Supplementary Materials: Signal and Noise Statistics Oblivious Orthogonal Matching Pursuit

Sreejith Kallummil [1] **Sheetal Kalyani** [2]

1. Proofs of Theorems 1-6

1.1. Appendix A: Proof of Theorem 1

Statement of Theorem 1:- Assume that the matrix $\mathbf{X}$ satisfies the RIC constraint $\delta_{k_0+1} < \dfrac{1}{\sqrt{k_0+1}}$ and $k_{max} > k_0$. Then
a). $RR(k_{min}) \overset{P}{\to} 0$ as $\sigma^2 \to 0$.
b). $\lim\limits_{\sigma^2 \to 0} \mathbb{P}(k_{min} = k_0) = 1$.

Proof. We first prove statement b) of Theorem 1. By Lemma 1, we have $k_{min} = k_0$ once $\|\mathbf{w}\|_2 \leq \epsilon_{omp}$. Hence, $\mathbb{P}(k_{min} = k_0) \geq \mathbb{P}(\|\mathbf{w}\|_2 \leq \epsilon_{omp})$. Since $\|\mathbf{w}\|_2 \overset{P}{\to} 0$ as $\sigma^2 \to 0$, it follows from the definition of convergence in probability that $\lim\limits_{\sigma^2 \to 0} \mathbb{P}(\|\mathbf{w}\|_2 \leq \epsilon_{omp}) = 1$ which implies statement b).

Next we prove statement a) of Theorem 1. When $\|\mathbf{w}\|_2 \leq \epsilon_{omp}$, we have $k_{min} = k_0$ which in turn implies that $\mathcal{S}^k_{omp} \subseteq \mathcal{S}$ for $k \leq k_0$. Following the discussions in the article, we have $\mathbf{r}^{k_0} = (\mathbf{I}_n - \mathbf{P}_{k_0})\mathbf{w}$ which in turn imply that $\|\mathbf{r}^{k_0}\|_2 = \|(\mathbf{I}_n - \mathbf{P}_{k_0})\mathbf{w}\|_2 \leq \|\mathbf{w}\|_2$. For $k < k_0$, we have $\mathbf{r}^k = (\mathbf{I}_n - \mathbf{P}_k)\mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} + (\mathbf{I}_n - \mathbf{P}_k)\mathbf{w}$. Since, $(\mathbf{I}_n - \mathbf{P}_k)\mathbf{X}_{\mathcal{S}^k_{omp}}\boldsymbol{\beta}_{\mathcal{S}^k_{omp}} = \mathbf{0}_n$, it follows that $(\mathbf{I}_n - \mathbf{P}_k)\mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}} = (\mathbf{I}_n - \mathbf{P}_k)\mathbf{X}_{\mathcal{S}/\mathcal{S}^k_{omp}}\boldsymbol{\beta}_{\mathcal{S}/\mathcal{S}^k_{omp}}$.

Lemma 1. *Let $\mathcal{S}_1 \subset \{1, \dots, p\}$ and $\mathcal{S}_2 \subset \{1, \dots, p\}$ be two disjoint index sets and $\mathbf{P}_{\mathcal{S}_1}$ be a projection matrix onto $span(\mathbf{X}_{\mathcal{S}_1})$. Then for every $\mathbf{b} \in \mathbb{R}^{card(\mathcal{S}_2)}$*

$$(1 - \delta_{card(\mathcal{S}_1 \cup \mathcal{S}_2)})\|\mathbf{b}\|_2^2 \leq \|(\mathbf{I}_n - \mathbf{P}_{\mathcal{S}_1})\mathbf{X}_{\mathcal{S}_2}\mathbf{b}\|_2^2 \leq (1 + \delta_{card(\mathcal{S}_1 \cup \mathcal{S}_2)})\|\mathbf{b}\|_2^2 \tag{1}$$

(Wen et al., 2016)

[1]Department of Electrical Engineering, IIT Madras, India [2]Department of Electrical Engineering, IIT Madras, India. Correspondence to: Sreejith Kallummil <sreejith.k.venugopal@gmail.com>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

It follows from Lemma 1 that

$$\begin{aligned}\|(\mathbf{I}_n - \mathbf{P}_k)\mathbf{X}_{\mathcal{S}/\mathcal{S}^k_{omp}}\boldsymbol{\beta}_{\mathcal{S}/\mathcal{S}^k_{omp}}\|_2 &\geq \sqrt{1-\delta_{k_0}}\|\boldsymbol{\beta}_{\mathcal{S}/\mathcal{S}^k_{omp}}\|_2 \\ &\geq \sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min},\end{aligned} \tag{2}$$

where $\boldsymbol{\beta}_{min} = \min\limits_{j \in \mathcal{S}} |\boldsymbol{\beta}_j|$. This along with the triangle inequality gives

$$\|\mathbf{r}^k\|_2 \geq \sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min} - \|\mathbf{w}\|_2 \tag{3}$$

for $k < k_0$. Consequently, $RR(k_{min})$ when $\|\mathbf{w}\|_2 \leq \epsilon_{omp}$ satisfies the bound

$$RR(k_{min}) \leq \frac{\|\mathbf{w}\|_2}{\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min} - \|\mathbf{w}\|_2} \tag{4}$$

When $\|\mathbf{w}\|_2 > \epsilon_{omp}$, it is likely that $k_{min} \geq k_0$. However, it is still true that $RR(k_{min}) \leq 1$. Hence,

$$RR(k_{min}) \leq \frac{\|\mathbf{w}\|_2}{\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min} - \|\mathbf{w}\|_2}\mathcal{I}_{\|\mathbf{w}\|_2 \leq \epsilon_{omp}} + \mathcal{I}_{\|\mathbf{w}\|_2 > \epsilon_{omp}}. \tag{5}$$

Here $\mathcal{I}_x$ is an indicator function taking value one when $x > 0$ and zero otherwise. Now $\|\mathbf{w}\|_2 \overset{P}{\to} 0$ as $\sigma^2 \to 0$ implies that $\dfrac{\|\mathbf{w}\|_2}{\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min} - \|\mathbf{w}\|_2} \overset{P}{\to} 0$, $\mathcal{I}_{\|\mathbf{w}\|_2 \leq \epsilon_{omp}} \overset{P}{\to} 1$ and $\mathcal{I}_{\|\mathbf{w}\|_2 > \epsilon_{omp}} \overset{P}{\to} 0$ as $\sigma^2 \to 0$. This along with $RR(k_{min}) \geq 0$ implies that $RR(k_{min}) \overset{P}{\to} 0$ as $\sigma^2 \to 0$. This proves statement a) of Theorem 1. □

1.2. Appendix B: Projection matrices and distributions (used in the proof of Theorem 2)

Consider two fixed index set $\mathcal{S}_1 \subset \mathcal{S}_2$ of cardinality k_1 and k_2. Let $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$ be two projection matrices projecting onto the column spaces $span(\mathbf{X}_{\mathcal{S}_1})$ and $span(\mathbf{X}_{\mathcal{S}_2})$. When $\mathbf{w} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$, it follows from standard results that $\|\mathbf{P}_{\mathcal{S}_1}\mathbf{w}\|_2/\sigma^2 \sim \chi^2_{k_1}$ and $\|(\mathbf{I}_n - \mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2/\sigma^2 \sim \chi^2_{n-k_1}$. Please note that χ^2_k is a central chi squared random variable with k degrees of freedom. Using the properties of projection matrices, one can show that $(\mathbf{I}_n - \mathbf{P}_{\mathcal{S}_2})(\mathbf{P}_{\mathcal{S}_2} - \mathbf{P}_{\mathcal{S}_1}) = O_n$, the $n \times n$ all zero matrix. This implies that $\|(\mathbf{I}_n - \mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2 = \|(\mathbf{I}_n - \mathbf{P}_{\mathcal{S}_2})\mathbf{w} + (\mathbf{P}_{\mathcal{S}_2} - \mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2 =$

$\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2+\|(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2$. Further, the orthogonality of $(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})$ and $(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})$ implies that the random variables $\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2$ and $\|(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2$ are uncorrelated and hence independent ($\mathbf{w}$ is Gaussian). Also note that $(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})$ is a projection matrix projecting onto the column space of $span(\mathbf{X}_{\mathcal{S}_2})\cap span(\mathbf{X}_{\mathcal{S}_1})^{\perp}$ of dimensions k_2-k_1. Hence, $\|(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2/\sigma^2\sim\chi^2_{k_2-k_1}$. It is well known in statistics that $X_1/(X_1+X_2)$, where $X_1\sim\chi^2_{n_1}$ and $X_2\sim\chi^2_{n_2}$ are two independent chi squared random variables have a $\mathbb{B}(\frac{n_1}{2},\frac{n_2}{2})$ distribution(Ravishanker & Dey, 2001). Applying these results to the ratio $\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2/\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2$ gives

$$\begin{aligned}\frac{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2}{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2}&=\frac{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2}{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2+\|(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2}\\&=\frac{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2/\sigma^2}{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}_2})\mathbf{w}\|_2^2/\sigma^2+\|(\mathbf{P}_{\mathcal{S}_2}-\mathbf{P}_{\mathcal{S}_1})\mathbf{w}\|_2^2/\sigma^2}\\&\sim\frac{\chi^2_{n-k_2}}{\chi^2_{n-k_2}+\chi^2_{k_2-k_1}}\\&\sim\mathbb{B}(\frac{n-k_2}{2},\frac{k_2-k_1}{2})\end{aligned}\tag{6}$$

1.3. Appendix C: Proof of Theorem 2

Statement of Theorem 2:- Let $F_{a,b}(x)$ denotes the cumulative distribution function of a $\mathbb{B}(a,b)$ random variable. Then $\forall\sigma^2>0$, $\Gamma^{\alpha}_{RRT}(k)=\sqrt{F^{-1}_{\frac{n-k}{2},0.5}\left(\frac{\alpha}{k_{max}(p-k+1)}\right)}$ satisfies

$$\mathbb{P}(RR(k)>\Gamma^{\alpha}_{RRT}(k),\forall k>k_{min})\geq 1-\alpha,. \tag{7}$$

Proof. Reiterating, $k_{min}=\min\{k:\mathcal{S}\subseteq\mathcal{S}^k_{omp}\}$, where $\mathcal{S}^k_{omp}$ is the support estimate returned by OMP at k^{th} iteration. k_{min} is a R.V taking values in $\{k_0,k_0+1,\dots,k_{max},\infty\}$. The proof of Theorem 2 proceeds by conditioning on the R.V k_{min} and by lower bounding $RR(k)$ for $k>k_{min}$ using artificially created random variables with known distribution.

Case 1:- Conditioning on $k_0\leq k_{min}=j<k_{max}$. Consider the step $k-1$ of the Alg where $k\geq j$. Current support estimate $\mathcal{S}^{k-1}_{omp}$ is itself a R.V. Let $\mathcal{L}_{k-1}\subseteq\{[p]/\mathcal{S}^{k-1}_{omp}\}$ represents the set of all all possible indices l at stage $k-1$ such that $\mathbf{X}_{\mathcal{S}^{k-1}_{omp}\cup l}$ is full rank. Clearly, $card(\mathcal{L}_{k-1})\leq p-card(\mathcal{S}^{k-1}_{omp})=p-k+1$. Likewise, let $\mathcal{K}^{k-1}$ represents the set of all possibilities for the set $\mathcal{S}^{k-1}_{omp}$ that would also satisfy the constraint $k\geq k_{min}=j$. Conditional on both $k_{min}=j$ and $\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp}$, the R.V $\|\mathbf{r}^{k-1}\|_2^2\sim\sigma^2\chi^2_{n-k+1}$ and $\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}^{k-1}_{omp}\cup l})\mathbf{w}\|_2^2\sim\sigma^2\chi^2_{n-k}$. Define the conditional R.V,

$$Z_k^l|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\}=\frac{\|(\mathbf{I}_n-\mathbf{P}_{\mathcal{S}^{k-1}_{omp}\cup l})\mathbf{w}\|_2^2}{\|\mathbf{r}^{k-1}\|_2^2},\tag{8}$$

for $l\in\mathcal{L}_{k-1}$. Following the discussions in Appendix B, one have

$$Z_k^l|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\}\sim\mathcal{B}\left(\frac{n-k}{2},\frac{1}{2}\right),\ \forall l\in\mathcal{L}_{k-1}.\tag{9}$$

Since the index selected in the $k-1^{th}$ iteration belongs to $\mathcal{L}_{k-1}$, it follows that conditioned on $\{\mathcal{S}^{k-1}_{omp},k_{min}\}$,

$$\min_{l\in\mathcal{L}_{k-1}}\sqrt{Z_k^l|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\}}\leq RR(k).\tag{10}$$

Note that $\Gamma^{\alpha}_{RRT}(k)=\sqrt{F^{-1}_{\frac{n-k}{2},0.5}\left(\frac{\alpha}{k_{max}(p-k+1)}\right)}$. It follows that

$$\begin{aligned}&\mathbb{P}(RR(k)<\Gamma^{\alpha}_{RRT}(k)|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\})\\&\leq\mathbb{P}(\min_{l\in\mathcal{L}_{k-1}}\sqrt{Z_k^l}|<\Gamma^{\alpha}_{RRT}(k)|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\})\\&\overset{(a)}{\leq}\sum_{l\in\mathcal{L}_{k-1}}\mathbb{P}(Z_k^l<(\Gamma^{\alpha}_{RRT}(k))^2|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\})\\&\overset{(b)}{\leq}\frac{\alpha}{k_{max}}\end{aligned}\tag{11}$$

(a) in Eqn.11 follows from the union bound. By the definition of $\Gamma^{\alpha}_{RRT}(k)$, $\mathbb{P}(Z_k^l<(\Gamma^{\alpha}_{RRT}(k))^2)=\dfrac{\alpha}{k_{max}(p-k+1)}$. (b) follows from this and the fact that $card(\mathcal{L}_{k-1})\leq p-k+1$. Next we eliminate the random set $\mathcal{S}^k_{omp}$ from (11) using the law of total probability, i.e.,

$$\begin{aligned}&\mathbb{P}(RR(k)<\Gamma^{\alpha}_{RRT}(k)|k_{min=j})\\&=\sum_{s^{k-1}_{omp}\in\mathcal{K}^{k-1}}\mathbb{P}(RR(k)<\Gamma^{\alpha}_{RRT}(k)|\{\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp},k_{min}=j\})\\&\quad\times\mathbb{P}(\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp}|k_{min}=j)\\&\leq\sum_{s^{k-1}_{omp}\in\mathcal{K}^{k-1}}\frac{\alpha}{k_{max}}\mathbb{P}(\mathcal{S}^{k-1}_{omp}=s^{k-1}_{omp}|k_{min}=j)\\&=\frac{\alpha}{k_{max}},\forall k>k_{min}=j.\end{aligned}\tag{12}$$

Now applying the union bound and (12) gives

$$\begin{aligned}&\mathbb{P}(RR(k)>\Gamma^{\alpha}_{RRT}(k),\forall k>k_{min}|k_{min}=j)\\&\geq 1-\sum_{k=j+1}^{k_{max}}\mathbb{P}(RR(k)<\Gamma^{\alpha}_{RRT}(k)|k_{min}=j)\\&\geq 1-\alpha\frac{k_{max}-j}{k_{max}}\geq 1-\alpha.\end{aligned}\tag{13}$$

Case 2:- Conditioning on $k_{min}=\infty$ **and** $k_{min}=k_{max}$. In both these cases, the set $\{k_0\leq k\leq k_{max}:k>k_{min}\}$ is empty. Applying the usual convention of assigning the minimum value of empty sets to ∞, one has for $j\in\{k_{max},\infty\}$

$$\begin{aligned}&\mathbb{P}(RR(k)>\Gamma_{RRT}^{\alpha}(k),\forall k>k_{min}|k_{min}=j)\\&\geq \mathbb{P}(\min_{k>j}RR(k)>\Gamma_{RRT}^{\alpha}(k),\forall k>k_{min}|k_{min}=j)\\&=1\geq 1-\alpha.\end{aligned} \tag{14}$$

Again applying law of total probability to remove the conditioning on k_{min} and bounds (13) and (14) give

$$\begin{aligned}&\mathbb{P}(RR(k)>\Gamma_{RRT}^{\alpha}(k),\forall k>k_{min})\\&=\sum_{j\in\{k_0,\dots,k_{max},\infty\}}\mathbb{P}(RR(k)>\Gamma_{RRT}^{\alpha}(k),\forall k>k_{min}|k_{min}=j)\\&\quad\times\mathbb{P}(k_{min}=j)\\&\geq\sum_{j\in\{k_0,\dots,k_{max},\infty\}}(1-\alpha)\mathbb{P}(k_{min}=j)=1-\alpha.\end{aligned} \tag{15}$$

Hence proved. □

Appendix D: Proof of Theorem 3

Statement of Theorem 3:- Let $k_{max}\geq k_0$ and matrix $\mathbf{X}$ satisfies $\delta_{k_0+1}<\frac{1}{\sqrt{k_0+1}}$. Then RRT can recover the true support $\mathcal{S}$ with probability greater than $1-1/n-\alpha$ provided that $\epsilon_{\sigma}<\min(\epsilon_{omp},\epsilon_{RRT})$, where

$$\epsilon_{RRT}=\frac{\Gamma_{RRT}^{\alpha}(k_0)\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min}}{1+\Gamma_{RRT}^{\alpha}(k_0)}. \tag{16}$$

Proof. RRT support estimate $\mathcal{S}_{omp}^{k_{RRT}}$ where $k_{RRT}=\max\{k:RR(k)\leq\Gamma_{RRT}^{\alpha}(k)\}$ will be equal to $\mathcal{S}$ if the following three events occurs simultaneously.
A1). $\mathcal{S}_{omp}^{k_0}=\mathcal{S}$, i.e., $k_{min}=k_0$.
A2). $RR(k_0)<\Gamma_{RRT}^{\alpha}(k_0)$.
A3). $RR(k)>\Gamma_{RRT}^{\alpha}(k),\forall k\geq k_{min}$.

By Lemma 1 of the article, A1) is true once $\|\mathbf{w}\|_2\leq\epsilon_{omp}$. Next consider $RR(k_0)$ assuming that $\|\mathbf{w}\|_2\leq\epsilon_{omp}$. Following the proof of Theorem 1, one has

$$RR(k_0)\leq\frac{\|\mathbf{w}\|_2}{\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min}-\|\mathbf{w}\|_2} \tag{17}$$

whenever $\|\mathbf{w}\|_2\leq\epsilon_{omp}$. Consequently, $RR(k_0)$ will be smaller than $\Gamma_{RRT}^{\alpha}(k_0)$ if $\frac{\|\mathbf{w}\|_2}{\sqrt{1-\delta_{k_0}}\boldsymbol{\beta}_{min}-\|\mathbf{w}\|_2}\leq\Gamma_{RRT}^{\alpha}(k_0)$ which in turn is true once $\|\mathbf{w}\|_2\leq\epsilon_{RRT}$. Hence, $\mathcal{A}_2$ is true once $\|\mathbf{w}\|_2\leq\min(\epsilon_{RRT},\epsilon_{omp})$. Consequently, $\epsilon_{\sigma}\leq\min(\epsilon_{RRT},\epsilon_{omp})$ implies that

$$\mathbb{P}(\mathcal{A}_1\cap\mathcal{A}_2)\geq 1-1/n. \tag{18}$$

By Theorem 2, it is true that $\mathbb{P}(\mathcal{A}_3)\geq 1-\alpha,\forall\sigma^2>0$. Together, we have $\mathbb{P}(\mathcal{A}_1\cap\mathcal{A}_2\cap\mathcal{A}_3)\geq 1-\alpha-1/n$ whenever $\epsilon_{\sigma}\leq\min(\epsilon_{RRT},\epsilon_{omp})$. □

1.4. Appendix E. Proof of Theorem 4

Statement of Theorem 4:- Let $k_{lim}=\lim\limits_{n\to\infty}k_0/n$, $p_{lim}=\lim\limits_{n\to\infty}\log(p)/n$, $\alpha_{lim}=\lim\limits_{n\to\infty}\log(\alpha)/n$ and $k_{max}=\min(p,[0.5(n+1)])$. Then $\Gamma_{RRT}^{\alpha}(k_0)=\sqrt{F^{-1}_{\frac{n-k_0}{2},0.5}\left(\frac{\alpha}{k_{max}(p-k_0+1)}\right)}$ satisfies the following asymptotic limits.
Case 1:-). $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0)=1$, whenever $k_{lim}<0.5$, $p_{lim}=0$ and $\alpha_{lim}=0$.
Case 2:-). $0<\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0)<1$, if $k_{lim}<0.5$, $\alpha_{lim}=0$ and $p_{lim}>0$. In particular, $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0)=\exp(\frac{-p_{lim}}{1-k_{lim}})$.
Case 3:- $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0)=0$ if $k_{lim}<0.5$, $\alpha_{lim}=0$ and $p_{lim}=\infty$.

Proof. Recall that $\Gamma_{RRT}^{\alpha}(k_0)=\sqrt{\Delta_{k_0}(n)}$, where $\Delta_{k_0}(n)=F^{-1}_{\frac{n-k_0}{2},\frac{1}{2}}\left(\frac{\alpha}{k_{max}(p-k_0+1)}\right)$ and $k_{max}=\min(p,[0.5(n+1)])$. Note that $q(x)=F_{a,b}^{-1}(x)$ is implicitly defined by the integral $\int_{t=0}^{q(x)}t^{a-1}(1-t)^{b-1}dt=x\int_{t=0}^{1}t^{a-1}(1-t)^{b-1}dt$. The R.H.S $\int_{t=0}^{1}t^{a-1}(1-t)^{b-1}dt$ is the famous Beta function $\mathcal{B}(a,b)$.

1.4.1. Proof of Case 1):-

We first consider the situation of $n\to\infty$ with $k_{lim}<0.5$, $p_{lim}=0$ and $\alpha_{lim}=0$. Define $x(n,p,k_0)=\frac{\alpha}{\min([0.5(n+1)],p)(p-k_0+1)}$. Depending on whether, $x(n,p,k_0)$ converges to zero with increasing n or not, we consider two special cases.

Special case 1: (fixed p, k_0, α and $n\to\infty$):- This regime has $p/n\to 0$ and $k_0/[0.5(n+1)]\to 0$ (since $k_0<p$), $\log(\alpha)/n\to 0$, however, $x(n,p,k_0)=\frac{\alpha}{\min([0.5(n+1)],p)(p-k_0+1)}$ is bounded away from zero. For $n>2p$, $x(n,p,k_0)=\frac{\alpha}{\min(p,[0.5(n+1)])(p-k_0+1)}$ reduces to $x(n,p,k_0)=\frac{\alpha}{p(p-k_0+1)}$. Using the standard limit $\lim\limits_{a\to\infty}F_{a,b}^{-1}(x)=1$ for every fixed $b\in(0,\infty)$ and $x\in(0,1)$ (see proposition 1, (Askitis, 2016)), it follows that $\lim\limits_{n\to\infty}\Delta_{k_0}(n)=\lim\limits_{n\to\infty}F^{-1}_{\frac{n-k_0}{2},0.5}(x(n,p,k_0)=1$. Since $\Delta_{k_0}(n)\to 1$ as $n\to\infty$, it follows that $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0)=\lim\limits_{n\to\infty}\sqrt{\Delta_{k_0}(n)}=1$.

Special Case 2: ($(n,p,k_0)\to\infty$ such that $\log(p)/n\to 0$, $\lim\limits_{n\to\infty}k_0/n<1$) and $\lim\limits_{n\to\infty}\log(\alpha)/n=0$:-

The sequence $x(n,p,k_0)$ converges to zero as $n \to \infty$. Expanding $F_{a,b}^{-1}(z)$ at $z=0$ using the expansion given in *http://functions.wolfram.com/GammaBetaErf /Inverse-BetaRegularized/06/01/02/* gives

$$F_{a,b}^{-1}(z) = (az\mathcal{B}(a,b))^{(1/a)} + \frac{b-1}{a+1}(az\mathcal{B}(a,b))^{(2/a)} + \frac{(b-1)(a^2+3ab-a+5b-4)}{2(a+1)^2(a+2)}(az\mathcal{B}(a,b))^{(3/a)} + O(z^{(4/a)}) \tag{19}$$

for all $a>0$. Here $\mathcal{B}(a,b)$ is the regular Beta function. For our case, we associate $a=\frac{n-k_0}{2}$, $b=1/2$ and $z=x(n,p,k_0)$.

We first evaluate the limit of the term $\rho(n,p,k_0,l) = (az\mathcal{B}(a,b))^{(l/a)} = \left(\frac{\frac{n-k_0}{2}\alpha\mathcal{B}(\frac{n-k_0}{2},0.5)}{\min(p,[0.5(n+1)])(p-k_0+1)}\right)^{\frac{2l}{n-k_0}}$ for $l \geq 1$. Then $\log(\rho(n,p,k_0,l))$ gives

$$\log(\rho(n,p,k_0,l)) = \frac{2l}{n-k_0}\log\left(\frac{\frac{n-k_0}{2}}{\min(p,[0.5(n+1)])}\right) + \frac{2l}{n-k_0}\log\left(\mathcal{B}(\frac{n-k_0}{2},0.5)\right) + \frac{2l}{n-k_0}\log(\alpha) - \frac{2l}{n-k_0}\log(p-k_0+1) \tag{20}$$

Clearly, the first, third and fourth term in the R.H.S of (20) converges to zero as $(n,p,k_0)\to\infty$ such that $\log(p)/n \to 0$, $\lim\limits_{n\to\infty} k_0/n < 1$ and $\lim\limits_{n\to\infty}\log(\alpha)/n = 0$. Using the asymptotic expansion $\mathcal{B}(a,b) = G(b)a^{-b}\left(1-\frac{b(b-1)}{2a}(1+O(\frac{1}{a}))\right)$ as $a\to\infty$ from [*http://functions.wolfram.com/GammaBetaErf/Beta/06/02/*] in the second[1] term of (20) gives

$$\lim_{n\to\infty}\frac{2l}{n-k_0}\log\left(\mathcal{B}(\frac{n-k_0}{2},0.5)\right) = 0. \tag{21}$$

whenever, $\lim\limits_{n\to\infty} k_0/n < 0.5$. Hence, when $(n,p,k_0)\to\infty$ such that $\log(p)/n \to 0$, $\lim\limits_{n\to\infty}k_0/n < 0.5$ and $\lim\limits_{n\to\infty}\log(\alpha)/n = 0$, one has $\lim\limits_{n\to\infty}\log(\rho(n,p,k_0,l)) = 0$ which in turn implies that $\lim\limits_{n\to\infty}\rho(n,p,k_0,l) = 1, \forall \ell$.

Note that the coefficient of $\rho(n,p,k_0,l)$ in (19) decays with $1/a = 2/(n-k_0)$ at large n. This along with $\lim\limits_{n\to\infty}\rho(n,p,k_0,l) = 1$ implies that all terms other than $l=1$ in (19) decays to zero as $n\to\infty$. Consequently, only the first term in (19), i.e., $\rho(n,p,k_0,1)$ is non zero as $n\to\infty$ and this term converges to one as $n\to\infty$. This implies that $\lim\limits_{n\to\infty}\Delta_{k_0}(n) = 1$. Since $\Delta_{k_0}\to 1$ as $n\to\infty$, it follows that $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0) = \lim\limits_{n\to\infty}\sqrt{\Delta_{k_0}(n)} = 1$.

[1] $G(b) = \int\limits_{t=0}^{\infty} e^{-x}x^{b-1}dx$ is the famous Gamma function.

1.4.2. Proof of Case 2):-

Next consider the situation where $n\to\infty$, $0<p_{lim}<\infty$ and $k_{lim}<0.5$. Here also the argument inside $F_{a,b}^{-1}(.)$, i.e., $x(n,p,k_0)$ converges to zero and hence the asymptotic expansion (19) and (20) is valid. Note that the limits $0<p_{lim}<\infty$ and $k_{lim}<0.5$ implies that $k_0/p\to 0$ as $n\to\infty$. Applying these limits and $\alpha_{lim}=0$ in (20) gives

$$-\infty < \lim_{n\to\infty}\log(\rho(n,p,k_0,l)) = -\frac{2lp_{lim}}{1-k_{lim}} < 0 \text{ and} \tag{22}$$

$$0 < \lim_{n\to\infty}\rho(n,p,k_0,l) = e^{-\frac{2lp_{lim}}{1-k_{lim}}} < 1. \tag{23}$$

for every $l<\infty$. Since the coefficients of $\rho(n,p,k_0,l)$ for $l>1$ decays at the rate $1/n$, it follows that $0<\lim\limits_{n\to\infty}\Delta_{k_0}(n) = \lim\limits_{n\to\infty}\rho(n,p,k_0,1) = e^{-\frac{2p_{lim}}{1-k_{lim}}} < 1$. This limit in turn implies that $0<\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0) = \lim\limits_{n\to\infty}\sqrt{\Delta_{k_0}(n)} = e^{-\frac{p_{lim}}{1-k_{lim}}} < 1$.

1.4.3. Proof of Case 3):-

Next consider the situation where $n\to\infty$, $p_{lim}=\infty$, $k_{lim}<0.5$ and $\alpha_{lim}=0$. Here also the argument inside $F_{a,b}^{-1}(.)$, i.e., $x(n,p,k_0)$ converges to zero and hence the asymptotic expansion (19) and (20) is valid. Applying the limits $p_{lim}=0$, $k_{lim}<0.5$ and $\alpha_{lim}=0$ in (20) gives

$$\lim_{n\to\infty}\log(\rho(n,p,k_0,l)) = -\infty \text{ and} \tag{24}$$

$$\lim_{n\to\infty}\rho(n,p,k_0,l) = 0. \tag{25}$$

for every $l<\infty$. Following the steps in previous two cases, it follows that $\lim\limits_{n\to\infty}\Delta_{k_0}(n) = 0$ and $\lim\limits_{n\to\infty}\Gamma_{RRT}^{\alpha}(k_0) = 0$. $\square$

1.5. Appendix F: Proof of Theorem 5

Statement of Theorem 5:- Suppose that the sample size $n\to\infty$ such that the matrix $\mathbf{X}$ satisfies $\delta_{k_0+1} < \frac{1}{\sqrt{k_0+1}}$, $\epsilon_\sigma \leq \epsilon_{omp}$ and $p_{lim}=0$. Then
a). OMP with *a priori* knowledge of k_0 or σ^2 is consistent, i.e.. $\lim\limits_{n\to\infty}\mathbb{P}(\hat{\mathcal{S}}=\mathcal{S}) = 1$.
b). RRT with hyper parameter α satisfying $\lim\limits_{n\to\infty}\alpha = 0$ and $\alpha_{lim}=0$ is consistent.

Proof. Statement a) of Theorem 5 follows directly from the bound $\mathbb{P}(\hat{\mathcal{S}}=\mathcal{S}) \geq 1-1/n$ in Lemma 1 of the article for OMP with k_0 iterations and SC $\|\mathbf{r}^k\|_2 \leq \epsilon^\sigma$ once $\epsilon_\sigma < \epsilon_{omp}$. Next we consider statement b) of Theorem 5. Following Theorem 3, we know that RRT support estimate satisfies $\mathbb{P}(\hat{\mathcal{S}} = \mathcal{S}) \geq 1 - 1/n - \alpha$ once $\epsilon_\sigma < \min(\epsilon_{omp}, \epsilon_{RRT})$. Hyper parameter α satisfying $\alpha_{lim} = 0$ implies that as $n \to \infty$, $\Gamma^\alpha_{RRT}(k_0) \to 1$ which in turn imply that $\min(\epsilon_{RRT}, \epsilon_{omp}) \to \epsilon_{omp}$. This along with $\alpha \to 0$ as $n \to \infty$ implies that RRT support estimate satisfies $\lim\limits_{n\to\infty} \mathbb{P}(\hat{\mathcal{S}} = \mathcal{S}) = 1$ once $\epsilon_\sigma < \epsilon_{omp}$. □

1.6. Appendix G: Proof of Theorem 6

Statement of Theorem 6:- Let $k_{max} > k_0$ and the matrix $\mathbf{X}$ satisfies $\delta_{k_0+1} < \dfrac{1}{\sqrt{k_0+1}}$. Then,
a). $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}) = 0$.
b). $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{E}) = \lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{F}) \leq \alpha$.

Proof. Note that the RRT support estimate is given by $\hat{\mathcal{S}} = \mathcal{S}_{omp}^{k_{RRT}}$. Consider the three events missed discovery $\mathcal{M} = card(\mathcal{S}/\mathcal{S}_{omp}^{k_{RRT}}) > 0$, false discovery $\mathcal{F} = card(\mathcal{S}_{omp}^{k_{RRT}}/\mathcal{S}) > 0$ and error $\mathcal{E} = \{\mathcal{S}_{omp}^{k_{RRT}} \neq \mathcal{S}\}$ separately.

$\mathcal{M} = card(\mathcal{S}/\mathcal{S}_{omp}^{k_{RRT}}) > 0$ occurs if any of these events occurs.
a).$\mathcal{M}_1 : k_{min} = \infty$: then any support in the support sequence produced by OMP suffers from missed discovery.
b).$\mathcal{M}_2 : k_{min} \leq k_{max}$ but $k_{RRT} < k_{min}$: then the RRT estimate misses atleast one entry in $\mathcal{S}$.
Since these two events are disjoint, it follows that $\mathbb{P}(\mathcal{M}) = \mathbb{P}(\mathcal{M}_1) + \mathbb{P}(\mathcal{M}_2)$. By Lemma 1, it is true that $k_{min} = k_0 \leq k_{max}$ whenever $\|\mathbf{w}\|_2 \leq \epsilon_{omp}$. Note that

$$\mathbb{P}(\mathcal{M}_1^C) \geq \mathbb{P}(k_{min} = k_0) \geq \mathbb{P}(\|\mathbf{w}\|_2 \leq \epsilon_{omp}). \quad (26)$$

Since $\|\mathbf{w}\|_2 \overset{P}{\to} 0$ as $\sigma^2 \to 0$, it follows that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\|\mathbf{w}\|_2 < \epsilon_{omp}) = 1$ and $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}_1^C) = 1$. This implies that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}_1) = 0$. Next we consider the event $\mathcal{M}_2$, i.e., $\{k_{min} \leq k_{max} \& k_{RRT} < k_{min}\}$. Using the law of total probability we have

$$\begin{aligned}\mathbb{P}(\{k_{min} \leq k_{max} \& k_{RRT} < k_{min}\}) = \mathbb{P}(k_{min} \leq k_{max}) \\ -\mathbb{P}(\{k_{min} \leq k_{max} \& k_{RRT} \geq k_{min}\})\end{aligned} \quad (27)$$

Following Lemma 1 we have $\mathbb{P}(k_{min} \leq k_{max}) \geq \mathbb{P}(k_{min} = k_0) \geq \mathbb{P}(\|\mathbf{w}\|_2 \leq \epsilon_{omp})$. This implies that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(k_{min} \leq k_{max}) = 1$. Following the proof of Theorem 3, we know that both $k_{min} = k_0$ and $RR(k_0) < \Gamma^\alpha_{RRT}(k_0)$ once $\|\mathbf{w}\|_2 \leq \min(\epsilon_{omp}, \epsilon_{RRT})$. Hence,

$$\begin{aligned}\mathbb{P}(\{k_{min} \leq k_{max} \& k_{RRT} \geq k_{min}\}) \\ \geq \mathbb{P}(\|\mathbf{w}\|_2 \leq \min(\epsilon_{omp}, \epsilon_{RRT}))\end{aligned} \quad (28)$$

which implies that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\{k_{min} \leq k_{max} \& k_{RRT} \geq k_{min}\}) = 1$. Applying these two limits in (27) give $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}_2) = 1$. Since $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}_1) = 0$ and $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}_2) = 0$, it follows that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}) = 0$.

Following the proof of Theorem 3, one can see that the event $\mathcal{E}^C = \{\hat{\mathcal{S}} = \mathcal{S}\}$ occurs once three events $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ occurs simultaneously, i.e., $\mathbb{P}(\mathcal{E}^C) \geq \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3)$. Of these three events, $\mathcal{A}_1 \cap \mathcal{A}_2$ occur once $\|\mathbf{w}\|_2 \leq \min(\epsilon_{omp}, \epsilon_{RRT})$. This implies that

$$\lim_{\sigma^2\to 0} \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) \geq \lim_{\sigma^2\to 0} \mathbb{P}(\|\mathbf{w}\|_2 \leq \min(\epsilon_{omp}, \epsilon_{RRT})) = 1. \quad (29)$$

At the same time $\mathbb{P}(\mathcal{A}_3) \geq 1 - \alpha, \forall \sigma^2 > 0$. Hence, it follows that

$$\lim_{\sigma^2\to 0} \mathbb{P}(\mathcal{E}^C) = \lim_{\sigma^2\to 0} \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) \geq 1 - \alpha \quad (30)$$

which in turn implies that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{E}) \leq \alpha$. Since $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{M}) + \mathbb{P}(\mathcal{F})$ and $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{M}) = 0$, it follows that $\lim\limits_{\sigma^2\to 0} \mathbb{P}(\mathcal{F}) \leq \alpha$. □

2. Numerical validation of Theorems

2.1. Numerically validating Theorems 1 and 2

In this section, we numerically validate the results in Theorem 1 and Theorem 2. The experiment setting is as follows. We consider a design matrix $\mathbf{X} = [\mathbf{I}_n, \mathbf{H}_n]$, where $\mathbf{H}_n$ is a $n \times n$ Hadamard matrix. This matrix is known to satisfy $\mu_{\mathbf{X}} = \dfrac{1}{\sqrt{n}}$. Hence, OMP can recover support exactly (i.e., $k_{min} = k_0$ and $\mathcal{S}_{omp}^{k_0} = \mathcal{S}$) at high SNR once $k_0 \leq \dfrac{1}{2}(1 + \dfrac{1}{\mu_{\mathbf{X}}}) = \dfrac{1}{2}(1 + \sqrt{n})$. In our simulations, we set $n = 32$ and $k_0 = 3$ which satisfies $k_0 \leq \dfrac{1}{2}(1 + \sqrt{n})$. The noise $\mathbf{w}$ is sampled according to $\mathcal{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$ with $\sigma^2 = 1$. The non zero entries of $\boldsymbol{\beta}$ are set at $\pm a$, where a is set to achieve the required value of $SNR = \dfrac{\|\mathbf{X}\boldsymbol{\beta}\|_2^2}{n}$.

In Fig.1, we plot values taken by $RR(k_{min})$ in 1000 runs of OMP. The maximum iterations k_{max} is set at $[0.5(n+1)]$. Recall that k_{min} is itself a random variable taking values in $\{k_0, \ldots, k_{max}, \infty\}$. As one can see from Fig.1, the values of k_{min} are spread out in the set $\{k_0, \ldots, k_{max}, \infty\}$ when SNR=1. Further, the values taken by $RR(k_{min})$ are close to one. However, with increasing SNR, the range of values taken by k_{min} concentrates around $k_0 = 3$. This validates the statement b) of Theorem 1, *viz.* $\lim\limits_{SNR\to\infty} \mathbb{P}(k_{min} = k_0) = 1$. Further, one can also see that the values taken by $RR(k_{min})$ decreases with increasing SNR. This validates the statement $RR(k_{min}) \overset{P}{\to} 0$ as $SNR \to \infty$.

Next we consider the behaviour of $RR(k)$ for $k > k_{min}$. From Fig.2, it is clear that the range of values taken by $RR(k)$ for $k > k_{min}$ is invariant w.r.t to the SNR. Indeed, the density of points near k_0 at SNR=1 is lower than that of SNR=10. This because of the fact that the k_{min} becomes more concentrated around k_0 with increasing SNR. Further, one can see that bulk of the values taken by $RR(k)$ for $k > k_{min}$ are above the deterministic curves $\Gamma^{\alpha}_{RRT}(k)$. This agrees with the $\mathbb{P}(RR(k) > \Gamma^{\alpha}_{RRT}(k)) \geq 1-\alpha$ for all $\sigma^2 > 0$ bound derived in Theorem 2.

2.2. Numerically validating Theorem 4

We next numerically validate the asymptotic behaviour of $\Gamma^{\alpha}_{RRT}(k_0)$ predicted by Theorem 4. In Fig.3, we plot the variations of $\Gamma^{\alpha}_{RRT}(k_0)$ for different choices of α and different sampling regimes. The quantities in the boxes inside the figures represent the values of α. All choices of α satisfy $\alpha_{lim} = 0$. Among the four sample regimes considered, three sampling regimes satisfies $p_{lim} = 0$, whereas, the fourth sampling regime with $n = 2k_0\log(p)$ and $k_0 = 10$ has $0 < p_{lim} < \infty$. As predicted by Theorem 4, all the three regimes with $p_{lim} = 0$ have $\Gamma^{\alpha}_{RRT}(k_0)$ converging to one with increasing n. However, when $p_{lim} > 0$, one can see from the right-bottom figure in Fig.3 that $\Gamma^{\alpha}_{RRT}(k_0)$ converges to a value smaller than one. For this particular sampling regime one has $p_{lim} = 1/20$ and $k_{lim} = 0$. The convergent value is in agreement with the value $\exp(-\frac{p_{lim}}{1-k_{lim}}) = 0.9512$ predicted by Theorem 4.

3. Numerical simulations

3.1. Details on the real life data sets

In this section, we provide brief descriptions on the four real life data sets, *viz*., Brownlee's Stack loss data set, Star data set, Brain and body weight data set and the AR2000 dataset used in the article.

Stack loss data set contains $n = 21$ observations and three predictors plus an intercept term. This data set deals with the operation of a plant that convert ammonia to nitric acid. Extensive previous studies(Rousseeuw & Leroy, 2005; Jin & Rao, 2010) reported that observations $\{1, 3, 4, 21\}$ are potential outliers.

Star data set explore the relationship between the intensity of a star (response) and its surface temperature (predictor) for 47 stars in the star cluster CYG OB1 after taking a log-log transformation(Rousseeuw & Leroy, 2005). It is well known that 43 of these 47 stars belong to one group, whereas, four stars *viz.* 11, 20, 30 and 34 belong to another group. Aforementioned observations are outliers can be easily seen from scatter plot itself. Please see Figure 4.

Brain body weight data set explores the interesting hypothesis that body weight (predictor) is positively correlated with brain weight (response) using the data available for 27 land animals(Rousseeuw & Leroy, 2005). Scatter plot after log-log transformation itself reveals three extreme outliers, *viz.* observations 6, 16 and 25 corresponding to three Dinosaurs (big body and small brains). However, extensive studies reported in literature also claims the presence of three more outliers, *viz.* 1 (Mountain Beaver), 14 (Human) and 17 (Rhesus monkey). These animals have smaller body sizes and disproportionately large brains. Please see Figure 4.

AR2000 is an artificial data set discussed in TABLE A.2 of (Atkinson & Riani, 2012). It has $n = 60$ observations and $p = 3$ predictors. Using extensive graphical analysis, it was shown in (Atkinson & Riani, 2012) that observations $\{9, 21, 30, 31, 38, 47\}$ are outliers.

3.2. More simulations on synthetic data sets

In this section, we provide some more simulation results demonstrating the superior performance of the proposed RRT algorithm. Reiterating," OMP1" represents the performance of OMP running exactly k_0 iterations, "OMP2" represents the performance of OMP with stopping rule $\|\mathbf{r}^k\|_2 \leq \sigma\sqrt{n + 2\sqrt{n\log(n)}}$, "CV" represents the performance of OMP with sparsity parameter k_0 estimated using five fold cross validation, "RRT1'" represents RRT with $\alpha = 1/\log(n)$, "RRT2" represents RRT with $\alpha = 1/\sqrt{n}$ and "LAT" represents the recently proposed least squares adaptive thresholding algorithm. The non zero entries in $\boldsymbol{\beta}$ are fixed at $\pm a$ where a is selected to achieve a specific SNR. The support $\mathcal{S}$ is sampled randomly from the set $\{1, 2, \dots, p\}$. The noise is Gaussian with zero mean and variance one. We consider three models for the matrix $\mathbf{X}$.

Model 1:- Model 1 has $\mathbf{X}$ formed by the concatenation of $n \times n$ identity and $n \times n$ Hadamard matrices. This matrix allows exact support recovery at high SNR once $k_0 \leq [\frac{1+\sqrt{n}}{2}]$. We set $n = 32$ and $k_0 = 3$.
Model 2:- Model 2 has entries of $\mathbf{X}$ sampled independently from a $\mathcal{N}(0, 1)$ distribution. This matrix allows exact support recovery at high SNR with a reasonably good probability once $k_0 = O(n/\log(p))$. We set $n = 32$, $p = 64$ and $k_0 = 3$.
Model 3:- Model 3 has rows of matrix $\mathbf{X}$ sampled independently from a $\mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$ distribution with $\mathbf{\Sigma} = (1-\kappa)\mathbf{I}_n + \kappa\mathbf{1}_n\mathbf{1}_n^T$. Here $\mathbf{1}_n$ is a $n \times 1$ vector of all ones. For $\kappa = 0$, this model is same as model 2. However, larger values of κ results in $\mathbf{X}$ having highly correlated columns. Such a matrix is not conducive for sparse recovery. We set $n = 32$, $p = 64$, $k_0 = 3$ and $\kappa = 0.7$.
Please note that all the matrices are subsequently normalised to have unit l_2 norm. Algorithms are evaluated in terms of mean squared error $MSE = \mathbb{E}(\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2^2)$ and support recovery error $PE = \mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S})$. All the results are presented

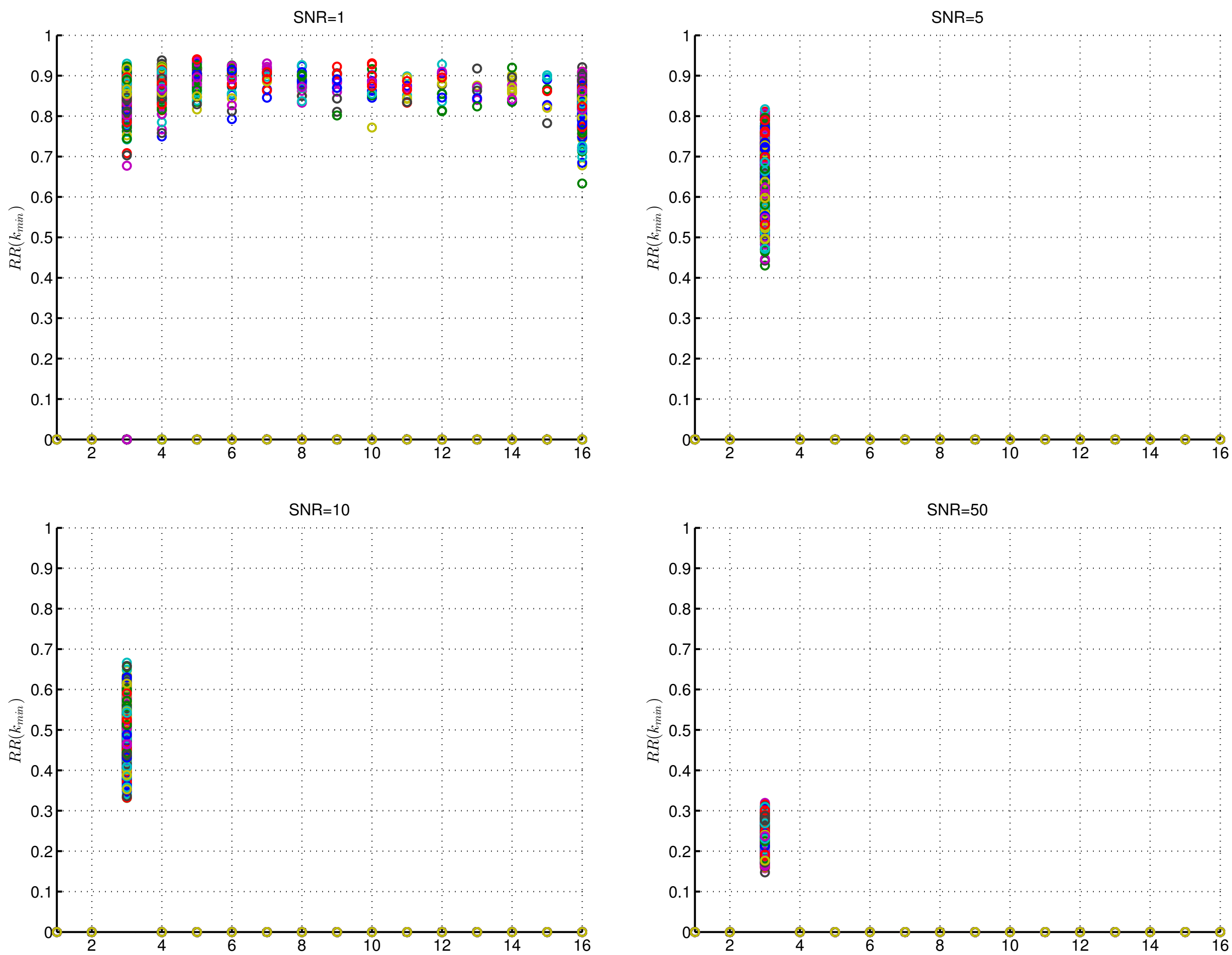

Figure 1. Validating Theorem 1: Evolution of $RR(k_{min})$ with increasing SNR. $k_{min} = k_0$ $368/1000$ times when SNR=1 and $1000/1000$ times for SNR=5, SNR=10 and SNR=50. $RR(k)$ for $k \neq k_{min}$ are set to zero for clarity.

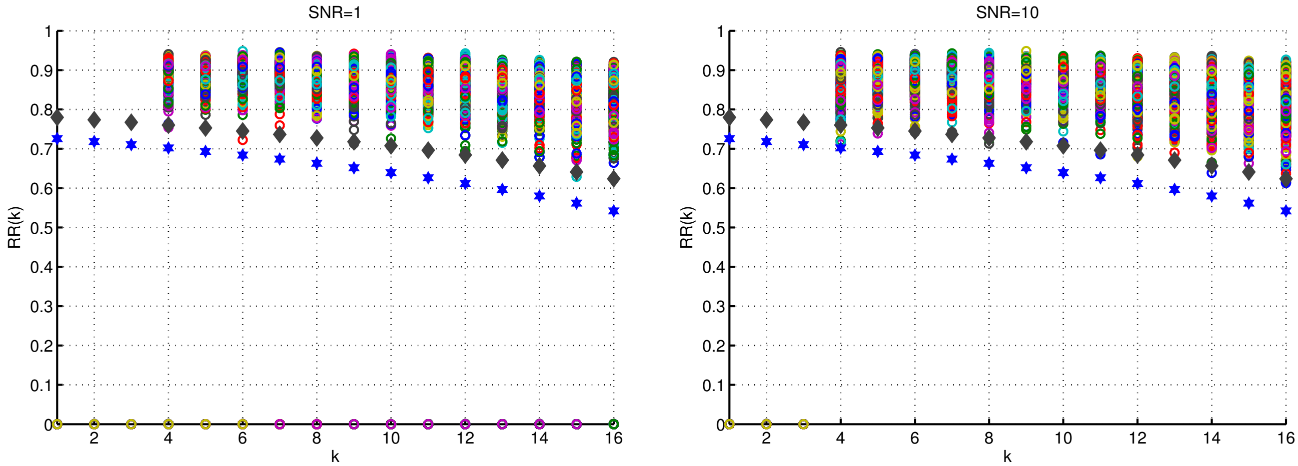

Figure 2. Validating Theorem 2: Evolution of $RR(k)$ for $k > k_{min}$ with increasing SNR. Circles are $RR(k)$ for $k > k_{min}$. Diamonds for Γ_{RRT}^{α} for $\alpha = 0.1$ and hexagons for $\alpha = 0.01$. $RR(k)$ for $k \leq k_{min}$ are set to zero for clarity.

after 10^3 iterations.

Figure 5 presents the performance of algorithms in matrix model 1. The best MSE and PE performance is achieved by OMP with *a priori* knowledge of k_0, i.e., OMP1. RRT1, RRT2 and OMP with *a priori* knowledge of σ^2 (i.e., OMP2) perform very similar to each other at all SNR in terms of MSE. Further, RRT1, RRT2 and OMP2 closely matches the MSE performance of OMP1 with increasing SNR. Please note that PE of RRT1 and RRT2 exhibits flooring at high SNR. The high SNR PE values of RRT1 and RRT2 are smaller than $\alpha = 1/\log(n) = 0.2885$ and $\alpha = 1/\sqrt(n) = 0.1768$ as predicted by Theorem 6. Further, RRT1 and RRT2 significantly outperform both CV and LAT at all SNR in terms of MSE and PE.

Figure 6 presents the performance of algorithms in matrix model 2. Here also OMP1 achieves the best performance. The MSE and PE performances of RRT1 and RRT2 are very close to that of OMP1. Also note that the performance gap between RRT1 and RRT2 versus LAT and CV diminishes in model 2 compared with model 1. Compared to model 1, model 2 is less conducive for sparse recovery and this is reflected in the relatively poor performance of all algorithms in model 2 compared with that of model 1.

Figure 7 presents the performance of algorithms in matrix model 3. As noted earlier, $\mathbf{X}$ in model 3 have highly coherent columns resulting in a very poor performance by all algorithms under consideration. Even in this highly non conducive environment, RRT1 and RRT2 delivered performances comparable or better compared to other algorithms under consideration.

To summarize, like the simulation results presented in the article, RRT1 and RRT2 delivered a performance very similar to the performance of OMP1 and OMP2. Please note that OMP1 and OMP2 are not practical in the sense that k_0 and σ^2 are rarely available *a priori*. Hence, RRT can be used as a signal and noise statistics oblivious substitute for OMP1 and OMP2. In many existing applications, CV is widely used to set OMP parameters. Note that RRT outperforms CV while employing only a fraction of computational effort required by CV.

References

Askitis, Dimitris. Asymptotic expansions of the inverse of the beta distribution. *arXiv preprint arXiv:1611.03573*, 2016.

Atkinson, Anthony and Riani, Marco. *Robust diagnostic regression analysis*. Springer Science & Business Media, 2012.

Jin, Y. and Rao, B. D. Algorithms for robust linear regression by exploiting the connection to sparse signal recovery. In *Proc. ICAASP*, pp. 3830–3833, March 2010. doi: 10.1109/ICASSP.2010.5495826.

Ravishanker, Nalini and Dey, Dipak K. *A first course in linear model theory*. CRC Press, 2001.

Rousseeuw, Peter J and Leroy, Annick M. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.

Wen, J., Zhou, Z., Wang, J., Tang, X., and Mo, Q. A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit. In *Proc. ISIT*, pp. 2364–2368, July 2016.

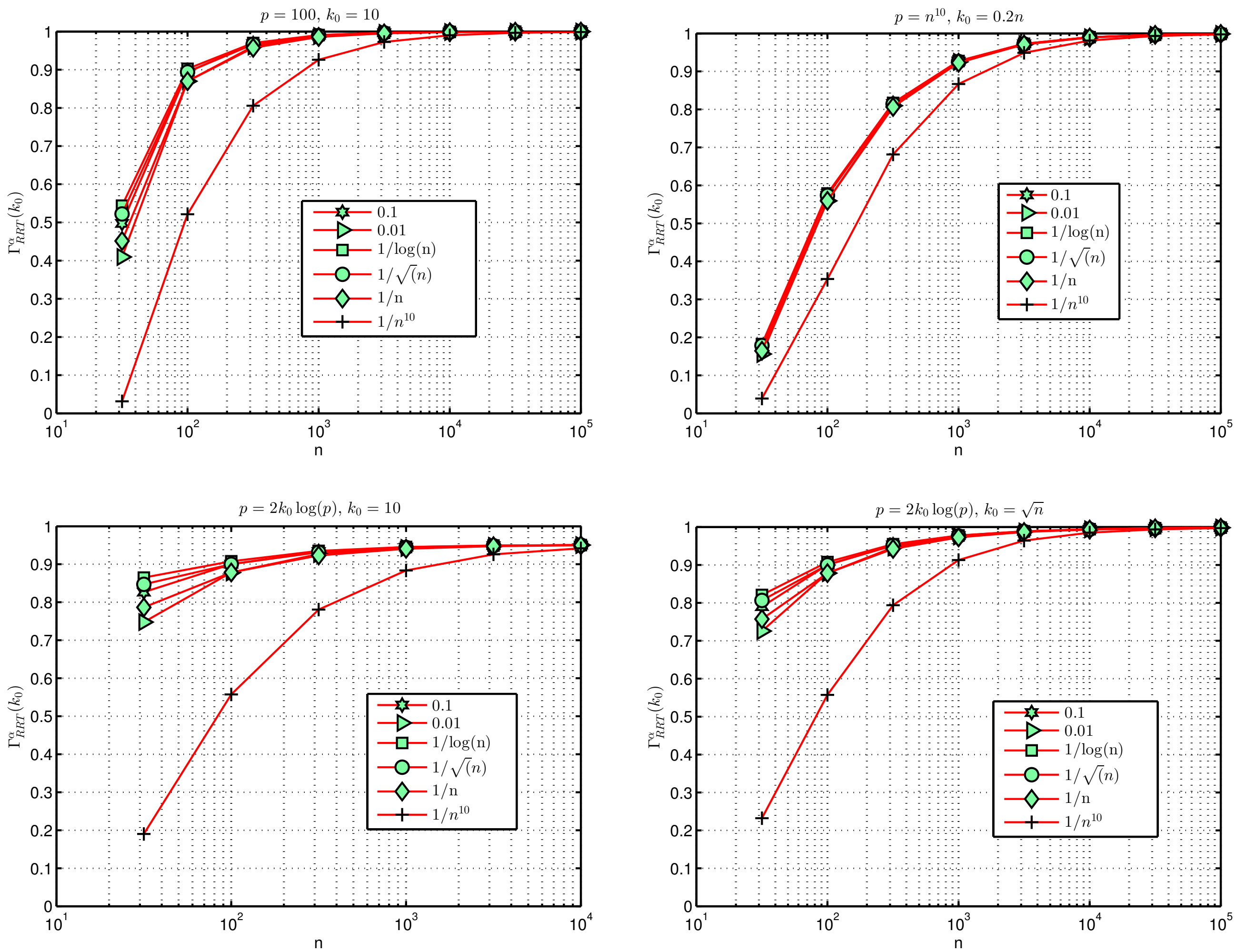

Figure 3. Validating Theorem 4. (Reading clockwise) i). plot the variations of $\Gamma^{\alpha}_{RRT}(k_0)$ when $n \to \infty$ and (p, k_0) are fixed at $(100, 10)$. ii). plot the variations of $\Gamma^{\alpha}_{RRT}(k_0)$ when $(n, p, k_0) \to (\infty, \infty, \infty)$ such that p increases polynomially with n, i.e., $p = n^{10}$ and $k_0 = 0.2n \to \infty$ increases linearly in n. iii). plot the variations of $\Gamma^{\alpha}_{RRT}(k_0)$ when $n \to \infty$, $k_0 = \sqrt{n} \to \infty$ sub linear in n and $p \to \infty$ as $p = 2k_0 \log(p)$. p is sub exponentially increasing w.r.t n in this case. iv). plot the variations of $\Gamma^{\alpha}_{RRT}(k_0)$ when $(n, p) \to (\infty, \infty)$ such that $k_0 = 10$ fixed and $p = 2k_0 \log(p)$. p is exponentially increasing w.r.t n in this case.

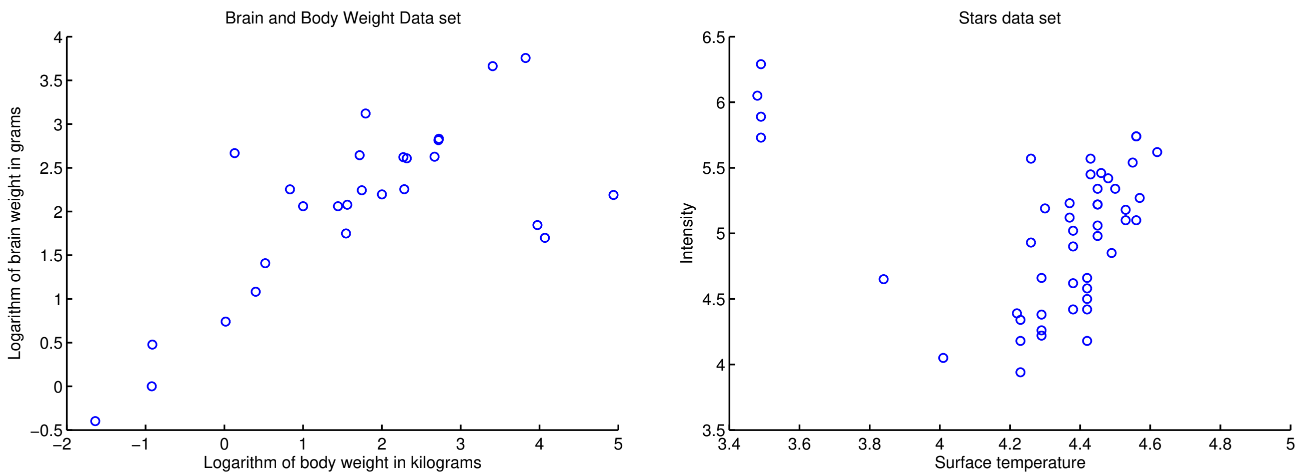

Figure 4. Scatter plots of Brain and body weight data set (left) and stars data set (right).

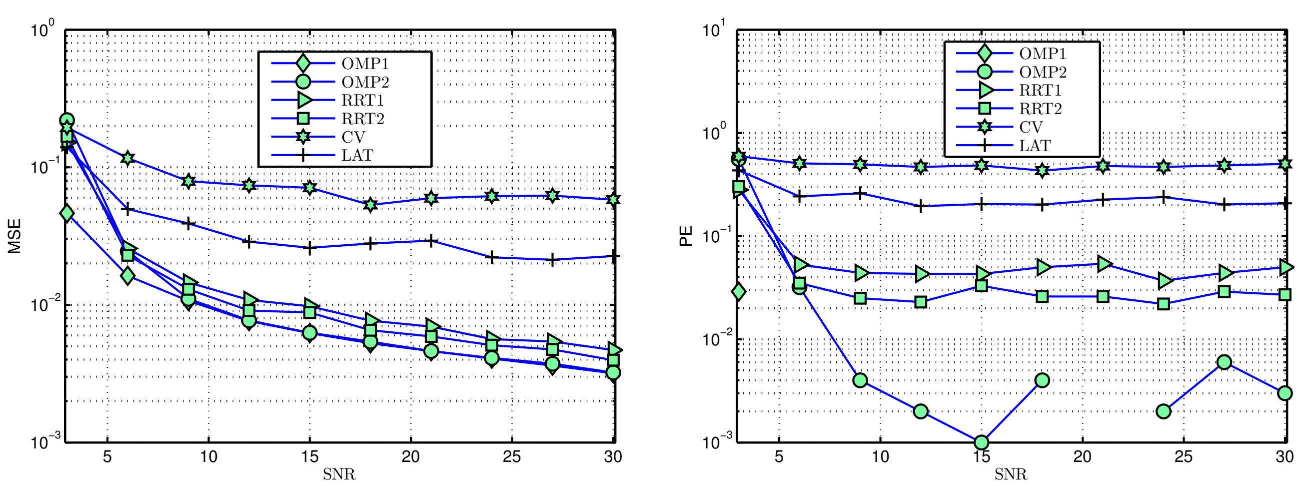

Figure 5. MSE and PE performances in matrix model 1.

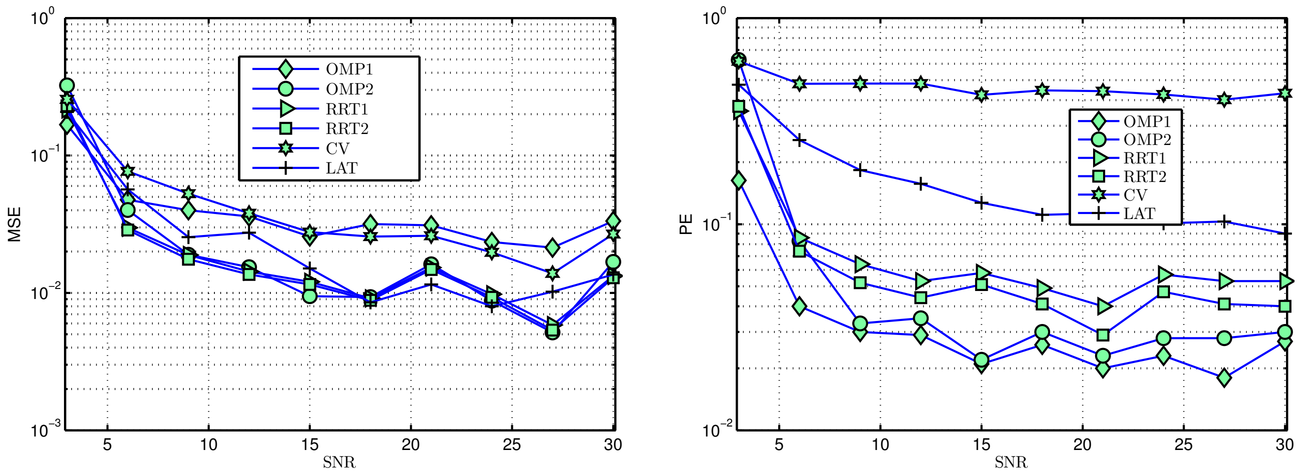

Figure 6. MSE and PE performances in matrix model 2.

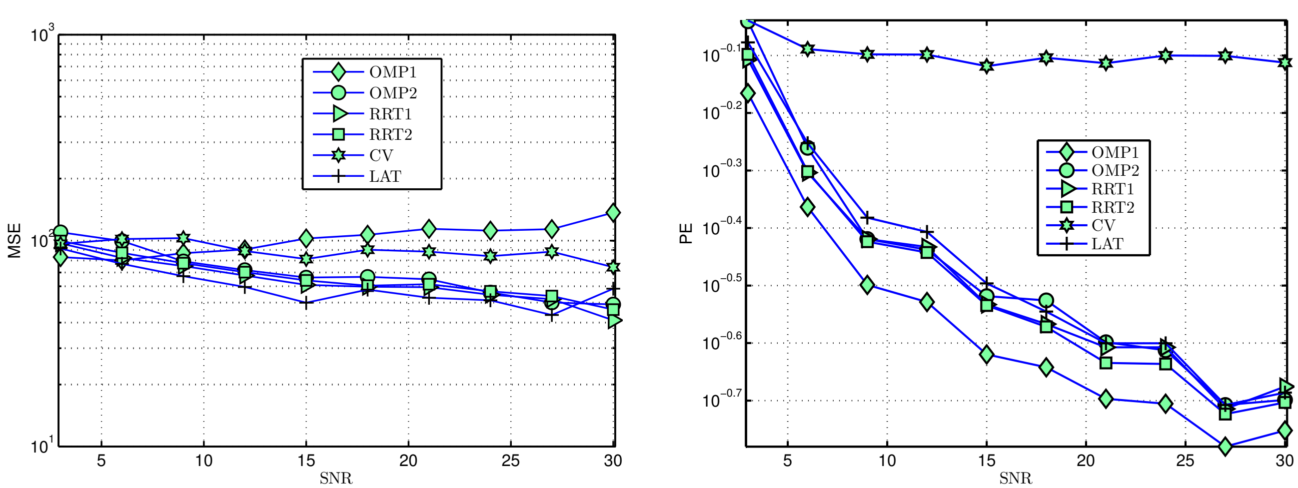

Figure 7. MSE and PE performances in matrix model 3.