if γ^k is chosen optimally at each iteration, then the upper bound on the error, $E(k)$ is consistently small for large k.

We also observe that if γ^k is fixed,(e.g. $= 10E(1)$), then $E(5)/\epsilon$ is much larger than if γ^k were chosen greedily at each iteration. So it is natural to ask if choosing γ^k adaptively at each iteration as a function of the error bound in the previous iteration would lead to improved numerical results. We have given results for RTH when the constraint set is restricted to be positive semidefinite. Further work could include extending this result when the constraint set is convex but not restricted to be positve semidefinite.

δ_{3r}	ν	$\frac{E(2)}{E(1)}$	$\frac{E(3)}{E(1)}$	$\frac{E(4)}{E(1)}$	$\frac{E(5)}{E(1)}$	$\frac{E(1)}{\epsilon}$	$\frac{E(5)}{\epsilon}$
0.2	5	0.83	0.75	0.712	0.694	5.702	3.955
	10	0.603	0.537	0.528	0.527	5.702	3.006
0.3	5	0.755	0.610	0.532	0.496	9.382	4.6543
	10	0.482	0.398	0.383	0.382	9.382	3.58
0.4	5	0.564	0.299	0.217	0.2	23.233	4.645
	10	0.275	0.187	0.181	0.18	23.233	4.187
0.45	5	0.284	0.075	0.06	0.06	77.05	4.59
	10	0.103	0.059	0.058	0.058	77.05	4.476

TABLE I
COMPARING UPPER BOUNDS ON RECOVERY ERROR AT DIFFERENT ITERATIONS OF REWEIGHTED NUCLEAR NORM MINIMIZATION WITH THE RECOVERY ERROR OF NUCLEAR NORM MINIMIZATION, $E(1)$

References

[1] J.F. Cai, E.J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. 2009. Available at http://arxiv.org/abs/0810.3286 , Submitted on 18 Oct 2008.

[2] T.T. Cai, G. Xu, and J. Zhang. New bounds for restricted isometry constants. 2009. Technical Report, Available Online.

[3] T.T. Cai, G. Xu, and J. Zhang. On recovery of sparse signals via ℓ_1 minimization. In *Proceedings of the 2009 IEEE Transactions on Information Theory*, pages 3388–3397, 2009.

[4] T.T. Cai, G. Xu, and J. Zhang. Shifting inequality and recovery of sparse signals. 2009. To appear in *IEEE Transactions on Signal Processing*.

[5] E.J. Candes. The restricted isometry property and its implications for compressed sensing. *Academie des Sciences*, 2008.

[6] E.J. Candes and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. 2009. Available at http://arxiv.org/abs/1001.0339.

[7] E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 2004.

[8] E.J. Candes, M.B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.

[9] M. Fazel, E.J. Candes, B. Recht, and P. Parrilo. Compressed sensing and robust recovery of low rank matrices. In *Proc. Asilomar Conference*, 2009.

[10] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proc. American Control Conference*, 2001.

[11] M. Fazel, H. Hindi, and S.Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proc. American Control Conference*, 2003.

[12] S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q minimization for $0 < q < 1$, 2009.

[13] K. Lee and Y. Bresler. Guaranteed minimum rank approximation from linear observations by nuclear norm minimization with an ellipsoidal constraint. Available online at http://arxiv.org/abs/0903.4742 Submitted on 27 Mar 2009.

[14] K. Lee and Y. Bresler. Admira:atomic decomposition for minimum rank approximation. 2009. Available at http://arxiv.org/abs/0905.0044.

[15] R. Meka, P. Jain, and I.S. Dhillon. Guaranteed rank minimization via singular value projection. Available at http://arxiv.org/abs/0909.5457 Submitted on 30 Sep, 2009.

[16] J.K. Merikoski and R. Kumar. Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes*, 4:150 – 159, 2004.

[17] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In Proc. ACC 2010.

[18] D. Needell. Noisy signal recovery via iterative reweighted l1-minimization. In *Proc. Asilomar conference on Signals, Systems and Computers*, 2009. Available at http://arxiv.org/abs/0904.3780.

[19] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. Accepted for publication, *SIAM Review*., 2007.