Block-sparse Solutions using Kernel Block RIP and its Application to Group Lasso (Supplementary material)

Rahul Garg
IBM T.J. Watson research center
grahul@us.ibm.com

Rohit Khandekar
IBM T.J. Watson research center
rohitk@us.ibm.com

1 Supplementary Material

1.1 Proof of Theorem 3.1

Here we outline a proof based on a reduction to the block RIP and a theorem of Eldar and Mishali [1]. We construct a matrix $\Psi \in \Re^{m\times n}$ as follows. For every block i, we orthogonalize the columns of Φ_i to obtain Ψ_i. We further normalize the columns in Ψ_i to have unit ℓ_2-norm. Thus the columns of Ψ_i form an orthonormal basis of the column space of Φ_i. Thus for any x_i, there exists x'_i such that $\Phi_i x_i = \Psi_i x'_i$ and vice versa. Now from the definition of kernel block RIP, it is clear that the kernel block isometry constants of Φ are identical to those of Ψ. Furthermore, since $\|\Psi_i x'_i\|_2 = \|x'_i\|_2$, the kernel block isometry constants of Ψ are identical to its block isometry constants. Thus the block isometry constant of Ψ satisfies $\delta_{2s} < \sqrt{2}-1$. We now consider the program

$$\min \sum_{i=1}^{k} \|x'_i\|_2 \quad \text{subject to} \quad y = \Psi x'.$$

From the theorem of Eldar and Mishali [1], this program has a unique optimum solution $\hat{x}'$ that forms a unique s-block-sparse solution to the program $y = \Psi x'$.

Now note that any s-block-sparse solution $\hat{x}$ to $y = \Phi x$ satisfies $\Phi_i \hat{x}_i = \Psi_i \hat{x}'_i$ for all i and vice versa. Furthermore, any optimum solution $\hat{x}$ to (4) also satisfies $\Phi_i \hat{x}_i = \Psi_i \hat{x}'_i$ for all i and vice versa. Thus the proof of Theorem 2.1 follows.

1.2 Proof of Theorem 3.2

We begin with some notation. For a positive integer n, let $[n] = \{1, 2, \ldots, n\}$. For a vector $x \in \Re^n$, we use $\|x\|$ to denote $\|x\|_2$. For a matrix $\Phi \in \Re^{m\times n}$ and a subset $T \in [k]$ of blocks $\mathcal{B}$ of Φ, let $\Phi_T \in \Re^{m\times \sum_{i\in T} n_i}$ be the matrix Φ restricted to blocks T. Similarly, let $x_T \in \Re^{\sum_{i\in T} n_i}$ denote the vector x restricted to blocks T and let I_T denote the identity matrix of size $\sum_{i\in T} n_i \times \sum_{i\in T} n_i$.

In order to simplify the presentation, we first assume that $\Phi_i^\top \Phi_i = I \in \Re^{n_i\times n_i}$, i.e., the columns in Φ_i form an orthonormal basis of their span. We can transform Φ_i to satisfy this property as follows. If a column in block i lies in the span of the other columns in block i, we can discard it. Therefore we assume that $\Phi_i^\top \Phi_i$ is a *full rank* symmetric matrix. Let $A_i \in \Re^{n_i\times n_i}$ be a symmetric matrix such that $A_i^{-2} = \Phi_i^\top \Phi_i$. We now apply change of basis by replacing Φ_i with $\Phi_i A_i$. This change of basis does not affect the original problem, since the system $\Phi_i x_i = y_i$ has a non-zero solution $x_i \in \Re^{n_i\times n_i}$ if and only if $\Phi_i A_i x'_i = y_i$ has a non-zero solution $x'_i = A_i^{-1} x_i \in \Re^{n_i\times n_i}$. Thus we have $\|\Phi_i x_i\| = \|x_i\|$ for all $x_i \in \Re^{n_i}$.

The above change of basis transformation, however, will affect the bound on $\|x^* - x\|_2^2$ as we describe later.

Our proof is similar to and motivated by [2]. Before we begin, we prove two important lemmas.

Lemma 1.1 *For any $x \in \Re^n$ and a subset $T \subseteq [k]$ of blocks, we have $\|(I_T - \Phi_T^\top \Phi_T)x_T\| \le \delta^{\mathcal{B}}_{|T|}\|x_T\|$.*

Proof. Observe that the largest and the smallest eigenvalues, $\sigma_{\max}$ and $\sigma_{\min}$, of the symmetric matrix $I_T - \Phi_T^\top \Phi_T$ can be bounded as

$$\begin{aligned}
\sigma_{\max} &= \max_{v:\|v\|=1} v^\top (I_T - \Phi_T^\top \Phi_T) v \\
&\le 1 - \min_{v:\|v\|=1} v^\top \Phi_T^\top \Phi_T v \le 1 - (1 - \delta^{\mathcal{B}}_{|T|}) = \delta^{\mathcal{B}}_{|T|}, \\
\sigma_{\min} &= \min_{v:\|v\|=1} v^\top (I_T - \Phi_T^\top \Phi_T) v \\
&\le 1 - \max_{v:\|v\|=1} v^\top \Phi_T^\top \Phi_T v \le 1 - (1 + \delta^{\mathcal{B}}_{|T|}) = -\delta^{\mathcal{B}}_{|T|}.
\end{aligned}$$

Thus all the eigenvalues lie in the range $[-\delta^{\mathcal{B}}_{|T|}, \delta^{\mathcal{B}}_{|T|}]$ and the lemma follows. ∎

Lemma 1.2 *For any $x \in \Re^n$ and two* disjoint *subsets $T, U \subseteq [k]$ of blocks, we have $\|\Phi_U^\top \Phi_T x_T\| \le \delta^{\mathcal{B}}_{|T\cup U|}\|x_T\|$.*

Proof. Let $S = T \cup U$. Note that $\Phi_U^\top \Phi_T$ is a submatrix of $\Phi_S^\top \Phi_S - I_S$. Since the spectral norm of a submatrix does not exceed the spectral norm of the entire matrix, we have $\|\Phi_U^\top \Phi_T\| \leq \|\Phi_S^\top \Phi_S - I_S\| \leq \delta^{\mathcal{B}}_{|S|}$, where the last inequality follows form Lemma 1.1. ∎

Let $x^{[t]}$ denote the value of vector x after t iterations. Let x^* be the optimum solution. Let $r^{[t]} = x^* - x^{[t]}$. We now state our key lemma which directly implies Theorem 2.2.

Lemma 1.3 *The error vector $r^{[t]}$ shrinks in each iteration. This shrinkage can be quantified in terms of $\delta^{\mathcal{B}}_{2s}$ and $\delta^{\mathcal{B}}_{3s}$ as follows.*

$$\|r^{[t+1]}\| \leq \min\{\sqrt{3} \cdot \delta^{\mathcal{B}}_{2s}, \delta^{\mathcal{B}}_{3s}\} \cdot \phi \cdot \|r^{[t]}\|.$$

Proof. Let $B_t = \mathtt{supp}^{\mathcal{B}}(x^*) \cup \mathtt{supp}^{\mathcal{B}}(x^{[t]})$. Let $\hat{r}^{[t+1]} = x^* - (x^{[t]} + \Phi^\top \Phi r^{[t]}) = (I - \Phi^\top \Phi) r^{[t]}$. The proof is based on the following two claims.

Claim 1.1 $\|\hat{r}^{[t+1]}_{B_{t+1}}\| \leq \min\{\sqrt{3} \cdot \delta^{\mathcal{B}}_{2s}, \delta^{\mathcal{B}}_{3s}\} \cdot \|r^{[t]}\|$.

Claim 1.2 $\|r^{[t+1]}\| \leq \phi \cdot \|\hat{r}^{[t+1]}_{B_{t+1}}\|$.

Proof of Claim 1.1. Since $\mathtt{supp}^{\mathcal{B}}(r^{[t]}) \subseteq B_t$, we have

$$\begin{aligned}
\hat{r}^{[t+1]}_{B_t \cup B_{t+1}} &= I_{B_t \cup B_{t+1}} r^{[t]}_{B_t \cup B_{t+1}} - \Phi^\top_{B_t \cup B_{t+1}} (\Phi r^{[t]}) \\
&= I_{B_t \cup B_{t+1}} r^{[t]}_{B_t \cup B_{t+1}} - \\
&\quad \Phi^\top_{B_t \cup B_{t+1}} \Phi_{B_t \cup B_{t+1}} r^{[t]}_{B_t \cup B_{t+1}} \\
&= (I_{B_t \cup B_{t+1}} - \Phi^\top_{B_t \cup B_{t+1}} \Phi_{B_t \cup B_{t+1}}) r^{[t]}_{B_t \cup B_{t+1}}.
\end{aligned}$$

Thus from Lemma 1.1, we have $\|\hat{r}^{[t+1]}_{B_{t+1}}\| \leq \|\hat{r}^{[t+1]}_{B_t \cup B_{t+1}}\| \leq \delta^{\mathcal{B}}_{3s} \cdot \|r^{[t]}_{B_t \cup B_{t+1}}\| = \delta^{\mathcal{B}}_{3s} \cdot \|r^{[t]}\|$. Thus we have established the bound in terms of $\delta^{\mathcal{B}}_{3s}$.

We now prove the bound in terms of $\delta^{\mathcal{B}}_{2s}$. Since $\mathtt{supp}^{\mathcal{B}}(r^{[t]}) \subseteq B_t$, we have

$$\begin{aligned}
\hat{r}^{[t+1]}_{B_t} &= I_{B_t} r^{[t]}_{B_t} - \Phi^\top_{B_t} (\Phi r^{[t]}) \\
&= I_{B_t} r^{[t]}_{B_t} - \Phi^\top_{B_t} \Phi_{B_t} r^{[t]}_{B_t} \\
&= (I_{B_t} - \Phi^\top_{B_t} \Phi_{B_t}) r^{[t]}_{B_t}.
\end{aligned}$$

Thus from Lemma 1.1, we have

$$\|\hat{r}^{[t+1]}_{B_t}\| \leq \delta^{\mathcal{B}}_{2s} \cdot \|r^{[t]}_{B_t}\| = \delta^{\mathcal{B}}_{2s} \cdot \|r^{[t]}\|. \quad (1)$$

Similarly, we have

$$\begin{aligned}
\hat{r}^{[t+1]}_{B_{t+1} \setminus B_t} &= I_{B_{t+1} \setminus B_t} r^{[t]}_{B_{t+1} \setminus B_t} - \Phi^\top_{B_{t+1} \setminus B_t} (\Phi r^{[t]}) \\
&= -\Phi^\top_{B_{t+1} \setminus B_t} \Phi_{B_t} r^{[t]}_{B_t} \\
&= -\Phi^\top_{B_{t+1} \setminus B_t} \Phi_{B_{t+1} \cap B_t} r^{[t]}_{B_{t+1} \cap B_t} \\
&\quad -\Phi^\top_{B_{t+1} \setminus B_t} \Phi_{B_t \setminus B_{t+1}} r^{[t]}_{B_t \setminus B_{t+1}}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\|\hat{r}^{[t+1]}_{B_{t+1} \setminus B_t}\|^2 &\leq 2\|\Phi^\top_{B_{t+1} \setminus B_t} \Phi_{B_{t+1} \cap B_t} r^{[t]}_{B_{t+1} \cap B_t}\|^2 + \\
&\quad 2\|\Phi^\top_{B_{t+1} \setminus B_t} \Phi_{B_t \setminus B_{t+1}} r^{[t]}_{B_t \setminus B_{t+1}}\|^2 \quad (2) \\
&\leq 2 \cdot (\delta^{\mathcal{B}}_{2s})^2 \cdot \left(\|r^{[t]}_{B_{t+1} \cap B_t}\|^2 + \|r^{[t]}_{B_t \setminus B_{t+1}}\|^2\right) \quad (3) \\
&= 2 \cdot (\delta^{\mathcal{B}}_{2s})^2 \cdot \|r^{[t]}\|^2. \quad (4)
\end{aligned}$$

The inequality (2) follows from the identity $\|u+v\|^2 \leq 2(\|u\|^2 + \|v\|^2)$. The inequality (3) follows from two applications of Lemma 1.2. Now combining (1) and (4), we get

$$\begin{aligned}
\|r^{[t+1]}_{B_{t+1}}\|^2 &= \|r^{[t+1]}_{B_t}\|^2 + \|r^{[t+1]}_{B_{t+1} \setminus B_t}\|^2 \\
&\leq (\delta^{\mathcal{B}}_{2s})^2 \cdot \|r^{[t]}\|^2 + 2 \cdot (\delta^{\mathcal{B}}_{2s})^2 \cdot \|r^{[t]}\|^2 \\
&= 3 \cdot (\delta^{\mathcal{B}}_{2s})^2 \cdot \|r^{[t]}\|^2.
\end{aligned}$$

Thus the proof of Claim 1.1 is complete.

Proof of Claim 1.2. Since $\phi^2 = 1 + \phi$, it is enough to prove

$$\|r^{[t+1]}\|^2 \leq (1 + \phi) \cdot \|\hat{r}^{[t+1]}_{B_{t+1}}\|^2. \quad (5)$$

Without loss of generality, we assume that $|\mathtt{supp}^{\mathcal{B}}(x^{[t+1]})| = |\mathtt{supp}^{\mathcal{B}}(x^*)| = s$. Let $v = x^{[t]} + \Phi^\top \Phi r^{[t]}$ and let $A = \mathtt{supp}^{\mathcal{B}}(x^{[t+1]}) \setminus \mathtt{supp}^{\mathcal{B}}(x^*), B = \mathtt{supp}^{\mathcal{B}}(x^{[t+1]}) \cap \mathtt{supp}^{\mathcal{B}}(x^*), C = \mathtt{supp}^{\mathcal{B}}(x^*) \setminus \mathtt{supp}^{\mathcal{B}}(x^{[t+1]})$. Note that $|A| = |C|$. Since $x^{[t+1]} = H_s(v)$, from the definition of hard-thresholding H_s, we get that $\|v_i\|^2 = \|\Phi_i v_i\|^2 \leq \|\Phi_j v_j\|^2 = \|v_j\|^2$ for all $i \in C$ and $j \in A$. Note that $r^{[t+1]} = x^* - H_s(v) = x^* - v_{A \cup B}$ and $\hat{r}^{[t+1]} = x^* - v$ and hence $\hat{r}^{[t+1]}_{B_{t+1}} = (x^* - v)_{A \cup B \cup C}$. Therefore the right-hand-side of (5) minus the left-hand-side of (5) is

$$\begin{aligned}
&(1+\phi)\left(\sum_{i \in A} \|v_i\|^2 + \sum_{i \in B} \|x^*_i - v_i\|^2 + \sum_{i \in C} \|x^*_i - v_i\|^2\right) \\
&\quad - \left(\sum_{i \in A} \|v_i\|^2 + \sum_{i \in B} \|x^*_i - v_i\|^2 + \sum_{i \in C} \|x^*_i\|^2\right).
\end{aligned}$$

The above expression is at least

$$\begin{aligned}
&\phi \sum_{i \in A} \|v_i\|^2 + \sum_{i \in C} \left((1+\phi)\|x^*_i - v_i\|^2 - \|x^*_i\|^2\right) \\
\geq &\sum_{i \in C} \left(\phi\|v_i\|^2 + (1+\phi)\|x^*_i - v_i\|^2 - \|x^*_i\|^2\right).
\end{aligned}$$

The inequality follows from the fact that $\sum_{i \in C} \|v_i\|^2 \leq \sum_{j \in A} \|v_j\|^2$ as observed above. Each term on the right-hand-side of the above inequality can be simplified to

$$\begin{aligned}
&(1+2\phi)\|v_i\|^2 - 2(1+\phi) x^*_i \cdot v_i + \phi\|x^*_i\|^2 \\
= &\left(1 + 2\phi - \frac{(1+\phi)^2}{\phi}\right)\|v_i\|^2 + \left\|\frac{1+\phi}{\sqrt{\phi}} v_i - \sqrt{\phi} x^*_i\right\|^2.
\end{aligned}$$

Thus a sufficient condition for this term to be non-negative for any value of v_i and x_i^* is $(1+2\phi)\phi \geq (1+\phi)^2$. This is equivalent to $1+\phi-\phi^2 \leq 0$. This condition holds since in fact $1+\phi = \phi^2$ for golden ratio $\phi = (1+\sqrt{5})/2$. Thus the proof of Claim 1.2 is complete.

Combining Claims 1.1 and 1.2, we get Lemma 1.3. ∎

Now recall that we transformed Φ_i in the beginning so that it satisfied $\Phi_i^\top \Phi_i = I \in \Re^{n_i \times n_i}$ and therefore $\|y - \Phi x\|_2^2 = \|x^* - x\|_2^2$. Thus the bound on the norm of the error vector $r^{[t]}$ proved in Lemma 1.1 in fact implies that

$$\|y - \Phi x\|_2^2 \leq 2\|y\|_2^2 \cdot \left[\phi \cdot \min\{\sqrt{3} \cdot \delta_{2s}^{\mathcal{B}}, \delta_{3s}^{\mathcal{B}}\}\right]^t$$

holds after t iterations as claimed in Theorem 3.2. For general Φ_i, from the definition of $\lambda_{\min}$ given just before Theorem 3.2, we have $\|x^* - x\|_2^2 \leq \frac{1}{\lambda_{\min}} \cdot \|y - \Phi x\|_2^2$. Thus

$$\|x^* - x\|_2^2 \leq \frac{2\|y\|_2^2}{\lambda_{\min}} \cdot \left[\phi \cdot \min\{\sqrt{3} \cdot \delta_{2s}^{\mathcal{B}}, \delta_{3s}^{\mathcal{B}}\}\right]^t$$

holds as well after t iterations as claimed in Theorem 3.2.

References

[1] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theor.*, 55(11):5302–5316, 2009.

[2] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2008.