**SUPPLEMENT MATERIAL**

**Chinmay Talegaonkar**, **Parthasarathi Khirwadkar** and **Ajit Rajwade**

1,2 Department of Electrical Engineering, IIT Bombay  
3 Department of Computer Science and Engineering, IIT Bombay

1. PERFORMANCE BOUNDS FOR Q1 AND Q2

**Theorem 1**: - Bounds for Q1  
Abiding by the notations in [1], let $B(\cdot)$ be an $(\delta, k)$ RIP obeying linear operator [1], i.e. $\forall X \neq 0$ and $\|X\|_0 \leq k$

$$\left| \frac{\|B(X)\|_2^2}{\|X\|_2^2} - 1 \right| < \delta$$

(1)

Let $X^*$ be the true rank 1 matrix that satisfies the constraints in Q1, and $\hat{X}$ be the solution to Q1 for an appropriate choice of parameters. We define $\Delta = \hat{X} - X^*$. For a matrix $X \in \mathbb{R}^{d \times d}$, $X_T$ denotes a matrix with all values zero except the indices in $T$, which are set to the corresponding values of $X$. $\forall \lambda \geq \frac{2d}{1-\rho} + \frac{\delta}{k^2}$ and $\delta \leq \sqrt{2} - 1$ we get

$$\|\Delta\|_2 \leq \frac{2\alpha(1-\rho)^{-1} + 2(1+\rho)(1-\rho)^{-1}k^{-\frac{1}{2}}\|\hat{X} - X^*_k\|_1}{1 - \frac{2d}{1-\rho} + \frac{\delta}{k^2} \frac{1}{\lambda}}$$

(2)

where $\rho = \frac{\sqrt{\delta}}{1-\delta} \leq 1$, $X^*_k$ is the matrix with the $k$ largest elements in $X^*$ at the corresponding indices and the others elements to be 0, $\varepsilon \sim O(\sqrt{n})$ and $\alpha = \frac{4\sqrt{(T_0+1)(1+\delta)}}{1-\delta}$. $\hat{X}$ is the true intensity of the underlying measurement, i.e. $\|B(X^*)\|_1$, which is assumed to be known for naturally acquired measurements.

**Theorem 2**: - Bounds for Q2  
Given that the assumptions in section 2.2 hold, and $\Psi$ has sufficiently small RIP constant $\delta$, then, there exist absolute constants $C_1$, $C_2$, and $C_3$ such that if $m \geq C_1 k \log \frac{\lambda}{\varepsilon}$, and $n \geq C_1 m$ then any estimate $\tilde{X}$ of the Algorithm obeys

$$\|\tilde{X} - X^*\|_F \leq C_2,$$

for all rank-one and $k \times k$-sparse matrices $X^* \neq 0$ with probability exceeding $1 - e^{-C_3 mn}$. The constant $C_2$ depends on $\sqrt{\lambda}$ and since $\varepsilon \sim O(\sqrt{n})$, we don’t get a dependence on $\frac{\lambda}{\sqrt{n}}$ like that in [2]. The proofs of these theorems are inspired by the performance bounds for compressive phase retrieval presented in [3] by Ohlsson et. al for the standard CPRL problem (P1), and in [2] by Bahamani et. al for the 2-stage sparse recovery method (P2). We adapted these proofs for Poisson corrupted signals using the techniques in [4] and [5].

2. PROVING PERFORMANCE BOUNDS

We briefly present here the proof outline for performance bounds of Q1 and Q2.

2.1. Proof of Q1

As defined in theorem 1, let $B(\cdot)$ be an $(\delta, k)$ RIP obeying linear operator [1], i.e. $\forall X \neq 0$ and $\|X\|_0 \leq k$

$$\left| \frac{\|B(X)\|_2^2}{\|X\|_2^2} - 1 \right| < \delta$$

(3)

Let $X^*$ be the true rank 1 matrix that satisfies the constraints in Q1, and $\hat{X}$ be the solution to Q1 for an appropriate choice of parameters. We define $\Delta = \hat{X} - X^*$. For a matrix $X \in \mathbb{R}^{d \times d}$, $X_T$ denotes a matrix with all values zero except the indices in $T$, which are set to the corresponding values of $X$. $T_0$ is the index set of the $k$ largest elements of $X^*$ in absolute value, and $T_0^c$ is the corresponding complement set. Let $T_1$ be the index set associated with the $k$ largest elements in absolute value of $\Delta_{T_0}$ and $T_{0,1} \triangleq T_0 \cup T_1$ be the union. Let $T_2$ be the index set associated with the $k$ largest elements in absolute value of $\Delta_{T_{0,1}}$, and so on. Using the fact that:

$$\|\Delta\|_2 = \|\Delta_{T_{0,1}} + \Delta_{T_{0,1}^c}\|_2 \leq \|\Delta_{T_{0,1}}\|_2 + \|\Delta_{T_{0,1}^c}\|_2.$$  

(4)

We independently derive bounds for both the terms on the RHS, and then combine the result to get a bound on $\|\Delta\|_2$.

Bounds for $\|\Delta_{T_{0,1}}\|_2$ can be derived exactly as in the proof of theorem 8 in [1]. For bounding $\|\Delta_{T_{0,1}}\|_2$, we first use a technique from [5] to upper bound $\|B(\Delta)\|_2$, which handles Poisson corruption for signals in Q1 and Q2. Some algebraic manipulation gives

$$\|B(\Delta)\|_2^2 = \|B(\hat{X} - X^*)\|_2^2 = \Sigma_{i=1}^{N}(\sqrt{(B(\hat{X}))_i + c - \sqrt{(B(X^*))_i + c}})^2$$

$$= \Sigma_{i=1}^{N}(\sqrt{(B(\hat{X}))_i + c + \sqrt{(B(X^*))_i + c}}^2).$$

1. Since $\|\sqrt{y + c} - \sqrt{B(X^*)} + c\|_2$ is upper bounded by $\varepsilon$, using triangular inequality we get $\|\sqrt{B(\hat{X})} + c - \sqrt{B(X^*)} + c\|_2 \leq \|\sqrt{y} + c - \sqrt{B(\hat{X})} + c\|_2 + \|\sqrt{y} + c - \sqrt{B(X^*)} + c\|_2 \leq 2\varepsilon$. 

chinmay0301@iitb.ac.in\textsuperscript{1}, parthasarathi.k@iitb.ac.in\textsuperscript{2}  
ajit0714@cse.iitb.ac.in\textsuperscript{3}, AR thanks IIT-B Seed Grant 14IRCCSG012
2. For scalars $v_1 \geq 0, v_2 \geq 0$, we have $(\sqrt{v_1} + \sqrt{v_2})^2 \leq 4\max(v_1, v_2)$. Likewise we also have $(B\hat{X})_i \leq 1$, where $I$ is the true intensity of the underlying measurement, i.e. $\|B(X^*)\|_1$, which is assumed to be known for naturally acquired measurements. Hence we get $(\sqrt{(B\hat{X})_i} + c + \sqrt{(B(X^*))_i} + c)^2 \leq 4(I + c)$.

3. Using the results of all the steps above, we have $\|B(\Delta)\|_2 \leq 4\varepsilon \sqrt{T + c}$.

By a derivation similar to that in [4], one can show that

$$\|\Delta r_{0,1}\|_2 \leq \frac{\alpha \varepsilon}{1 - \rho} + \frac{2\rho}{1 - \rho} \frac{k^{-1/2}}{\lambda} \|X - X^*\|_1 - \frac{1}{\lambda} \frac{1}{1 - \rho} \text{Tr } \Delta,$$

(5)

where $\rho = \sqrt{\frac{2d}{1 - \rho}} \leq 1$, $X^*_k$ is the matrix with the $k$ largest elements in $X^*$ at the corresponding indices and the others elements to be 0, and $\alpha = \frac{1}{\lambda} \frac{1}{1 - \rho}$. The $\text{Tr } \Delta$ term makes the bound a bit unpractical, and can be eliminated using the inequality $\text{Tr } \Delta \leq d \|\Delta\|_2$. This inequality can obtained by upper bounding trace by $l_1$ norm and then using the Cauchy Schwartz inequality subsequently as done in [1]. We observe that this inequality differs slightly from that in the proof of the corollary 9 in [1] since $\Delta$ is not $k$-sparse. Using this inequality, and the bounds obtained for $\|\Delta r_{0,1}\|_2$ and $\|\Delta r_{0,1}\|_2$,

$$\forall \lambda \geq \frac{2d}{1 - \rho} + \frac{d}{k^2} \text{ and } \delta \leq \sqrt{2} - 1 \text{ we get}$$

$$\|\Delta\|_2 \leq \frac{2\alpha \varepsilon (1 - \rho)^{-1} + 2(1 + \rho)(1 - \rho)^{-1} k^{-\frac{1}{2}}}{1 - \left(\frac{2d}{1 - \rho} + \frac{d}{k^2}\right)} \|X - X^*\|_1.$$

(6)

### 2.2. Proof for Q2

First we find an upper bound on $\|\hat{A} - A^*\|_2$. $\hat{A}$ is the solution to Q2 and $A^*$ is the true low rank matrix. The performance bounds for estimation of $A$ can be derived exactly as done in [2], with a modification to the proof of theorem 1.2 presented in [6] to account for the VST based constraint to handle Poisson noise. In section 6 of [6], on taking into account our constraint, and following the steps to bound $\|B(\Delta)\|_2$ as done for Q1 above, equation 6.2 in section 6 of [6] changes to $\|A(H)\|_2 \leq 4\varepsilon \sqrt{(T + c)}$, where $H = \Delta$ and $\Delta$ corresponds to the linear operator $B(\cdot)$ in Q1. Since the upper bound for $\|A(H)\|_2$ is changed only by a multiplicative constant, the rest of the proof proceeds in a similar manner to [6], giving us a bound of the form similar to that in [2] with a different constant on the R.H.S., i.e.

$$\|\hat{A} - A^*\|_F \leq \frac{C_n \varepsilon}{\sqrt{n}},$$

(7)

due to the VST constraint. We see that the bound we obtained has a dependence on $\sqrt{n}$ as opposed to the theorem 1.2 in [6]. This difference arises due to the fact that the noise vector in [6] is assumed to be a sphere of radius $\sqrt{n}$, and hence we see a dependence on $\sqrt{n}$ in the 2-stage sparse recovery algorithm as no such assumption is taken for the noise vector in this case. The $\hat{A}$ estimated from above step now can be used to get the sparse estimate $\hat{X}$. The upper bound on $\|\hat{X} - X^*\|_2$ can now be derived exactly as done in [2]. This completes the proof sketch for Q1 and Q2.

### 3. REFERENCES


