MITIGATING SEVERE ILL-CONDITIONING IN SPARSE SIGNAL RECOVERY VIA RANDOM GAUSSIAN ROTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recovering a sparse signal from an underdetermined linear system, known as compressed sensing (CS), is central to many signal processing and machine learning applications. Low inter-column coherence of the sensing matrix is a standard condition for identifiable recovery. In several practical settings, such as magnetic resonance imaging (MRI) reconstruction and gene-based disease classification, the sensing matrix can be very fat and contain many nearly proportional columns, making direct recovery fragile. This work studies CS under such highly ill-conditioned sensing matrices. We propose a diagonal-dominant Gaussian random rotator (DogRot) as a plug-and-play right-side preconditioner. DogRot mixes sensing columns to reduce mutual coherence and uses the mixed problem to obtain a more reliable support ordering. The final amplitudes are then refitted on the original sensing matrix. We analyze invertibility, geometry-dependent coherence reduction, and exact-inverse dominant-support preservation, and evaluate the method on MRI, localized statistical channel modeling, and gene-based disease classification. The experiments demonstrate improved support localization and substantially reduced coefficient-amplitude errors with classical and modern subset-selection solvers, and identify the trade-off between decorrelation and excessive mixing.

1 INTRODUCTION

Sparse signal recovery, or compressed sensing (CS) (Donoho, 2006), has long been used in signal processing and machine learning. Given the observation $\boldsymbol{y} \in \mathbb{R}^p$ and a known fat sensing matrix $\boldsymbol{A} \in \mathbb{R}^{p\times n}$, CS seeks to solve the following *underdetermined* system (Donoho, 2006; Candes et al., 2006):

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{n} \in \mathbb{R}^p, p < n, \tag{1}$$

where $\boldsymbol{x}$ is the target *sparse* signal, and $\boldsymbol{n}$ denotes measurement noise. We write its support as $\mathcal{S}$, with sparsity level $s = |\mathcal{S}| \ll n$. Formulation (1) appears in real-time visual tracking (Zhang et al., 2012), remote sensing (Herman & Strohmer, 2009), accelerated magnetic resonance imaging (MRI) (Lustig et al., 2007), and sparse channel estimation (Gao et al., 2015), where the number of samples is much smaller than the dimension of the underlying signal.

A standard sufficient condition for stable CS recovery is the so-called *restricted isometry property* (RIP) (Candes & Tao, 2005; Candes, 2008; Candes & Wakin, 2008). RIP requires $\boldsymbol{A}$ to approximately preserve the norms of sufficiently sparse vectors, under which the recovery problem in (1) can be reliably solved via iterative optimization methods (Tibshirani, 1996; Beck & Teboulle, 2009; Boyd et al., 2011), greedy pursuit algorithms (Pati et al., 1993; Needell & Vershynin, 2010; Needell & Tropp, 2009), or Bayesian learning approaches (Ji et al., 2008; Baron et al., 2010). However, verifying RIP is NP-hard (Tillmann & Pfetsch, 2014), where NP denotes nondeterministic polynomial time.

The *mutual coherence* (MC) of $\boldsymbol{A}$ provides tractable worst-case sufficient conditions for recovery in (1) (Elad, 2007; Donoho & Elad, 2003). The MC of the sensing matrix $\boldsymbol{A}$ is defined as $\mu(\boldsymbol{A}) = \max_{i\neq j} \left| \left[\hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}\right]_{i,j} \right|$, where $\hat{\boldsymbol{A}}$ denotes the column-normalized version of $\boldsymbol{A}$, with $\hat{\boldsymbol{A}}(:,i) = \boldsymbol{A}(:,i)/\|\boldsymbol{A}(:,i)\|_2$. Closely related to MC is the Gram matrix $\boldsymbol{G}_A = \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}}$, whose (i,j)th entry equals

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

$\hat{\boldsymbol{a}}_i^T \hat{\boldsymbol{a}}_j$ and quantifies the correlation between the ith and jth columns of $\boldsymbol{A}$, where $\hat{\boldsymbol{a}}_i$ denotes the ith column of $\hat{\boldsymbol{A}}$. The largest absolute off-diagonal element of $\boldsymbol{G}_A$ is exactly the MC $\mu(\boldsymbol{A})$. We use MC and the Gram matrix as the main conditioning metrics in the analysis and experiments.

While effective CS approaches under ideal RIP or low MC are extensive, this work focuses on a challenging family of (1) where $\boldsymbol{A}$ is a *highly ill-conditioned matrix* (HIM) and the RIP is severely violated. Such cases typically arise when $p \ll n$ and the columns of $\boldsymbol{A}$ are strongly correlated or nearly proportional. In *MRI reconstruction* (Lustig et al., 2007; Eksioglu, 2016; Ryu et al., 2019; Ding et al., 2023), sub-sampling makes the number of image pixels far exceed the acquired k-space samples and may introduce aliasing under common sampling patterns. In *localized statistical channel modeling* (LSCM) (Luo et al., 2023; Zhang et al., 2024; 2025), different beams can have similar responses, while the angular power spectrum (APS) dimension is much larger than the number of measured beams. In *gene-based disease classification* (GDC) (Golub et al., 1999; Zou & Hastie, 2005; Li & Li, 2008), thousands of genes are observed from only tens of patient samples.

We propose a preconditioning framework for such highly ill-conditioned CS problems. Our main contribution is introducing a right-side mixer $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, which rewrites (1) as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{x} + \boldsymbol{n}. \tag{2}$$

We instantiate $\boldsymbol{Q}$ as a *diagonal-dominant Gaussian random rotator* (DogRot). We establish invertibility, a geometry-dependent coherence-reduction bound, and a dominant-support guarantee for the exact inverse action. An explicit auxiliary-error condition connects this ordering to the output of a sparse solver. The final coefficients are reconstructed by refitting on the original sensing matrix. Extensive experiments verify that DogRot consistently reduces the MC of sensing matrices across MRI, LSCM, and GDC, and that the resulting screening and refitting procedures improve support localization and substantially reduce coefficient-amplitude errors in the tested HIM settings.

2 Background: Sparse Signal Recovery Under Highly Ill-Conditioned Sensing Matrices

In this section, we first characterize problem (1) under *highly ill-conditioned sensing matrix* (HIM), and then briefly review some of its close relatives.

2.1 Highly Ill-Conditioned Sensing Matrix

Definition 2.1 (Highly Ill-conditioned Matrix). We use HIM to describe sensing matrices with nearly collinear columns, characterized by the last two properties below and often accompanied by the first two:

- **High Aspect Ratio:** $p \ll n$, or equivalently $n/p \geq \alpha$ for some sufficiently large constant α. Typical values include $\alpha \approx 10$ in MRI reconstruction and $\alpha \approx 100$ in GDC.
- **Large-scale dimension:** $n > 10^3$, for example, $n = 6552$ in Zhang et al. (2024) and $n = 7129$ in Golub et al. (1999).
- **High MC:** $\mu(\boldsymbol{A}) \approx 1$, indicating strong collinearity among certain columns of $\boldsymbol{A}$.
- **Heavy-tailed Gram correlations:** The off-diagonal entries of the Gram matrix $\boldsymbol{G}_A = \hat{\boldsymbol{A}}^T\hat{\boldsymbol{A}}$ exhibit a heavy-tailed distribution near unity. Specifically, for some small $d > 0$, a non-negligible fraction of column pairs satisfy $|[\boldsymbol{G}_A]_{i,j}| > 1 - d$.

An illustration of a HIM is provided in Figure 4a in Appendix G. The large problem scale and limited observations make it difficult for conventional algorithms to distinguish between similar columns, such as sub-sampled Fourier bases, closely spaced beams, and genes with similar expression profiles. Consequently, recovering $\boldsymbol{x}$ from (1) is severely hindered by these HIM properties. In particular, a high aspect ratio confines many columns to a low-dimensional observation space and enlarges the feasible search space. Meanwhile, the latter two properties imply that a considerable fraction of columns are nearly indistinguishable in the observation space. In extreme cases, successful recovery is only guaranteed when $\boldsymbol{x}$ is 1-sparse (Elad, 2007), which is too restrictive for most applications. These observations highlight the detrimental impact of HIMs on sparse recovery.

2.2 Prior Work

Many approaches have been explored to enhance recovery performance by manipulating $\boldsymbol{A}$ to reduce $\mu(\boldsymbol{A})$ (Elad, 2007; Bruckstein et al., 2008; Abolghasemi et al., 2010; Li et al., 2017; Kilic et al., 2022; Wang et al., 2024). Among them, projection is the most straightforward and commonly adopted preconditioning strategy, as an invertible left transformation preserves the noiseless solution set while reshaping the sensing matrix. This is achieved by imposing an invertible left transformation $\boldsymbol{P}$ on $\boldsymbol{A}$ and reducing the MC of the effective observation matrix $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{A}$. Relaxation (Abolghasemi et al., 2010) and decomposition-based methods (Elad, 2007; Kilic et al., 2022) are two main branches for finding an effective projector. Relaxation-based methods aim to minimize a smooth surrogate of $\mu(\boldsymbol{D})$ because directly expressing MC as a tractable objective is difficult. Gradient-based methods, including gradient descent (GD), are often adopted to obtain a stationary solution. On the other hand, decomposition-based approaches exploit structured matrix factorizations, such as triangular decomposition and singular value decomposition (SVD), to improve the conditioning of $\boldsymbol{A}$ through designed transformations.

Although these approaches have shown theoretical and empirical success in various applications, their gains can be limited for HIM, as illustrated in Appendix G. Left multiplication changes the measurement-space metric and can alter column angles, but exact proportionality is preserved. For nearly proportional columns, substantial decorrelation may require strongly anisotropic weighting, which can amplify measurement noise. This motivates a complementary approach that changes the column representation before support screening. Modern subset selectors such as adaptive best subset selection (ABESS) (Zhu et al., 2020) and optimal pursuit–best subset selection (OP-BESS) (Zhu et al., 2025) improve the support search itself. We examine their combination with DogRot in Appendix G.3.2.

Although perfect recovery is favorable in general, in challenging cases with HIM it may be more realistic to first identify the active support and then estimate its coefficients. In gene-based disease classification (Golub et al., 1999), for example, selecting informative genes can be more important than estimating their individual contribution factors precisely. Motivated by this, our objective is to improve support identification under HIM, followed by coefficient refitting on the original sensing matrix.

3 DogRot Preconditioning for HIM Compressed Sensing

In this section, we introduce a new preconditioning paradigm designed to address HIM-induced instances of (1). The key idea, expressed as (2), is to apply a structured matrix $\boldsymbol{Q}$ that refines the conditioning of $\boldsymbol{A}$ while preserving the essential support structure of (1). In what follows, we provide the theoretical foundations and implementation details of this idea.

3.1 Column Mixing as Preconditioning

By definition, MC is a column-related metric, making column mixing a direct way to alter it. The heavy-tailed correlation distribution of HIM indicates that some highly correlated columns in $\boldsymbol{A}$ already contribute to the ill-conditioning, while many others may exhibit weak correlations. This observation motivates the idea of *mixing*, i.e., by combining highly correlated columns with other less correlated ones, the MC of $\boldsymbol{A}$ can be effectively reduced. Such a mixing operation can be realized through right multiplication with an invertible matrix, which we refer to as a *mixer* in this paper. In this way, the original problem (1) is reformulated as (2), where the new sensing matrix $\boldsymbol{A}\boldsymbol{Q}$ is expected to exhibit reduced MC, thereby enabling improved recovery.

To ensure recoverability of the original sparse signal, $\boldsymbol{Q}$ must be invertible. Since the new problem remains underdetermined, the transformed signal $\boldsymbol{Q}^{-1}\boldsymbol{x}$ should also preserve the support structure. Thus, an effective mixer $\boldsymbol{Q}$ should decorrelate the columns of $\boldsymbol{A}$ while preserving the support through its inverse.

Finding an appropriate mixer can be formulated as:

$$\min_{\boldsymbol{Q}} \quad \mu\left(\boldsymbol{A}\boldsymbol{Q}\right) \tag{3}$$

$$\text{s.t.} \quad \boldsymbol{Q}^{-1}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^{-1} = \boldsymbol{I}_n, \tag{3a}$$

$$\operatorname{supp}\left(\boldsymbol{Q}^{-1}\boldsymbol{x}\right) = \operatorname{supp}\left(\boldsymbol{x}\right) = \mathcal{S}, \tag{3b}$$

where $\operatorname{supp}\left(x\right) = \left\{i \,|x_i \neq 0\right\}$ denotes the support set of $\boldsymbol{x}$. The objective and support constraint make this a non-convex design problem. We use it as the ideal formulation and develop a computationally simple approximate approach below, rather than solving for an exact feasible optimizer.

3.2 Diagonal-Dominant Gaussian Random Rotator

Finding an optimizer for (3) is highly non-trivial. We instead structure a family of preconditioning matrices, termed the diagonal-dominant Gaussian random rotator (DogRot), that balances decorrelation with preservation of the signal's salient coordinates. Its definition is as follows:

Definition 3.1 (**Diagonal-Dominant Gaussian Random Rotator**). Given an integer $n \geq 2$ and a small positive parameter $\epsilon > 0$, the DogRot $\boldsymbol{Q}_{\epsilon,n} \in \mathbb{R}^{n\times n}$ is defined as:

$$\boldsymbol{Q}_{\epsilon,n} = \begin{bmatrix} 1 & -q_{2,1} & \cdots & -q_{n,1} \\ q_{2,1} & 1 & \cdots & -q_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n,1} & q_{n,2} & \cdots & 1 \end{bmatrix}, \tag{4}$$

where the lower-triangular entries $q_{i,j}, i > j$ are independently drawn from the Gaussian distribution $\mathcal{N}\left(0,\epsilon\right)$. The upper-triangular entries satisfy the skew-symmetry relation $q_{j,i} = -q_{i,j}$ for all $i > j$, while all diagonal elements are fixed to one.

By definition, DogRot intuitively encompasses three key properties:
• **Diagonal-dominant:** The diagonal entries are set to 1, while the off-diagonal entries are random variables with a small variance. The name refers to this entrywise scale separation, not strict row-sum diagonal dominance or spectral proximity to the identity;
• **Gaussian Random:** Each off-diagonal element below the main diagonal is independently drawn from a Gaussian distribution, introducing controlled randomness to the matrix;
• **Rotator:** DogRot is a combination of the identity matrix and a skew-symmetric matrix. This structure aligns with the Lie algebra $\mathfrak{so}(n)$, which consists of real skew-symmetric matrices and serves as the infinitesimal generator of the special orthogonal group $\mathrm{SO}(n)$. Consequently, right multiplication by a DogRot can be viewed as applying a small, random rotational perturbation to each column of the sensing matrix.

We next establish the properties supporting this approximate approach to (3).

3.2.1 Invertibility and Column Concentration

Write $\boldsymbol{Q}_{\epsilon,n} = \boldsymbol{I}_n + \boldsymbol{S}$, where $\boldsymbol{S}^T = -\boldsymbol{S}$. This decomposition gives an exact invertibility result.

Lemma 3.2. *Every realization of $\boldsymbol{Q}_{\epsilon,n}$ is invertible, and $\|\boldsymbol{Q}_{\epsilon,n}^{-1}\|_2 \leq 1$.*

Proof. Since $\boldsymbol{Q}_{\epsilon,n}^T\boldsymbol{Q}_{\epsilon,n} = \boldsymbol{I}_n + \boldsymbol{S}^T\boldsymbol{S} \succeq \boldsymbol{I}_n$, every singular value is at least one. □

The distribution of $\boldsymbol{Q}_{\epsilon,n}^T$ is the same as that of $\boldsymbol{Q}_{\epsilon,n}$, because the independent Gaussian entries of $\boldsymbol{S}$ are symmetric about zero. In what follows, let $m_n = 1 + \epsilon(n-1)$.

Theorem 3.3 (**Norm concentration**). *For a column $\boldsymbol{q}_i$ of $\boldsymbol{Q}_{\epsilon,n}$ and $0 < t < 1$,*

$$\Pr\left\{\left|\frac{\|\boldsymbol{q}_i\|_2^2}{m_n} - 1\right| \geq t\right\} \leq 2\exp\left(-\frac{(n-1)t^2}{8}\right). \tag{5}$$

The identity $\|\boldsymbol{q}_i\|_2^2 = 1 + \epsilon\chi_{n-1}^2$ yields this bound (Appendix A). For distinct columns, skew symmetry cancels the two linear terms in their inner product:

$$\boldsymbol{q}_i^T\boldsymbol{q}_j = \sum_{k\neq i,j} S_{ki}S_{kj}. \tag{6}$$

The summands in (6) are independent for a fixed pair (i, j), although the columns themselves are dependent.

Theorem 3.4 (**Column inner-product concentration**). *For $n > 2$ and $t > 0$, there is a universal constant $c_0 > 0$ such that*

$$\Pr\{|\boldsymbol{q}_i^T \boldsymbol{q}_j| \geq t\} \leq 2\exp\left[-c_0 \min\left\{\frac{t^2}{(n-2)\epsilon^2}, \frac{t}{\epsilon}\right\}\right], \qquad i \neq j. \tag{7}$$

Combining the two bounds over all entries gives the following result, proved in Appendix A.2.

Corollary 3.5 (**Entrywise scaled Gram concentration**). *For any positive sequence $\epsilon = \epsilon_n$,*

$$\left\|m_n^{-1}\boldsymbol{Q}_{\epsilon,n}^T\boldsymbol{Q}_{\epsilon,n} - \boldsymbol{I}_n\right\|_{\max} \xrightarrow{P} 0, \qquad n \to \infty, \tag{8}$$

where $\|\boldsymbol{D}\|_{\max} = \max_{i,j}|D_{ij}|$.

This is an entrywise statement, rather than spectral-norm convergence. The inverse action used for support recovery is analyzed directly below.

3.2.2 Reduction in Mutual Coherence

Let $\boldsymbol{A}$ be fixed independently of the mixer and have unit-norm columns, $\boldsymbol{G}_A = \boldsymbol{A}^T\boldsymbol{A}$, and $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{Q}_{\epsilon,n}$. Expanding the Gram entries in the independent lower-triangular Gaussian variables gives

$$\mathbb{E}[\boldsymbol{b}_i^T\boldsymbol{b}_j] = (1-\epsilon)[\boldsymbol{G}_A]_{ij}\ (i \neq j), \qquad \mathbb{E}\|\boldsymbol{b}_i\|_2^2 = m_n. \tag{9}$$

Each mixed column combines its original component with random contributions from the remaining sensing columns. Equation (9) shows that these contributions increase the expected column energy while shrinking the expected cross term. This separates the two quantities that determine normalized coherence. Their ratio suggests a decorrelation scale of $|1-\epsilon|/m_n$, but the realized reduction also depends on fluctuations in both quantities. The next theorem controls them jointly through the geometry of the original Gram matrix.

Theorem 3.6 (**Geometry-dependent coherence reduction**). *For $0 < \delta < 1$, let R_{ij} be the deterministic Gaussian fluctuation radii defined in* (20) *of Appendix B. If $R_{ii} < m_n$ for every i, then, with probability at least $1-\delta$,*

$$\mu(\boldsymbol{B}) \leq \max_{i<j} \frac{|(1-\epsilon)[\boldsymbol{G}_A]_{ij}| + R_{ij}}{\sqrt{(m_n - R_{ii})(m_n - R_{jj})}} =: U_\delta(\boldsymbol{A}, \epsilon). \tag{10}$$

In particular, $U_\delta(\boldsymbol{A}, \epsilon) < \mu(\boldsymbol{A})$ is sufficient for strict MC reduction with the same probability.

The radii depend on the geometry of $\boldsymbol{G}_A$, accounting for correlations shared by different column pairs. When they are small relative to m_n, the bound reflects a contraction close to $|1-\epsilon|/m_n$. This condition also distinguishes severe coherence from complete geometric collapse: for a rank-one sensing matrix, all nonzero mixed columns remain collinear and MC cannot decrease (Appendix B).

3.2.3 The Essence of DogRot: Preserving Dominant Support

Column mixing changes the representation of the inverse problem without changing its measurements. The exact mixed system is

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{z} + \boldsymbol{n}, \qquad \boldsymbol{z} = \boldsymbol{Q}_{\epsilon,n}^{-1}\boldsymbol{x}. \tag{11}$$

Although $\boldsymbol{z}$ is generally dense, support screening depends on the ordering of its coordinates rather than exact zeros. DogRot spreads the off-support leakage over many coordinates, allowing the active entries to remain distinguishable. The following result quantifies this effect without approximating the inverse.

Theorem 3.7 (**Exact-inverse dominant-support preservation**). *Let $n \geq 3$ and let $\boldsymbol{x} \neq \boldsymbol{0}$ be a fixed s-sparse vector, $1 \leq s < n$, independent of $\boldsymbol{Q}_{\epsilon,n}$, with support $\mathcal{S}$ and $x_{\min} = \min_{j\in\mathcal{S}}|x_j|$. For $0 < \delta < 1$, define*

$$\Gamma_\delta = \epsilon\left[(n-1) + 2\sqrt{(n-1)\log(2/\delta)} + 2\log(2/\delta)\right], r_\delta = \sqrt{\frac{2\log(4n/\delta)}{n-2}}. \tag{12}$$

If $x_{\min} > 2\|\boldsymbol{x}\|_2\sqrt{\Gamma_\delta}\, r_\delta$, then $\mathrm{Top}_s(|\boldsymbol{Q}_{\epsilon,n}^{-1}\boldsymbol{x}|) = \mathcal{S}$ with probability at least $1-\delta$.

Algorithm 1 Sparse signal recovery with HIM

Input Observations $\boldsymbol{y}$, HIM $\boldsymbol{A}$, DogRot variance ϵ, sparsity level s if known, otherwise path length $k_{\max}$;
Output Estimated sparse signal $\hat{\boldsymbol{x}}$;
1: Generate a DogRot matrix $\boldsymbol{Q}_{\epsilon,n}$;
2: Normalize the columns of $\boldsymbol{A}$ to obtain $\hat{\boldsymbol{A}}$ and compute the scaling matrix $\boldsymbol{W}$;
3: Form the mixed sensing matrix $\hat{\boldsymbol{A}}\boldsymbol{Q}_{\epsilon,n}$;
4: Apply the chosen auxiliary recovery algorithm to $\boldsymbol{y}$ and $\hat{\boldsymbol{A}}\boldsymbol{Q}_{\epsilon,n}$ to obtain an auxiliary estimate $\hat{\boldsymbol{z}}$;
5: **if** s is known **then**
6: Estimate the support by $\widehat{\mathcal{S}} = \mathrm{Top}_s(|\hat{\boldsymbol{z}}|)$;
7: **else**
8: Construct $\widehat{\mathcal{S}}_k = \mathrm{Top}_k(|\hat{\boldsymbol{z}}|)$ for $k = 1, \ldots, k_{\max}$, refit each candidate, and select $\widehat{\mathcal{S}}$ by the Bayesian information criterion (BIC);
9: **end if**
10: Refit the coefficients on the original sensing matrix: $\hat{\boldsymbol{x}}_{\widehat{\mathcal{S}}} \leftarrow \arg\min_{\boldsymbol{u}} \|\boldsymbol{y} - \boldsymbol{A}_{:,\widehat{\mathcal{S}}}\boldsymbol{u}\|_2^2$, and $\hat{\boldsymbol{x}}_{\widehat{\mathcal{S}}^c} \leftarrow \boldsymbol{0}$.

Here Top_s returns the indices of the s largest magnitudes. The proof in Appendix C uses the exact representation $\boldsymbol{Q}_{\epsilon,n}^{-1}\boldsymbol{x} = a\boldsymbol{x} + \|\boldsymbol{x}\|_2\sqrt{a(1-a)}\,\boldsymbol{u}$, where $a \in (0,1]$ and, conditionally on a, $\boldsymbol{u}$ is uniform on the unit sphere orthogonal to $\boldsymbol{x}$. Thus, coordinatewise leakage scales as $\sqrt{\log n/n}$ relative to the signal when $\epsilon(n-1) = O(1)$. For comparable nonzero amplitudes, $s = o(n/\log n)$ suffices at this constant mixing scale.

The common factor a changes the amplitude scale without changing the ordering of the active coordinates. The remaining component spreads across coordinates, so support preservation depends on its largest entry rather than its total energy. The separation condition compares this coordinatewise perturbation with the weakest active amplitude. It allows a dense auxiliary vector to retain the original dominant support even when its total off-support energy is not negligible.

The algorithm screens an estimate $\hat{\boldsymbol{z}}$ rather than the exact transformed vector. Its error enters the separation condition explicitly.

Corollary 3.8 (**Screening from an auxiliary estimate**). *Under Theorem 3.7, on the same probability-$1-\delta$ event, any estimate satisfying $\|\hat{\boldsymbol{z}} - \boldsymbol{Q}_{\epsilon,n}^{-1}\boldsymbol{x}\|_\infty \le \eta$ recovers $\mathcal{S}$ by top-s screening if*

$$x_{\min} > 2\|\boldsymbol{x}\|_2\sqrt{\Gamma_\delta}\, r_\delta + 2(1+\Gamma_\delta)\eta. \tag{13}$$

This guarantee is conditional on the auxiliary estimation error. A solver-specific bound on η is not assumed to follow from MC reduction alone. Once $\widehat{\mathcal{S}} = \mathrm{Top}_s(|\hat{\boldsymbol{z}}|)$ is selected, coefficients are refitted in the original system:

$$\hat{\boldsymbol{x}}_{\widehat{\mathcal{S}}} = \arg\min_{\boldsymbol{u}\in\mathbb{R}^{|\widehat{\mathcal{S}}|}} \|\boldsymbol{y} - \boldsymbol{A}_{:,\widehat{\mathcal{S}}}\boldsymbol{u}\|_2^2, \qquad \hat{\boldsymbol{x}}_{\widehat{\mathcal{S}}^c} = \boldsymbol{0}. \tag{14}$$

Proposition 3.9 (**Refit error after correct support selection**). *If $\widehat{\mathcal{S}} = \mathcal{S}$ and $\boldsymbol{A}_{:,\mathcal{S}}$ has full column rank, then $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \le \|\boldsymbol{A}_{:,\mathcal{S}}^\dagger\|_2\|\boldsymbol{n}\|_2$. In particular, noiseless refitting recovers the exact amplitudes.*

The proof is in Appendix D. This screen-and-refit formulation assigns distinct roles to the two representations. The auxiliary entries provide a ranking of candidate locations, not the final coefficient values. Once the support is selected, refitting uses the original column scales and measurement model. This is particularly relevant when nearly parallel sensing columns have different norms, since a useful support estimate need not provide an accurate amplitude estimate in the mixed coordinates.

The transformation retains the measurements and the exact solution relation in (11). Its benefit comes from changing the representation presented to the recovery algorithm. Increasing ϵ can strengthen decorrelation, but also increases leakage and the separation needed for reliable screening. A useful mixing level balances these effects. The resulting amplitude accuracy still depends on the selected columns of $\boldsymbol{A}$, as quantified by Proposition 3.9, rather than on the MC of $\boldsymbol{AQ}$ alone.

3.3 Practical Implementation

Algorithm 1 summarizes direct auxiliary screening. For a general matrix with nonzero columns, write $\boldsymbol{A} = \hat{\boldsymbol{A}}\boldsymbol{W}$, where $\boldsymbol{W} = \mathrm{diag}(\|\boldsymbol{a}_1\|_2, \ldots, \|\boldsymbol{a}_n\|_2)$. The exact relation is

$$\boldsymbol{y} = \hat{\boldsymbol{A}}\boldsymbol{Q}_{\epsilon,n}\boldsymbol{z} + \boldsymbol{n}, \qquad \boldsymbol{z} = \boldsymbol{Q}_{\epsilon,n}^{-1}\boldsymbol{W}\boldsymbol{x}. \tag{15}$$

The diagonal scaling preserves $\mathcal{S}$, and Theorem 3.7 applies to $\boldsymbol{W}\boldsymbol{x}$ in this implementation. If the mixed columns are normalized again inside a solver, their coefficient scaling is undone before screening. Refitting on $\boldsymbol{A}_{:,\widehat{\mathcal{S}}}$ restores the original scale. Penalized, greedy, and Bayesian estimators provide candidate auxiliary estimates. Integration with a base solver follows its coefficient representation and update rule. For ABESS and OP-BESS, mixed-domain candidates are mapped to the original coordinates for support initialization, with an additional refinement step for OP-BESS (Appendix G.3.2).

When s is unknown, we form $\widehat{\mathcal{S}}_k = \mathrm{Top}_k(|\hat{\boldsymbol{z}}|)$, refit each candidate to obtain $\hat{\boldsymbol{u}}_k$, and select $k \le k_{\max}$ by BIC (Neath & Cavanaugh, 2012):

$$\mathrm{BIC}(k) = p \log\left(\frac{\|\boldsymbol{y} - \boldsymbol{A}_{:,\widehat{\mathcal{S}}_k}\hat{\boldsymbol{u}}_k\|_2^2}{p}\right) + k \log p. \tag{16}$$

This selects the model size, separately from the mixing variance. Our main experiments use $\sqrt{\epsilon} = 0.025$, with application-specific sensitivity studies in the appendices. In a new setting, a small validation grid can identify a stable reconstruction region, within which a smaller perturbation limits unnecessary leakage.

3.4 Complexity

DogRot requires one preprocessing step for a fixed sensing matrix. The preconditioning stage requires generating $\frac{n(n-1)}{2}$ Gaussian random variables and forming the mixed matrix $\boldsymbol{A}\boldsymbol{Q}_{\epsilon,n}$, which costs $\mathcal{O}(pn^2)$ for a dense implementation. The post-processing stage ranks $\hat{\boldsymbol{z}}$ and solves restricted least squares on $|\widehat{\mathcal{S}}|$ columns, costing $\mathcal{O}(n \log n + ps^2 + s^3)$ for a fixed support size. Independently refitting every model up to $k_{\max}$ costs $\mathcal{O}(pk_{\max}^3 + k_{\max}^4)$ in total, or less when nested factorizations are reused. Dense storage is $\mathcal{O}(n^2)$. DogRot itself requires no specialized hardware and can run on a standard workstation. Accelerators are only needed if the chosen base solver requires them.

4 Experiments

The problem of sparse signal recovery with HIM arises in a wide range of applications in signal processing and machine learning. We examine MRI reconstruction, the combination with modern support selectors in LSCM, and GDC. The full application studies are provided in Appendices F–H.

4.1 Experiment settings

MRI reconstruction recovers images from under-sampled k-space data by exploiting sparsity in the wavelet or discrete cosine transform (DCT) domain (Lustig et al., 2007; Eksioglu, 2016; Ryu et al., 2019; Ding et al., 2023). Although the Fourier and DCT bases are orthogonal, under-sampling can introduce aliasing, particularly under Cartesian sampling. Low sampling rates can thus produce a HIM. We use the iterative shrinkage-thresholding algorithm (ISTA), the alternating direction method of multipliers (ADMM), and a variational network (VarNet), and measure image quality by peak signal-to-noise ratio (PSNR). Appendix F gives the detailed settings.

4.2 Reduction in mutual coherence

Table 1 compares MC before and after mixing across sampling patterns and rates. Under Cartesian sampling, the original MC reaches 0.99, whereas DogRot reduces it substantially. Radial operators have lower initial MC but also exhibit a reduction after mixing.

Table 1: MC before and after DogRot across MRI sampling patterns and rates.

Pattern	Matrix	10%	20%	30%	40%	50%
Cartesian	$\boldsymbol{A}$	0.99	0.90	0.74	0.99	0.99
	$\boldsymbol{AQ}$	0.37	0.33	0.26	0.34	0.33
Radial	$\boldsymbol{A}$	0.53	0.42	0.22	0.19	0.15
	$\boldsymbol{AQ}$	0.29	0.21	0.19	0.13	0.11

4.3 Improvement in signal reconstruction

Table 2: PSNR (dB) at 30% sampling. Entries report baseline/DogRot, with improvements in bold.

Pattern	Solver	Phantom	Notumor	Pituitary	Glioma	Mening.1	Mening.2
Cartesian	ISTA	16.26/**18.19**	12.31/**14.62**	12.17/**14.79**	15.29/**17.16**	13.12/**15.05**	15.06/**16.98**
	ADMM	16.26/**18.19**	12.50/**14.77**	13.10/**14.79**	15.29/**17.17**	13.21/**15.14**	15.08/**17.05**
	VarNet	25.65/**25.70**	23.78/**24.10**	25.34/**25.77**	26.15/**26.32**	23.39/**23.43**	24.69/**24.71**
Radial	ISTA	21.40/**22.39**	22.08/**24.12**	24.51/**25.19**	26.26/**27.22**	25.17/**26.09**	28.32/**29.39**
	ADMM	21.40/**22.39**	22.15/**24.12**	24.60/**25.22**	26.26/**27.25**	25.21/**26.12**	28.92/**30.42**

Table 2 reports PSNR for the six test images at 30% sampling. DogRot improves both ISTA and ADMM, with the largest gains under Cartesian sampling, where the original operator produces strong aliasing. VarNet also improves on all six images, with smaller gains of 0.02–0.43 dB. Its learned reconstruction prior provides a substantially stronger baseline, while the additional gains demonstrate compatibility with a learned solver in this setting. The 20% results and representative reconstructions are given in Appendix F. Under Cartesian sampling, overlap with the reference image's largest 1% DCT coefficients also improves for both classical solvers at 20% and 30% sampling (Appendix F.3.1).

To examine how the benefit varies with the sensing conditions, we sweep radial sampling rates from 10% to 50%. The sweep changes both the measurement count and column geometry, while each baseline/DogRot pair uses identical measurements. Table 3 reports support overlap and PSNR alongside the MC values in Table 1. Appendix F.3.2 gives the averaging protocol.

Table 3: Radial sweep: baseline/DogRot overlap (%) and PSNR (dB). Better values are bold.

Sampling	ISTA		ADMM	
	Overlap	PSNR	Overlap	PSNR
10%	52.13/**66.87**	20.18/**24.04**	53.76/**67.02**	20.34/**24.32**
20%	76.52/**79.16**	22.59/**23.79**	77.54/**79.41**	22.80/**24.00**
30%	87.50/**89.38**	24.62/**25.73**	87.91/**89.48**	24.76/**25.92**
40%	**95.83**/94.79	**26.81**/26.68	**95.93**/95.07	**26.91**/26.73
50%	**98.07**/95.72	**28.92**/28.06	**98.27**/95.74	**29.02**/28.11

Both solvers improve in support overlap and PSNR at 10%–30% sampling. At 10%, for example, ADMM overlap rises from 53.76% to 67.02% and PSNR from 20.34 to 24.32 dB. At 40%–50%, the original MC is lower and baseline overlap already exceeds 95%. Mixing still reduces MC, but lowers both recovery metrics. Within this acquisition family, the diminishing benefit is consistent with the decorrelation–leakage trade-off: when support screening is already reliable, further mixing can degrade the auxiliary representation without a compensating recovery gain.

4.4 Combination with Modern Support Selectors

We evaluate ABESS and OP-BESS with DogRot on LSCM, using three array configurations and sparsity levels $s = 5, 10$. P@0 measures exact-grid precision, while P@1 allows one-to-one matching within one grid step. We report normalized mean squared error (NMSE) for coefficient recovery and Earth mover's distance (EMD) for spatial error in the recovered angular power spectrum. The mixed-domain candidates are mapped to the original coordinates before refitting, with an additional support-refinement step for OP-BESS. The detailed protocols and a matched two-stage $\boldsymbol{Q} = \boldsymbol{I}$ control are given in Appendix G.3.2.

Table 4: LSCM: baseline/DogRot precision (%), NMSE, and EMD. Better values are bold.

Selector	s	P@0	P@1	NMSE	EMD
ABESS	5	**17.11**/16.67	**23.78**/22.89	222.66/**11.81**	29.99/**29.95**
	10	6.00/6.00	**10.56**/10.11	518.65/**72.77**	32.36/**30.99**
OP-BESS	5	34.89/**39.11**	40.44/**44.00**	103.16/**8.27**	24.61/**20.73**
	10	13.44/**15.56**	18.78/**22.22**	465.19/**32.99**	31.50/**25.94**

Table 4 shows substantially lower amplitude errors for both selectors. OP-BESS also improves precision and EMD, whereas ABESS has mixed precision results. With initialization and refinement fixed, DogRot reduces NMSE from 138.01 to 8.27 ($s = 5$) and 467.11 to 32.99 ($s = 10$) relative to $\boldsymbol{Q} = \boldsymbol{I}$ (Appendix G.3.2), isolating the contribution of mixing. Appendix G.3.3 includes left-preconditioner comparisons under measurement noise.

4.5 Gene-Based Disease Classification

For GDC, we use a leukemia dataset with 7,129 genes and 72 samples (Golub et al., 1999). Table 5 compares the least absolute shrinkage and selection operator (LASSO) and elastic net with and without DogRot at $\sqrt{\epsilon} = 0.025$. Both classifiers improve across all four training-set sizes despite the highly correlated features and limited observations. Appendix H provides the protocol, MC, gene-selection stability, and sensitivity results.

Table 5: Gene classification accuracy across training-set sizes. DogRot improvements are bold.

Training samples	LASSO		Elastic net	
	Baseline	DogRot	Baseline	DogRot
15	53.33%	**72.04%**	75.95%	**83.46%**
20	73.19%	**82.29%**	82.33%	**87.71%**
25	82.19%	**86.73%**	83.71%	**91.85%**
30	83.21%	**87.13%**	86.72%	**94.62%**

5 Conclusion

This paper studied sparse signal recovery under highly ill-conditioned sensing matrices. DogRot mixes sensing columns to reduce mutual coherence, uses the mixed representation for support screening or initialization, and refits coefficients in the original system. The analysis establishes geometry-dependent decorrelation and dominant-support preservation for the exact inverse, with an explicit auxiliary-error condition for screening an estimated signal. Experiments on MRI, LSCM, and gene-based disease classification demonstrate gains in support localization and reconstruction, including substantially reduced amplitude errors with modern subset selectors. The sensitivity studies identify the balance between decorrelation and leakage. Structured and implicit mixers offer a direction for extending this approach to larger problems.

ARTIFICIAL INTELLIGENCE (AI) USE STATEMENT

We used generative AI tools to assist with language polishing for potential grammar errors. The authors reviewed all suggested edits and take full responsibility for the final manuscript.

REFERENCES

Vahid Abolghasemi, Saideh Ferdowsi, Bahador Makkiabadi, and Saeid Sanei. On optimization of the measurement matrix for compressive sensing. In *2010 18th Eur. Signal Process. Conf.*, pp. 427–431, 2010.

Cagla D. Bahadir, Alan Q. Wang, Adrian V. Dalca, and Mert R. Sabuncu. Deep-learning-based optimization of the under-sampling pattern in MRI. *IEEE Transactions on Computational Imaging*, 6:1139–1152, 2020. doi: 10.1109/TCI.2020.3006727.

Dror Baron, Shriram Sarvotham, and Richard G. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE Transactions on Signal Processing*, 58(1):269–280, 2010. doi: 10.1109/TSP.2009.2027773.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi: 10.1561/2200000016.

Alfred M. Bruckstein, Michael Elad, and Michael Zibulevsky. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Trans. Inf. Theory*, 54(11):4813–4820, 2008. doi: 10.1109/TIT.2008.929920.

E.J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9–10):589–592, 2008. ISSN 1631-073X. doi: https://doi.org/10.1016/j.crma.2008.03.014. URL https://www.sciencedirect.com/science/article/pii/S1631073X08000964.

E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005. doi: 10.1109/TIT.2005.858979.

E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006. doi: 10.1109/TIT.2005.862083.

Emmanuel J. Candes and Michael B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008. doi: 10.1109/MSP.2007.914731.

Yanyun Ding, Peili Li, Yunhai Xiao, and Haibin Zhang. Efficient dual ADMMs for sparse compressive sensing MRI reconstruction. *Mathematical Methods of Operations Research*, 97(2):207–231, 2023.

David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci.*, 100(5):2197–2202, 2003. doi: 10.1073/pnas.0437847100. URL https://www.pnas.org/doi/abs/10.1073/pnas.0437847100.

D.L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006. doi: 10.1109/TIT.2006.871582.

Chenhe Du, Xuanyu Tian, Qing Wu, Muyu Liu, Jingyi Yu, Hongjiang Wei, and Yuyao Zhang. Plug-and-play diffusion meets ADMM: Dual-variable coupling for robust medical image reconstruction, 2026. URL https://arxiv.org/abs/2602.23214.

Ender M Eksioglu. Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI. *Journal of Mathematical Imaging and Vision*, 56(3):430–440, 2016.

Michael Elad. Optimized projections for compressed sensing. *IEEE Trans. Signal Process.*, 55(12): 5695–5702, 2007. doi: 10.1109/TSP.2007.900760.

Zhen Gao, Linglong Dai, Zhaocheng Wang, and Sheng Chen. Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO. *IEEE Trans. Signal Process.*, 63(23):6169–6183, 2015. doi: 10.1109/TSP.2015.2463260.

Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A Caligiuri, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 286(5439):531–537, 1999.

Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1–3):389–422, 2002.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018. doi: 10.1002/mrm.26977.

Matthew A. Herman and Thomas Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009. doi: 10.1109/TSP.2009.2014277.

Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008. doi: 10.1109/TSP.2007.914345.

Berkan Kilic, Alper Güngör, Mert Kalfa, and Orhan Arıkan. Adaptive measurement matrix design in direction of arrival estimation. *IEEE Trans. Signal Process.*, 70:4742–4756, 2022. doi: 10.1109/TSP.2022.3209880.

Bo Li, Liang Zhang, Thia Kirubarajan, and Sreeraman Rajan. Projection matrix design using prior information in compressive sensing. *Signal Processing*, 135:36–47, 2017. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2016.11.024. URL https://www.sciencedirect.com/science/article/pii/S0165168416303401.

Caiyan Li and Hongzhe Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008.

Zhi-Quan Luo, Xi Zheng, David López-Pérez, Qi Yan, Xin Chen, Nanbin Wang, Qingjiang Shi, Tsung-Hui Chang, and Adrian Garcia-Rodriguez. SRCON: A data-driven network performance simulator for real-world wireless networks. *IEEE Communications Magazine*, 61(6):96–102, 2023.

Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, 2007. doi: https://doi.org/10.1002/mrm.21391. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.21391.

Andrew A Neath and Joseph E Cavanaugh. The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2): 199–203, 2012.

D. Needell and J.A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2008.07.002.

Deanna Needell and Roman Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Sel. Top. Signal Process.*, 4(2):310–316, 2010.

Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–44 vol.1, 1993. doi: 10.1109/ACSSC.1993.342465.

Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pp. 5546–5557. PMLR, 2019.

Julian Tachella, Matthieu Terris, Samuel Hurault, Andrew Wang, Dongdong Chen, Minh-Hai Nguyen, Maxime Song, Thomas Davies, Leo Davy, Jonathan Dong, Paul Escande, Johannes Hertrich, Zhiyuan Hu, Tobias I. Liaudat, Nils Laurent, Brett Levac, Mathurin Massias, Thomas Moreau, Thibaut Modrzyk, Brayan Monroy, Sebastian Neumayer, Jeremy Scanvic, Florian Sarron, Victor Sechaud, Georg Schramm, Romain Vo, and Pierre Weiss. DeepInverse: A Python package for solving imaging inverse problems with deep learning, 2025. URL https://arxiv.org/abs/2505.20160.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Andreas M. Tillmann and Marc E. Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2014. doi: 10.1109/TIT.2013.2290112.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Jianyu Wang, Liu Yang, Chaoyuan Bai, and Heng Liu. Improved sparse activity detection based on optimized measurement matrix in mMTC. In *2024 IEEE 100th Veh. Technol. Conf. (VTC2024-Fall)*, pp. 1–7, 2024. doi: 10.1109/VTC2024-Fall63153.2024.10757915.

Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang. Real-time compressive tracking. In *Computer Vision – ECCV 2012*, pp. 864–877. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33712-3.

Shutao Zhang, Xinzhi Ning, Xi Zheng, Qingjiang Shi, Tsung-Hui Chang, and Zhi-Quan Luo. A physics-based and data-driven approach for localized statistical channel modeling. *IEEE Trans. Wireless Commun.*, 23(6):5409–5424, 2024. doi: 10.1109/TWC.2023.3326209.

Shutao Zhang, Ye Xue, Zhiwei Tang, Hao Wang, Chao Shen, Qingjiang Shi, and Tsung-Hui Chang. Robust network optimization by deep generative models and stochastic optimization. *IEEE Transactions on Wireless Communications*, 2025.

Jin Zhu, Xueqin Wang, Liyuan Hu, Junhao Huang, Kangkang Jiang, Yanhang Zhang, Shiyun Lin, and Junxian Zhu. abess: A fast best-subset selection library in Python and R. *Journal of Machine Learning Research*, 23(202):1–7, 2022. URL https://www.jmlr.org/papers/v23/21-1060.html.

Junxian Zhu, Canhong Wen, Jin Zhu, Heping Zhang, and Xueqin Wang. A polynomial algorithm for best-subset selection problem. *Proceedings of the National Academy of Sciences*, 117(52):33117–33123, 2020. doi: 10.1073/pnas.2014241117.

Zhihan Zhu, Yanhao Zhang, and Yong Xia. Best subset selection: Optimal pursuit for feature selection and elimination. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 80398–80426. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhu25ad.html.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005.

A Column Concentration and Entrywise Orthogonality

A.1 Proof of Theorem 3.3

Only the lower-triangular entries of $\boldsymbol{S} = \boldsymbol{Q}_{\epsilon,n} - \boldsymbol{I}_n$ are independent. Within a fixed column, however, its $n-1$ off-diagonal entries are independent $\mathcal{N}(0,\epsilon)$ variables, up to deterministic signs. Consequently,

$$\|\boldsymbol{q}_i\|_2^2 = 1 + \epsilon X, \qquad X \sim \chi^2_{n-1}.$$

Writing $d = n-1$, for $0 < t < 1$ the standard chi-square bound (Vershynin, 2018) gives

$$\Pr\{|X/d - 1| \geq t\} \leq 2e^{-dt^2/8}.$$

Since $\epsilon d/m_n \leq 1$, $|\|\boldsymbol{q}_i\|_2^2/m_n - 1| = (\epsilon d/m_n)|X/d - 1|$, which proves the stated result.

A.2 Proof of Theorem 3.4 and Corollary 3.5

For $i \neq j$, skew symmetry gives

$$\boldsymbol{q}_i^T \boldsymbol{q}_j = S_{ij} + S_{ji} + \sum_k S_{ki} S_{kj} = \sum_{k \neq i,j} S_{ki} S_{kj}.$$

Each summand uses two distinct independent lower-triangular Gaussian variables, and these pairs are disjoint across k. No independence across different column pairs is needed. If X, Y are independent $\mathcal{N}(0,\epsilon)$ variables, then

$$\mathbb{E}e^{\lambda XY} = (1 - \epsilon^2\lambda^2)^{-1/2}, \qquad |\lambda| < 1/\epsilon.$$

For $|\lambda| \leq 1/(\sqrt{2}\epsilon)$, the logarithm is at most $\epsilon^2\lambda^2$. Applying the Chernoff bound to the sum and optimizing over this interval proves the two-regime tail in Theorem 3.4, with a universal positive constant c_0. When $n = 2$, the inner product is identically zero.

For completeness, take $n \geq 3$, $\epsilon > 0$, and $t \in (0,1)$. A union bound over the diagonal and off-diagonal entries yields

$$\begin{aligned}
&\Pr\left\{\|m_n^{-1}\boldsymbol{Q}_{\epsilon,n}^T\boldsymbol{Q}_{\epsilon,n} - \boldsymbol{I}_n\|_{\max} \geq t\right\} \\
&\quad \leq 2ne^{-(n-1)t^2/8} + n(n-1)\exp\left[-c_0 \min\left\{\frac{m_n^2 t^2}{(n-2)\epsilon^2}, \frac{m_n t}{\epsilon}\right\}\right].
\end{aligned}$$

Because $m_n/\epsilon \geq n-1$, both exponents dominate a positive multiple of n for fixed t. The right side tends to zero, including when ϵ varies with n. Entrywise convergence does not imply convergence in spectral norm when the dimension increases.

B Geometry-Dependent Coherence Bound

B.1 Independent Gaussian Parameterization

Let $d = n(n-1)/2$ and write

$$\boldsymbol{S} = \sqrt{\epsilon}\sum_{r=1}^{d} g_r \boldsymbol{F}_r, \qquad \boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d),$$

where the canonical skew basis consists of $\boldsymbol{F}_{ab} = \boldsymbol{e}_a\boldsymbol{e}_b^T - \boldsymbol{e}_b\boldsymbol{e}_a^T$ for $a > b$. This parameterization respects all dependencies imposed by skew symmetry. Suppose that $\boldsymbol{A}$ is fixed independently of the mixer, has unit-norm columns, and set $\boldsymbol{G} = \boldsymbol{A}^T\boldsymbol{A}$. Define the deterministic vector $\boldsymbol{\ell}_{ij}$ and symmetric matrix $\boldsymbol{H}_{ij}$ by

$$[\boldsymbol{\ell}_{ij}]_r = \boldsymbol{e}_i^T\boldsymbol{F}_r^T\boldsymbol{G}\boldsymbol{e}_j + \boldsymbol{e}_i^T\boldsymbol{G}\boldsymbol{F}_r\boldsymbol{e}_j, \tag{17}$$

$$[\boldsymbol{K}_{ij}]_{rt} = \boldsymbol{e}_i^T\boldsymbol{F}_r^T\boldsymbol{G}\boldsymbol{F}_t\boldsymbol{e}_j, \qquad \boldsymbol{H}_{ij} = \tfrac{1}{2}(\boldsymbol{K}_{ij} + \boldsymbol{K}_{ij}^T). \tag{18}$$

The exact expansion is

$$\boldsymbol{b}_i^T\boldsymbol{b}_j = G_{ij} + \sqrt{\epsilon}\,\boldsymbol{\ell}_{ij}^T\boldsymbol{g} + \epsilon\boldsymbol{g}^T\boldsymbol{H}_{ij}\boldsymbol{g}. \tag{19}$$

The identity $\sum_r \boldsymbol{F}_r^T\boldsymbol{G}\boldsymbol{F}_r = \mathrm{tr}(\boldsymbol{G})\boldsymbol{I}_n - \boldsymbol{G}$ and $\mathrm{tr}(\boldsymbol{G}) = n$ give

$$\mathbb{E}[\boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}] = (1-\epsilon)\boldsymbol{G} + \epsilon n\boldsymbol{I}_n,$$

and hence (9). These are expectations of unnormalized Gram entries, not expectations of their ratios.

B.2 Joint Fluctuation Control

For a deterministic vector $\boldsymbol{\ell}$, a symmetric matrix $\boldsymbol{H}$, and $t > 0$, Gaussian concentration gives

$$\Pr\{|\boldsymbol{\ell}^T\boldsymbol{g}| > \|\boldsymbol{\ell}\|_2\sqrt{2t}\} \le 2e^{-t},$$

and

$$\Pr\{|\boldsymbol{g}^T\boldsymbol{H}\boldsymbol{g} - \operatorname{tr}(\boldsymbol{H})| > 2\|\boldsymbol{H}\|_F\sqrt{t} + 2\|\boldsymbol{H}\|_2 t\} \le 2e^{-t}.$$

To obtain the second inequality, diagonalize $\boldsymbol{H}$ and write the centered quadratic form as $X = \sum_r \lambda_r(h_r^2 - 1)$ with independent standard Gaussian h_r. For $2|\theta|\|\boldsymbol{H}\|_2 < 1$,

$$\log \mathbb{E}e^{\theta X} \le \frac{\theta^2\|\boldsymbol{H}\|_F^2}{1 - 2|\theta|\|\boldsymbol{H}\|_2}.$$

The Chernoff bound with $\theta = \sqrt{t}/(\|\boldsymbol{H}\|_F + 2\|\boldsymbol{H}\|_2\sqrt{t})$, applied to X and $-X$, proves the claim, including for indefinite $\boldsymbol{H}$. The zero-matrix case is deterministic.

Let $N = n(n+1)/2$ and $t_\delta = \log(4N/\delta)$. Define

$$R_{ij} = \sqrt{2\epsilon t_\delta}\,\|\boldsymbol{\ell}_{ij}\|_2 + 2\epsilon\left(\|\boldsymbol{H}_{ij}\|_F\sqrt{t_\delta} + \|\boldsymbol{H}_{ij}\|_2 t_\delta\right). \tag{20}$$

A union bound over the N distinct Gram entries controls both terms in (19) simultaneously, with probability at least $1 - \delta$. On this event,

$$|\boldsymbol{b}_i^T\boldsymbol{b}_j - \mathbb{E}[\boldsymbol{b}_i^T\boldsymbol{b}_j]| \le R_{ij} \quad \text{for every } i \le j.$$

If all $m_n - R_{ii}$ are positive, division by the lower bounds for the column norms proves (10). The additional deterministic condition $U_\delta(\boldsymbol{A}, \epsilon) < \mu(\boldsymbol{A})$ proves strict reduction. Neither the numerator and denominator nor the individual Gram entries have been treated as independent.

B.3 Geometry and the Degenerate Boundary

The radii in (20) depend on the full Gram geometry, so dimension alone does not imply an $\exp(-cn^2)$ concentration rate. In particular, $\operatorname{rank}(\boldsymbol{A}) > 1$ alone is not a sufficient decorrelation condition.

If $\boldsymbol{A} = \boldsymbol{u}\boldsymbol{v}^T$ has rank one, then $\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{u}(\boldsymbol{v}^T\boldsymbol{Q})$ has collinear nonzero columns. For a fixed $\boldsymbol{v} \neq \boldsymbol{0}$, the jth coordinate of $\boldsymbol{v}^T\boldsymbol{Q}$ has mean v_j and variance $\epsilon(\|\boldsymbol{v}\|_2^2 - v_j^2)$. It is either a nondegenerate Gaussian or a nonzero constant. All mixed columns are therefore nonzero almost surely, and $\mu(\boldsymbol{A}\boldsymbol{Q}) = 1$. Column mixing cannot resolve this loss of information.

More generally, the bound characterizes a sufficient event for decorrelation, not a necessary condition for success of a recovery algorithm. Support screening also depends on signal separation and on the auxiliary estimation error, as specified in Theorem 3.7 and Corollary 3.8.

C Exact-Inverse Support Analysis

C.1 Inverse Identities and Rotational Symmetry

Let $\boldsymbol{R} = (\boldsymbol{I}_n + \boldsymbol{S})^{-1}$ and $\boldsymbol{v} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$. Since $\boldsymbol{S}^T = -\boldsymbol{S}$,

$$\boldsymbol{R}^T\boldsymbol{R} = \frac{\boldsymbol{R} + \boldsymbol{R}^T}{2} = (\boldsymbol{I}_n + \boldsymbol{S}^T\boldsymbol{S})^{-1}. \tag{21}$$

Put $a = \boldsymbol{v}^T\boldsymbol{R}\boldsymbol{v}$. Equation (21) implies $a = \|\boldsymbol{R}\boldsymbol{v}\|_2^2 \in (0, 1]$. Thus $\boldsymbol{R}\boldsymbol{v} - a\boldsymbol{v}$ is perpendicular to $\boldsymbol{v}$ and has squared norm $a(1-a)$.

The Gaussian skew-symmetric ensemble has density proportional to $\exp(-\|\boldsymbol{S}\|_F^2/(4\epsilon))$ on the space of skew-symmetric matrices. Its law is invariant under orthogonal conjugation. Conjugating by any orthogonal matrix that fixes $\boldsymbol{v}$ leaves a unchanged and rotates $\boldsymbol{R}\boldsymbol{v} - a\boldsymbol{v}$ within $\boldsymbol{v}^\perp$. Consequently, conditionally on a, its direction is uniform on the unit sphere of $\boldsymbol{v}^\perp$, giving the exact representation

$$\boldsymbol{R}\boldsymbol{x} = a\boldsymbol{x} + \|\boldsymbol{x}\|_2\sqrt{a(1-a)}\,\boldsymbol{u}. \tag{22}$$

If $a = 1$, the second term is zero and its direction is immaterial. For fixed nonzero $\boldsymbol{x}$ and $\epsilon > 0$, this exceptional event has probability zero.

C.2 Finite-Dimensional Separation

The inverse quadratic-form inequality applied to $\boldsymbol{I}_n + \boldsymbol{S}^T\boldsymbol{S}$ gives

$$a \geq \frac{1}{1 + \|\boldsymbol{S}\boldsymbol{v}\|_2^2}.$$

Orthogonal invariance also implies $\|\boldsymbol{S}\boldsymbol{v}\|_2^2/\epsilon \sim \chi^2_{n-1}$. The chi-square upper-tail bound at $t = \log(2/\delta)$ therefore gives $a \geq (1 + \Gamma_\delta)^{-1}$ except on an event of probability $\delta/2$.

For a uniform unit vector $\boldsymbol{u}$ in the $(n-1)$-dimensional subspace $\boldsymbol{v}^\perp$, spherical concentration gives, for each coordinate and $r > 0$,

$$\Pr\{|u_j| > r \mid a\} \leq 2\exp\left(-\frac{(n-2)r^2}{2}\right).$$

Indeed, a coordinate of a uniform vector on the sphere in $\mathbb{R}^{n-1}$ has the distribution $Z/\sqrt{Z^2 + W}$, with independent $Z \sim \mathcal{N}(0,1)$ and $W \sim \chi^2_{n-2}$. For $0 < r < 1$, conditioning on W bounds its two-sided tail by $2(1-r^2)^{(n-2)/2} \leq 2e^{-(n-2)r^2/2}$. The orthogonal projection of a coordinate vector onto $\boldsymbol{v}^\perp$ has norm at most one, so the same bound holds regardless of $\boldsymbol{v}$. For $r \geq 1$ the event is empty. Taking a union bound over j with $r = r_\delta$ controls $\|\boldsymbol{u}\|_\infty$ with conditional probability at least $1 - \delta/2$. Combining these two events does not require independence. On their intersection,

$$\frac{\|\boldsymbol{R}\boldsymbol{x} - a\boldsymbol{x}\|_\infty}{a} \leq \|\boldsymbol{x}\|_2\sqrt{\frac{1-a}{a}}\, r_\delta \leq \|\boldsymbol{x}\|_2\sqrt{\Gamma_\delta}\, r_\delta. \tag{23}$$

An active coordinate has magnitude at least $ax_{\min} - \|\boldsymbol{R}\boldsymbol{x} - a\boldsymbol{x}\|_\infty$, while every inactive coordinate has magnitude at most the latter error. The strict separation in Theorem 3.7 proves its conclusion.

For $\epsilon(n-1) = O(1)$ and $\delta = n^{-q}$ with fixed $q > 0$, $\Gamma_\delta = O(1)$ and $r_\delta = O(\sqrt{\log n/n})$. If the nonzero amplitudes have bounded dynamic range, $x_{\min}/\|\boldsymbol{x}\|_2$ is bounded below by a constant times $s^{-1/2}$. Hence $s = o(n/\log n)$ suffices. For a finite-dimensional illustration, $n = 6552$, $\epsilon = 0.000625$, and $\delta = 0.05$ give $2\sqrt{\Gamma_\delta}r_\delta \approx 0.263$. A flat $s = 10$ signal has ratio $1/\sqrt{10} \approx 0.316$ and satisfies the condition. This calculation concerns the exact transformed vector, not an unmeasured error bound for a particular solver.

C.3 Proof of Corollary 3.8

On the event used above, if $\|\hat{\boldsymbol{z}} - \boldsymbol{R}\boldsymbol{x}\|_\infty \leq \eta$, then

$$\min_{j\in\mathcal{S}} |\hat{z}_j| - \max_{j\notin\mathcal{S}} |\hat{z}_j| \geq ax_{\min} - 2\|\boldsymbol{R}\boldsymbol{x} - a\boldsymbol{x}\|_\infty - 2\eta.$$

Dividing by a, using (23) and $a^{-1} \leq 1 + \Gamma_\delta$, yields (13). If the solver-error event itself holds with probability at least $1 - \delta_{\mathrm{sol}}$, the intersection has probability at least $1 - \delta - \delta_{\mathrm{sol}}$ by a union bound. No independence between the solver and the mixer is assumed. This result applies to direct auxiliary screening. A mapped support initializer followed by original-domain search, as used in Appendix G.3.2, is a separate algorithmic variant.

D Restricted Refitting

On the correct support, unrestricted least squares gives $\hat{\boldsymbol{x}}_{\mathcal{S}} - \boldsymbol{x}_{\mathcal{S}} = \boldsymbol{A}_{:,\mathcal{S}}^\dagger\boldsymbol{n}$, proving Proposition 3.9. For nonnegative APS recovery, the refit instead minimizes the same residual over $\boldsymbol{u} \geq \boldsymbol{0}$. If $\boldsymbol{x}_{\mathcal{S}} \geq \boldsymbol{0}$ and the support is correct, projection of $\boldsymbol{y}$ onto the closed convex cone $\{\boldsymbol{A}_{:,\mathcal{S}}\boldsymbol{u} : \boldsymbol{u} \geq \boldsymbol{0}\}$ is nonexpansive. Therefore $\|\boldsymbol{A}_{:,\mathcal{S}}(\hat{\boldsymbol{x}}_{\mathcal{S}} - \boldsymbol{x}_{\mathcal{S}})\|_2 \leq \|\boldsymbol{n}\|_2$, and the same singular-value error bound holds when $\boldsymbol{A}_{:,\mathcal{S}}$ has full column rank. Neither refit removes noise amplification from a nearly singular selected submatrix.

E Additional Finite-Dimensional Properties

The exact-inverse analysis does not require a scaled-transpose approximation. Its accuracy, when of independent interest, can be characterized directly. Let $\lambda_k \geq 0$ be the eigenvalues of $\boldsymbol{S}^T\boldsymbol{S}$, and set

$\boldsymbol{D} = \boldsymbol{Q}^{-1} - m_n^{-1}\boldsymbol{Q}^T$. Orthogonal block diagonalization of the real skew-symmetric matrix $\boldsymbol{S}$ gives

$$\|\boldsymbol{D}\|_2 = \max_k \frac{|m_n - (1+\lambda_k)|}{m_n\sqrt{1+\lambda_k}}, \qquad \kappa_2(\boldsymbol{Q}) = \sqrt{\frac{1+\lambda_{\max}}{1+\lambda_{\min}}}. \tag{24}$$

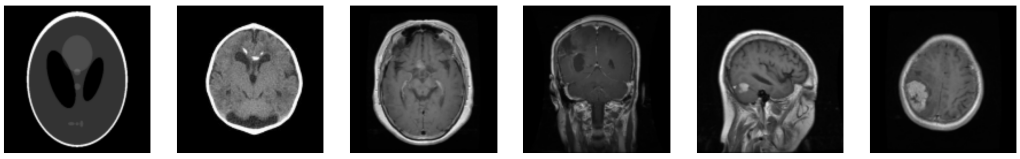
These formulas are valid in finite dimensions. Entrywise Gram concentration alone does not control the maximum in (24). In contrast, Theorem 3.7 controls the coordinatewise action of the exact inverse on a fixed signal, which is the quantity needed for support ranking.

F Supplementary Materials for MRI reconstruction

F.1 Experimental Settings

Dataset: The test images are drawn from the public brain-tumor MRI dataset available at `https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset` and from the Shepp-Logan phantom. All images are resized to 128×128.
Sampling: We consider Cartesian and radial sampling patterns, both commonly used in MRI. The sampling rate varies between 10% and 50%. Sparsity is imposed in the DCT domain. The test images are shown in Figure 1.
Metric: We use MC $\mu(\boldsymbol{A}) \in [0,1]$ to quantify the conditioning of the sensing matrix $\boldsymbol{A}$, and peak signal-to-noise ratio (PSNR), reported in dB, to measure reconstruction quality. Dominant-support overlap is $100\,|\widehat{\mathcal{T}} \cap \mathcal{T}|/|\mathcal{T}|$, where $\mathcal{T}$ contains the reference image's largest 1% DCT coefficients and $\widehat{\mathcal{T}}$ has the same cardinality, ranked directly from the auxiliary estimate. It measures dominant coefficients rather than all nonzero image coefficients. For ISTA and ADMM, we use the six images and 1,000 independent mixers per image. At each sampling setting, the mask and solver parameters are fixed, and each run records both dominant-support overlap and final-image PSNR.
Methods: We use ISTA and ADMM as classical MRI reconstruction solvers, and the image-domain VarNet of Hammernik et al. (2018), implemented through DeepInverse (Tachella et al., 2025), as a learned solver under single-coil Cartesian sampling. Their performance without mixing serves as the baseline for evaluating DogRot. We set $\sqrt{\epsilon} = 0.025$.

Learned priors and acquisition design address complementary parts of MRI reconstruction. Diffusion plug-and-play (PnP) methods (Du et al., 2026) modify the reconstruction prior and updates, while learning-based optimization of the under-sampling pattern (LOUPE) (Bahadir et al., 2020) jointly learns the sampling pattern and reconstruction. Our paired comparisons keep the acquired measurements fixed. A pretrained image-domain prior would require an explicit mapping between image and mixed-coefficient variables, rather than treating the auxiliary coefficients as ordinary images.



(a) Phantom (b) Notumor (c) Pituitary (d) Glioma (e) Meningioma1 (f) Meningioma2

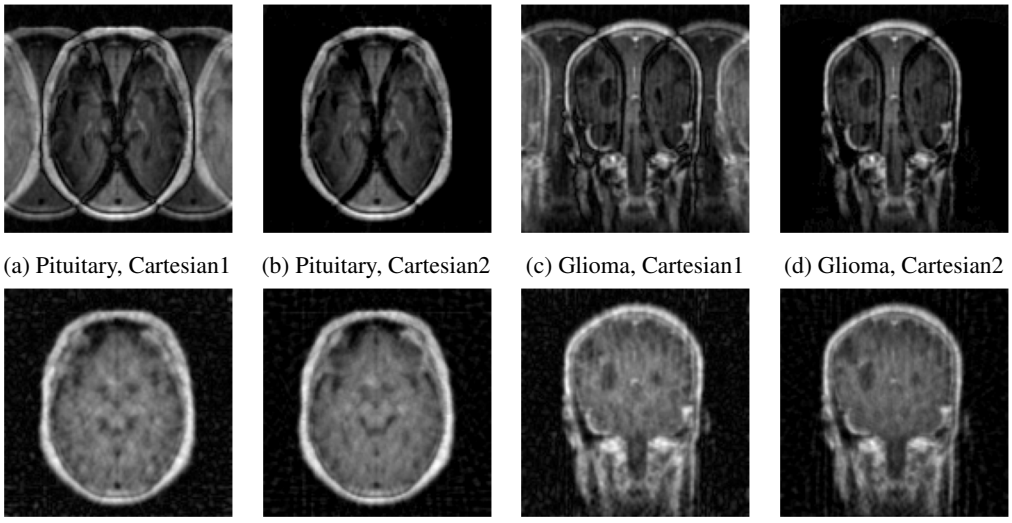
Figure 1: Six different experimental MRIs acquired from the public dataset

F.2 Reduction in mutual coherence

The mutual-coherence reduction for MRI reconstruction is reported in Table 1. The same sensing construction is used throughout the supplementary MRI experiments. Under Cartesian sampling, the original sensing matrix is nearly maximally coherent, whereas DogRot lowers the MC after column mixing. Under radial sampling, the original MC is lower but still decreases after applying DogRot. These results show that right-side mixing improves the conditioning of the MRI sensing matrix across sampling patterns and sampling rates, with the downstream effect evaluated separately below.

F.3 IMPROVEMENT IN SIGNAL RECONSTRUCTION

The PSNR comparison for six test images under a 20% sampling rate is provided in Table 6. Consistent with the 30% sampling case reported in Table 2, DogRot improves the reconstruction quality achieved by existing algorithms, even under challenging conditions. Comparing Tables 2 and 6, the PSNR obtained at 20% sampling with DogRot can in some cases match or approach that at 30% sampling without DogRot. This indicates that better conditioning can reduce the sampling burden in MRI reconstruction. A visual comparison of MRI reconstructions before and after applying DogRot is provided in Figure 2. In this example, ADMM is employed under a 30% sampling rate with both Cartesian and radial sampling patterns. The pituitary and glioma images are selected as representative cases. In Figure 2, the label "1" corresponds to ADMM alone, while "2" denotes the result after incorporating DogRot.



(a) Pituitary, Cartesian1 (b) Pituitary, Cartesian2 (c) Glioma, Cartesian1 (d) Glioma, Cartesian2

(e) Pituitary, Radial1 (f) Pituitary, Radial2 (g) Glioma, Radial1 (h) Glioma, Radial2

Figure 2: Comparison of images reconstructed by ADMM before and after DogRot under different sampling patterns with a 30% sampling rate

Table 6: PSNR comparison at 20% sampling rate. Each entry reports baseline/DogRot, and bold values indicate improvements achieved by DogRot. VarNet is evaluated under Cartesian sampling.

Pattern	Solver	Phantom	Notumor	Pituitary	Glioma	Mening.1	Mening.2
Cartesian	ISTA	15.03/**16.08**	11.71/**12.95**	11.82/**13.30**	14.91/**16.04**	12.51/**13.99**	14.34/**15.98**
	ADMM	15.03/**16.14**	11.81/**13.06**	12.56/**13.50**	14.91/**16.12**	12.60/**14.12**	14.37/13.50
	VarNet	22.02/**22.48**	20.02/**20.86**	22.11/**23.01**	23.60/**23.89**	20.49/**21.84**	22.72/**23.22**
Radial	ISTA	19.71/**20.79**	20.53/**22.15**	22.77/**23.60**	24.36/**25.22**	22.95/**24.88**	25.19/**26.11**
	ADMM	19.71/**20.79**	20.55/**22.15**	22.78/**23.62**	24.39/**25.27**	22.95/**24.94**	26.44/**27.22**

Figure 2a and Figure 2c illustrate that severe aliasing arises under Cartesian sampling when the sampling rate is low. After applying DogRot, this aliasing is largely mitigated, as shown in Figure 2b and Figure 2d. Radial sampling inherently avoids strong aliasing, but DogRot still enhances reconstruction by recovering finer details in the MRI.

F.3.1 DOMINANT-SUPPORT SCREENING

Table 7 evaluates dominant-support overlap under Cartesian sampling using the largest 1% DCT coefficients. Both ISTA and ADMM improve at 20% and 30% sampling, consistent with the PSNR gains in Tables 2 and 6.

Table 7: Dominant-support overlap (%) under Cartesian sampling, using 1,000 independent mixers per image and averaging across six images. Better values for each solver are bold.

Sampling	ISTA		ADMM	
	Baseline	DogRot	Baseline	DogRot
20%	10.68	**25.19**	13.84	**27.27**
30%	18.39	**29.69**	19.46	**30.92**

F.3.2 Sampling-Rate Sweep

Table 3 in the main text reports radial sampling rates of 10%, 20%, 30%, 40%, and 50%. Each baseline/DogRot pair uses the same image and acquired measurements. Dominant-support overlap is computed from the largest 1% DCT coefficients, using the definition above. The corresponding MC values are reported in Table 1. Changing the sampling rate changes both the number of measurements and the operator geometry, so the transition in recovery performance is specific to this acquisition family rather than a universal MC or sampling-rate threshold.

F.3.3 Additional MRI Images

We also evaluate four images from the Augmented Alzheimer MRI Dataset, one from each diagnostic category. The dataset is online available at the Kaggle Augmented Alzheimer MRI Dataset (https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset). Table 8 reports reconstruction at 30% Cartesian sampling. Both PSNR and structural similarity index (SSIM) improve for each image. This small cross-dataset experiment tests reconstruction on additional anatomy examples, not diagnostic performance.

Table 8: Alzheimer MRI examples at 30% Cartesian sampling. Entries report baseline/DogRot, with better values in bold.

Image category	Solver	PSNR (dB)	SSIM
Very mild dementia	ISTA	25.82/**27.25**	0.32/**0.52**
	ADMM	26.06/**27.36**	0.32/**0.54**
Mild dementia	ISTA	26.09/**26.62**	0.33/**0.57**
	ADMM	26.12/**26.81**	0.34/**0.57**
Moderate dementia	ISTA	24.42/**26.17**	0.26/**0.48**
	ADMM	24.65/**26.19**	0.28/**0.49**
No dementia	ISTA	26.22/**26.45**	0.36/**0.46**
	ADMM	26.24/**26.46**	0.37/**0.46**

F.3.4 Dominant-Support Screening with VarNet

We additionally evaluate support screening in a separate structured-mixer VarNet experiment, distinct from the six-image PSNR comparison. Both models use DCT latent coefficients. The mixed forward operator is $\boldsymbol{M}\boldsymbol{F}\,\mathrm{IDCT}_2(\boldsymbol{Q}_{\mathrm{str}}\boldsymbol{z})$, with a shared 128×128 DogRot applied along one DCT axis. The experiment uses 256 training images and 64 test images from the glioma, meningioma, no-tumor, and pituitary categories. The paired models share the Cartesian mask, data split, initial weights, three cascades, and 20 training epochs. The mixed model uses one fixed mixer with $\sqrt{\epsilon} = 0.025$ during training and testing.

Table 9 reports the mean overlap between the 164 largest-magnitude auxiliary coefficients and the reference image's largest 1% DCT coefficients. Nominal sampling rates of 20% and 30% give actual rates of 20.31% and 29.69% after discretizing the mask. Hence, the structured-mixer VarNet experiment improves dominant-support overlap.

Table 9: Mean top-1% DCT-support overlap (%) over 64 test images in the structured-mixer VarNet experiment. Better values are bold.

Sampling rate	VarNet	DogRot + VarNet
20%	38.41	**46.28**
30%	43.07	**52.72**

F.4 Sensitivity to ϵ

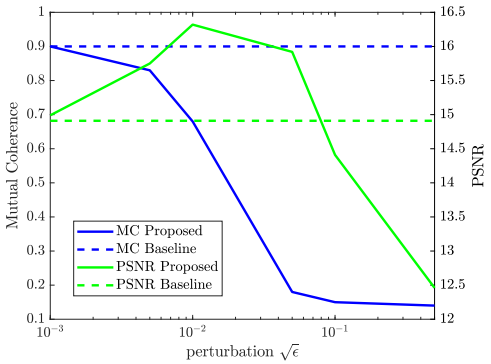


Figure 3: Sensitivity of MRI reconstruction to the DogRot perturbation level $\sqrt{\epsilon}$.

Figure 3 reports the sensitivity of MRI reconstruction to the perturbation level $\sqrt{\epsilon}$. As $\sqrt{\epsilon}$ increases, DogRot more strongly mixes the sensing columns and the MC of the effective sensing matrix decreases monotonically. However, the reconstruction quality does not improve monotonically. When $\sqrt{\epsilon}$ is too small, the mixer is close to the identity matrix, and the coherence reduction is insufficient to noticeably improve recovery. In the intermediate regime, the PSNR is consistently above the baseline, indicating that the coherence reduction is strong enough to improve support screening while the DogRot-induced perturbation remains controlled. When $\sqrt{\epsilon}$ becomes too large, the PSNR drops even though the MC continues to decrease, consistent with stronger leakage in the auxiliary signal and less reliable support screening. This trade-off is consistent with the theoretical role of ϵ: larger ϵ strengthens decorrelation, whereas smaller ϵ better preserves the dominant support. The value $\sqrt{\epsilon} = 0.025$ used in the experiments lies in the stable intermediate regime.

G Experimental results for Localized Statistical Channel Modeling

LSCM (Luo et al., 2023; Zhang et al., 2024; 2025) serves as a concrete example of CS with HIM. In this scenario, the reference signal received power (RSRP), measured using a limited number of beams, is exploited to estimate the angular power spectrum (APS), which is sparse in the spatial domain in high-frequency communication systems, at the base station (BS). Typically, fewer than 40 beams are available for RSRP measurements, while the APS is discretized over the spatial plane into $91 \times 72 = 6552$ grids. Because different beams can have similar responses, the corresponding 40×6552 sensing matrix is highly ill-conditioned.

G.1 Experimental Settings

Dataset: We use the sensing-matrix generation procedure in Zhang et al. (2024) for fifth-generation (5G) base-station (BS) configurations. The BS status is represented by a triple (x, y, z), where x and y denote the tilt and azimuth, respectively, and z specifies the beam pattern selected to satisfy coverage requirements. Random APS instances are generated to evaluate reconstruction performance under multiple BS configurations and path-number settings.
Metrics: MC measures the largest normalized column correlation. We report NMSE for coefficient

accuracy and Earth mover's distance (EMD) between unit-mass APS distributions on the 91×72 angular grid, using Euclidean transport cost. Support precision P@r is the fraction of selected grid points matched to a true point within radius r; P@0 requires an exact match. The modern-selector and algebraic-preconditioner experiments use maximum-cardinality one-to-one matching in the two-dimensional Euclidean grid. Each evaluated support contains s indices, so precision and recall coincide under the same matching rule.

Methods: We use nonnegative orthogonal matching pursuit (NNOMP), weighted nonnegative orthogonal matching pursuit (WNNOMP), and ADMM. WNNOMP is the application-specific weighted pursuit method of Zhang et al. (2024) for RSRP-to-APS recovery. Nonnegative refitting respects the physical power constraint. The main mixing scale is $\sqrt{\epsilon} = 0.025$. The GD projector is optimized for the operator to which it is applied, following Abolghasemi et al. (2010). Thus $\boldsymbol{PA}$ and $\boldsymbol{PAQ}$ denote separately optimized left-preconditioning pipelines, not necessarily the same numerical projector. Additional subset-selector and noisy experiments specify their own parameters below.

G.2 Reduction in mutual coherence

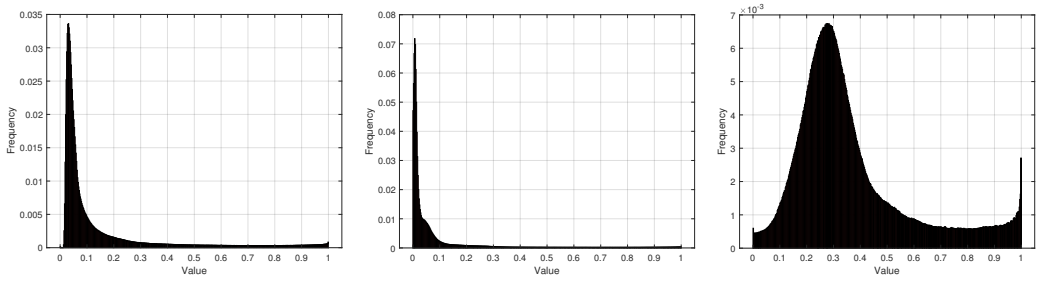
Table 10: Reduction in MC achieved through the proposed mixing strategy.

Matrix	(2, 243, 0)	(4, 240, 0)	(5, 240, 0)	(7, 127, 0)
$\boldsymbol{A}$	> 0.99	> 0.99	> 0.99	> 0.99
$\boldsymbol{AQ}$	0.90	0.90	0.90	0.90
$\boldsymbol{PAQ}$	0.79	0.78	0.79	0.79
Matrix	(8, 103, 0)	(8, 314, 0)	(6, 141, 1)	(6, 214, 1)
$\boldsymbol{A}$	> 0.99	> 0.99	> 0.99	> 0.99
$\boldsymbol{AQ}$	0.90	0.90	0.90	0.90
$\boldsymbol{PAQ}$	0.79	0.79	0.80	0.80

Table 10 summarizes the MC of the sensing matrices after applying DogRot with $\sqrt{\epsilon} = 0.025$. Here, $\boldsymbol{A}$ denotes the original sensing matrix, while $\boldsymbol{AQ}$ and $\boldsymbol{PAQ}$ denote the matrices after DogRot mixing and after an additional conventional projection, respectively. DogRot reduces MC even when the original columns are nearly parallel. Combining projection with DogRot gives an additional reduction, showing that the mixer can also be used with existing coherence-reduction techniques.

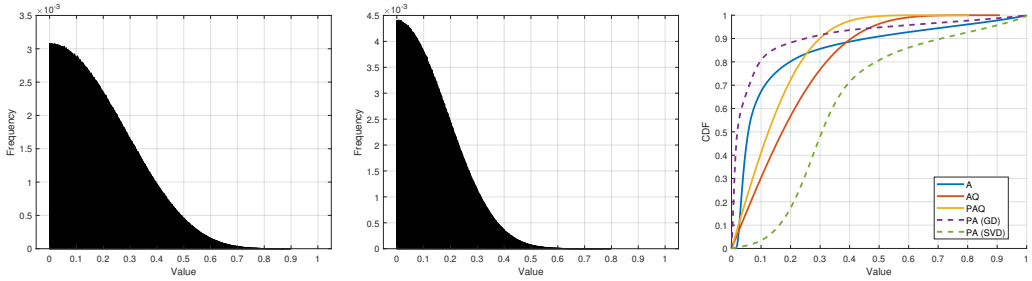
We use the $(2, 243, 0)$ sensing matrix as a representative HIM case. It is an extremely fat matrix with $\mu(\boldsymbol{A}) = 1 - 1 \times 10^{-7}$. Figure 4a shows the histogram of the off-diagonal entries of its Gram matrix $\boldsymbol{G}_A$, whose mass near one reflects strong inter-column correlations.

Figures 4b and 4c show the results of two existing projection methods. The GD-based projector (Abolghasemi et al., 2010) slightly reduces some correlations, but the heavy tail in $\boldsymbol{G}_D$ remains. The SVD-based projector (Elad, 2007) yields an even larger maximum correlation in this case. Quantitatively, the MC changes from $1 - 1 \times 10^{-7}$ to $1 - 5 \times 10^{-7}$ and $1 - 2 \times 10^{-10}$ after GD and SVD projection, respectively. These changes are marginal compared with the reductions obtained by



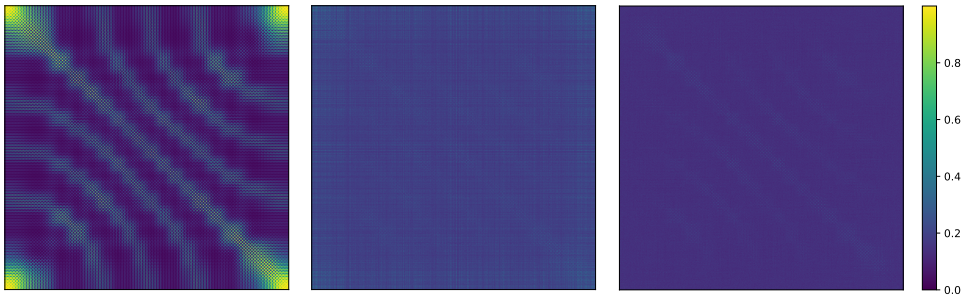
(a) Histogram of $\boldsymbol{G}_A$ (b) Histogram of GD-based $\boldsymbol{G}_D$ (c) Histogram of SVD-aided $\boldsymbol{G}_D$

Figure 4: Histograms of Gram matrices before and after conventional projection



(a) Histogram of G_{AQ} (b) Histogram of G_{PAQ} (c) Cumulative distributions

Figure 5: Histograms and cumulative distribution functions (CDFs) for the Gram matrices after mixing and projection



(a) Heatmap of G_A (b) Heatmap of G_{AQ} (c) Heatmap of G_{PAQ}

Figure 6: Heatmap for the Gram matrices before and after mixing and projection

DogRot in Table 10. Figure 5 compares the histograms and cumulative distribution functions (CDFs) of the Gram entries.

Figures 5 and 6 show the distribution and spatial pattern of normalized Gram entries. A CDF describes the bulk of column correlations, whereas MC is their largest absolute off-diagonal value. Improving most pairs need not remove a small number of nearly collinear pairs. In all eight configurations of Table 10, GD projection alone leaves $\mu(\boldsymbol{PA}) > 0.99$. DogRot reduces these extreme correlations, and subsequent projection reduces them further. Downstream NMSE and EMD provide a separate assessment of the recovery benefit.

G.3 Improvement in signal reconstruction

We next evaluate whether the coherence reduction translates into improved APS estimation. To cover different BS configurations and sparsity levels, we consider three representative BS parameter settings and two path-number scenarios. Table 11 reports the results for $s = |\mathcal{S}| = 5$, while Table 12 reports the corresponding results for $s = |\mathcal{S}| = 10$.

Table 11: APS reconstruction for $s = 5$. Each cell reports NMSE/EMD, with the lowest value for each row and metric in bold.

Configuration	Solver	$\boldsymbol{A}$	$\boldsymbol{PA}$	$\boldsymbol{AQ}$	$\boldsymbol{PAQ}$
$(2, 243, 0)$	NNOMP	0.97/29.47	0.96/27.63	**0.92**/24.02	**0.92/23.78**
	WNNOMP	42.75/23.81	23.15/23.69	0.97/23.76	**0.96/23.46**
	ADMM	0.96/27.17	0.90/24.23	**0.84/22.31**	**0.84/22.31**
$(4, 240, 0)$	NNOMP	0.98/29.78	0.97/27.98	0.94/24.12	**0.93/23.28**
	WNNOMP	48.15/23.15	23.69/**23.08**	**0.95**/23.13	**0.95/23.08**
	ADMM	0.88/27.13	0.86/25.97	**0.81/22.28**	**0.81/22.28**
$(6, 141, 1)$	NNOMP	0.99/29.36	0.98/28.76	0.95/24.02	**0.93/23.78**
	WNNOMP	141.48/27.14	50.12/27.02	**0.99**/25.92	**0.99/25.53**
	ADMM	0.92/22.53	0.92/20.39	**0.90/15.90**	**0.90/15.90**

Table 12: APS reconstruction for $s = 10$. Each cell reports NMSE/EMD, with the lowest value for each row and metric in bold.

Configuration	Solver	$\boldsymbol{A}$	$\boldsymbol{PA}$	$\boldsymbol{AQ}$	$\boldsymbol{PAQ}$
$(2, 243, 0)$	NNOMP	1.01/32.76	1.01/30.18	0.99/22.09	**0.98/20.83**
	WNNOMP	113.52/27.27	30.24/26.15	1.04/22.59	**1.03/22.28**
	ADMM	1.00/32.95	1.00/30.06	**0.97/25.55**	**0.97/25.55**
$(4, 240, 0)$	NNOMP	1.00/32.94	1.00/30.01	**0.98/21.63**	**0.98**/22.10
	WNNOMP	119.49/27.46	32.65/26.31	**1.01**/21.61	**1.01/21.55**
	ADMM	0.99/32.69	0.99/26.99	**0.97**/24.82	**0.97/24.46**
$(6, 141, 1)$	NNOMP	1.06/29.29	1.05/27.77	1.02/21.51	**1.00/21.04**
	WNNOMP	157.47/31.71	49.62/28.77	1.05/20.66	**1.03/20.34**
	ADMM	0.92/32.95	0.90/27.69	**0.79/25.55**	**0.79/25.55**

Tables 11 and 12 separate the effects of GD projection, DogRot, and their combination. GD alone can improve recovery, but the DogRot variants generally achieve lower spatial error and substantially reduce the large coefficient errors of WNNOMP. Applying GD after DogRot provides additional improvements in several settings. The full comparison thus supports using the two transformations together rather than treating them as interchangeable.

Spatial accuracy and amplitude accuracy remain distinct. In LSCM, columns with similar directions can have different norms, making path-strength estimation unstable even when the selected locations are close to the truth (Zhang et al., 2024). DogRot improves localization, as also measured by Table 13, while original-domain refitting limits the coefficient distortion introduced by the mixed representation. The APS visualization below shows the remaining amplitude bias.

G.3.1 Support Localization

Table 13 evaluates the selected indices before coefficient refitting. Each cell reports P@0/P@1 in percent, with 1,000 independently drawn mixers. Averaged over the three configurations and three solvers, the baseline/DogRot precisions are 5.50/12.28 for P@0 and 5.56/28.05 for P@1 at $s = 5$. At $s = 10$, they are 3.24/10.90 and 3.31/26.25. Neighborhood precision improves more than exact-grid precision, consistent with the lower EMD and the APS visualization.

Table 13: LSCM support localization. Entries report P@0/P@1 (%). Best values within each row and metric are bold.

Configuration	Solver	$\boldsymbol{A}$	$\boldsymbol{AQ}$	$\boldsymbol{PAQ}$
		$s = 5$		
(2,243,0)	NNOMP	5.26/5.30	10.32/24.18	**10.34/24.22**
	WNNOMP	5.78/5.80	**10.32**/24.44	**10.32/24.48**
	ADMM	6.22/6.30	**14.04/25.08**	13.98/24.96
(4,240,0)	NNOMP	5.34/5.42	13.88/**30.12**	**13.92/30.12**
	WNNOMP	5.62/5.74	**14.00**/29.54	13.96/**29.86**
	ADMM	6.08/6.14	13.82/**30.46**	**13.88**/30.42
(6,141,1)	NNOMP	4.98/5.02	11.28/**29.26**	**11.32/29.26**
	WNNOMP	5.14/5.22	**11.48**/29.30	11.42/**29.32**
	ADMM	5.08/5.08	11.36/**30.04**	**11.40**/29.96
		$s = 10$		
(2,243,0)	NNOMP	3.17/3.32	9.75/23.69	**9.77/23.72**
	WNNOMP	3.24/3.33	**9.83**/23.88	9.81/**23.91**
	ADMM	3.28/3.33	**10.12**/24.01	9.99/**24.09**
(4,240,0)	NNOMP	3.35/3.43	**11.85**/27.86	11.84/**27.93**
	WNNOMP	3.37/3.43	11.92/27.88	**11.99/27.89**
	ADMM	3.42/3.46	11.79/27.91	**11.99/28.11**
(6,141,1)	NNOMP	2.98/3.13	10.89/26.87	**10.93/27.01**
	WNNOMP	3.12/3.16	**10.95**/26.92	10.94/**26.98**
	ADMM	3.19/3.22	11.03/**27.25**	**11.07**/27.23

G.3.2 Modern Subset Selectors

We compare ABESS (Zhu et al., 2020; 2022) and OP-BESS (Zhu et al., 2025), both with and without DogRot. ABESS uses the official Python implementation; OP-BESS uses a Python port of the authors' fixed-cardinality algorithm. The experiment uses noiseless observations, nonzero APS amplitudes drawn uniformly from $[5, 10]$, and 30 paired trials for each of the three array configurations and each s. Every method ends with nonnegative least squares (NNLS) on the selected columns of the original matrix.

For the DogRot combinations, we mix the physical matrix $\boldsymbol{A}$ and normalize the columns of $\boldsymbol{AQ}$ for the subset search. If $\boldsymbol{D}_B$ contains these column norms and the search returns $\hat{\boldsymbol{u}}$, the mapped candidate is $\boldsymbol{Q}\boldsymbol{D}_B^{-1}\hat{\boldsymbol{u}}$. DogRot + ABESS uses auxiliary cardinality s, takes the s largest mapped entries, and refits without a second support search. DogRot + OP-BESS uses auxiliary cardinality 10, maps back, and uses the top-s candidate to initialize OP-BESS on the original system before refitting. The auxiliary support is therefore oversized for $s = 5$, but not for $s = 10$. A matched two-stage control sets $\boldsymbol{Q} = \boldsymbol{I}$ and retains the same initialization and refinement steps.

The mixing scales are $\sqrt{\epsilon} = 2 \times 10^{-5}$ for ABESS and 5×10^{-5} for OP-BESS. They were selected on configuration $(2, 243, 0)$ and fixed for the other two configurations. Table 14 averages all three, including the development configuration. Both selectors impose fixed cardinality in the auxiliary solve; the mapping and refinement treat this solve as a candidate generator rather than asserting that the exact transformed signal is sparse.

Table 14: Modern subset selectors on LSCM. P@0/P@1/P@2 are percentages. Bold and underlined values are best and second-best per sparsity and metric, with ties sharing a rank. The two-stage $\boldsymbol{Q} = \boldsymbol{I}$ row is a matched ablation constructed for this study.

s	Method	P@0	P@1	P@2	NMSE	EMD
5	ABESS	17.11	23.78	27.33	222.66	29.99
	DogRot + ABESS	16.67	22.89	27.78	<u>11.81</u>	29.95
	OP-BESS	34.89	40.44	44.44	103.16	24.61
	Two-stage OP-BESS, $\boldsymbol{Q} = \boldsymbol{I}$	<u>38.89</u>	<u>42.67</u>	<u>46.44</u>	138.01	<u>22.10</u>
	DogRot + OP-BESS	**39.11**	**44.00**	**48.44**	**8.27**	**20.73**
10	ABESS	6.00	10.56	14.67	518.65	32.36
	DogRot + ABESS	6.00	10.11	16.33	<u>72.77</u>	<u>30.99</u>
	OP-BESS	13.44	18.78	<u>22.00</u>	465.19	31.50
	Two-stage OP-BESS, $\boldsymbol{Q} = \boldsymbol{I}$	<u>13.56</u>	<u>18.89</u>	<u>22.00</u>	467.11	31.70
	DogRot + OP-BESS	**15.56**	**22.22**	**26.44**	**32.99**	**25.94**

DogRot + OP-BESS improves support localization and EMD over both OP-BESS and the matched identity control. Its most pronounced benefit is the reduction in amplitude error: NMSE falls from 103.16 to 8.27 at $s = 5$ and from 465.19 to 32.99 at $s = 10$. The identity control does not reproduce this reduction. DogRot + ABESS also reduces NMSE substantially, although its support precision is mixed. Large amplitude errors are therefore suppressed without a uniform improvement in every localization metric. Their remaining magnitude shows that stable, high-accuracy coefficient recovery is still difficult under this sensing geometry.

Support precision counts matched locations but does not control the conditioning or column scales of the selected submatrix. Nearly parallel beam responses can support a low-residual fit with severely distorted powers, even after nonnegative refitting. DogRot changes the support candidates before the original-domain fit. The joint support and reconstruction results quantify the resulting balance, rather than treating exact-grid precision alone as the recovery objective.

G.3.3 Algebraic Left Preconditioning and Measurement Noise

We further compare classical pseudo-inverse and regularized left transformations at a signal-to-noise ratio (SNR) of 20 dB. Let $\widetilde{\boldsymbol{A}}$ be column-normalized and $\boldsymbol{G} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}^T$. The evaluated matrices have full row rank, so $\boldsymbol{G}$ is invertible. The two transforms are

$$\boldsymbol{P}_0 = \widetilde{\boldsymbol{A}}^T\boldsymbol{G}^{-1}, \qquad \boldsymbol{P}_\lambda = \widetilde{\boldsymbol{A}}^T(\boldsymbol{G} + \lambda\boldsymbol{I})^{-1}.$$

Both act on the operator and observations. Columns of the effective operator are normalized before NNOMP screening, and the selected coefficients are refitted by NNLS on the original matrix. The regularization is $\lambda = \rho\ \mathrm{tr}(\boldsymbol{G})/p$. We select ρ from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ on $(2, 243, 0)$ using 200 validation signals per sparsity level and mean post-refit NMSE. The chosen $\rho = 0.1$ gives $\lambda = 16.38$ and is fixed on the remaining configurations.

Table 15: MC of the original, algebraically preconditioned, and DogRot sensing matrices. The lowest and second-lowest values per configuration are bold and underlined. Extra digits distinguish near-collinear operators.

Configuration	Original	Pseudo-inverse	Regularized	DogRot
$(2, 243, 0)$	0.99999988	<u>0.99999893</u>	0.99999934	**0.89815718**
$(4, 240, 0)$	0.99999982	<u>0.99999875</u>	0.99999923	**0.89752768**
$(6, 141, 1)$	0.99999999	<u>0.99999923</u>	0.99999940	**0.90639657**

Table 15 shows that algebraic preconditioning leaves the largest normalized column correlations close to one, although it changes measurement-space weighting. DogRot yields a much larger reduction in MC.

Table 16: NNOMP with original-domain NNLS refitting at 20 dB SNR. Each row averages 2,000 trials from the two configurations not used to select ρ. Bold and underlined values are best and second-best per sparsity and metric.

s	Method	P@0 (%)	P@1 (%)	NMSE	EMD
5	Original	5.50	5.56	<u>1.18</u>	34.31
	Pseudo-inverse	4.99	5.38	201.70	35.80
	Regularized	<u>6.12</u>	<u>9.94</u>	94.89	<u>33.52</u>
	DogRot	**11.12**	**23.68**	**1.001**	**27.82**
10	Original	3.06	3.83	<u>1.09</u>	33.86
	Pseudo-inverse	2.57	3.79	300.72	34.07
	Regularized	<u>4.62</u>	<u>5.24</u>	215.57	<u>32.43</u>
	DogRot	**8.62**	**11.24**	**1.08**	**26.83**

At 20 dB, the unregularized pseudo-inverse worsens all four metrics relative to the original system. Regularization improves localization and reduces the amplitude error relative to the pseudo-inverse, but its NMSE remains much larger than the original baseline. DogRot attains the best localization and EMD in this comparison while retaining much smaller amplitude error. These results complement the optimized GD projection in Tables 11–12: left weighting and right mixing have different effects, and an additional optimized projector can improve the DogRot system.

G.4 Visualization of the effect of the proposed DogRot

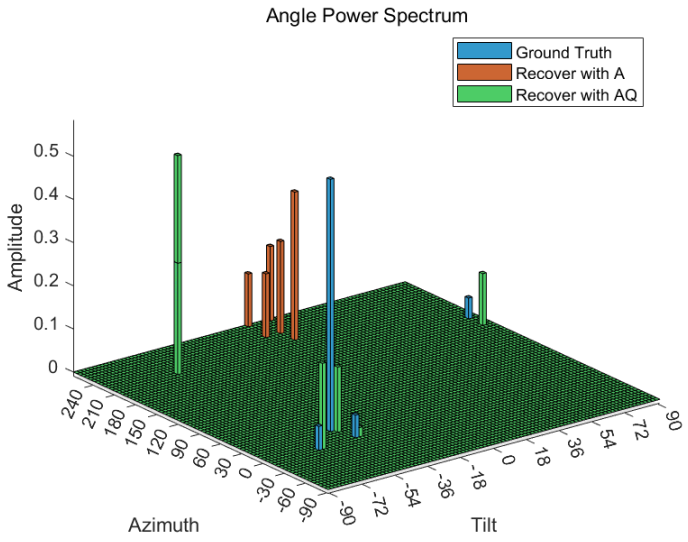


Figure 7: Visualization of recovered APS before and after applying DogRot.

Figure 7 provides a representative visualization of the recovered APS. When the original sensing matrix $\boldsymbol{A}$ is used directly, the recovered spectrum is dominated by displaced or spurious peaks and fails to localize the true angular support in a meaningful way. After applying DogRot, the dominant peaks move much closer to the ground-truth locations, and the recovered APS preserves the main sparse structure of the target spectrum. This visual behavior is consistent with the EMD reductions reported in Tables 11 and 12: DogRot primarily improves the spatial localization of the sparse support. The remaining amplitude mismatch is also expected from the preceding analysis, since column-norm fluctuations in LSCM can still distort path-strength estimation even when the support is substantially improved. Thus, **the effect of DogRot should be interpreted as a qualitative transition from**

near-failed recovery under the original HIM to a lossy but informative reconstruction whose support structure is already close to the ground truth.

G.5 SENSITIVITY TO ϵ

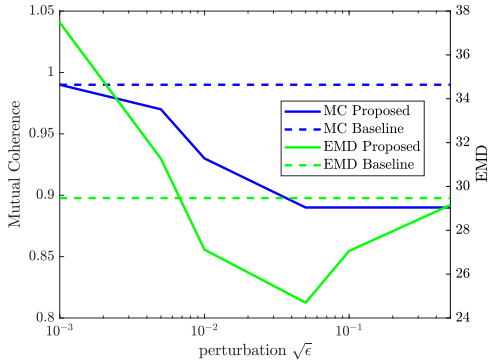


Figure 8: Sensitivity of LSCM APS estimation to the DogRot perturbation level $\sqrt{\epsilon}$.

Figure 8 shows the corresponding sensitivity analysis for LSCM. Similar to the MRI case, increasing $\sqrt{\epsilon}$ reduces the MC of the mixed sensing matrix, confirming that stronger random mixing more aggressively decorrelates the highly correlated beam responses. The EMD curve, however, exhibits a U-shaped behavior. For very small $\sqrt{\epsilon}$, the DogRot matrix is too close to the identity matrix, so the recovered APS remains close to the baseline result and the support localization is still poor. As $\sqrt{\epsilon}$ enters an intermediate range, EMD decreases substantially, indicating that the reduced coherence translates into more accurate spatial support recovery. When $\sqrt{\epsilon}$ becomes too large, EMD increases again, consistent with stronger off-support leakage and less reliable screening. These results again demonstrate that ϵ controls a trade-off between coherence reduction and support preservation. The choice $\sqrt{\epsilon} = 0.025$ used in the LSCM experiments falls in the low-EMD region and provides a favorable balance between these two effects.

H GENE-BASED DISEASE CLASSIFICATION

Another representative example of CS with HIM is disease classification based on gene expression monitoring with deoxyribonucleic acid (DNA) microarrays (Golub et al., 1999; Zou & Hastie, 2005; Li & Li, 2008). A typical microarray dataset contains thousands of genes but fewer than 100 samples, with strongly correlated expression profiles. Many existing works address this issue by feature elimination, relying on expert knowledge or statistical tools (Guyon et al., 2002) to reduce the number of genes used in the final model. Such preprocessing may remove informative genes. We further evaluate DogRot in this challenging application.

H.1 EXPERIMENTAL SETTINGS

Dataset: We evaluate DogRot using the leukemia dataset, which contains 7129 genes and 72 samples (Golub et al., 1999). Many genes exhibit similar expression patterns, leading to strong correlations and high MC in the expression matrix. This dataset, publicly available at `https://github.com/pulkitmehta/Molecular-Cancer-Classification-by-ML`, includes two classes of cancer, i.e., acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The objective is to construct a diagnostic rule based on the expression levels of the 7129 genes to predict the leukemia type. We combine all 72 samples and repeatedly draw stratified random training sets of sizes 15, 20, 25, and 30, using the remaining samples for testing. Baseline and DogRot use the same split in each run, and we report mean accuracy across runs. Preprocessing and parameter selection use training data only.

Metric: We use MC $\mu(\boldsymbol{A}) \in [0, 1]$ to quantify the conditioning of the sensing matrix $\boldsymbol{A}$, and classification accuracy to indirectly reflect the quality of gene selection.

Methods: We use two classical regression models: the least absolute shrinkage and selection operator

(LASSO), which uses ℓ_1 regularization, and elastic net (Zou & Hastie, 2005), which combines both ℓ_1 and ℓ_2 regularization. $\sqrt{\epsilon}$ is set to 0.025 in this experiment.

H.2 Reduction in mutual coherence

Table 17: Reduction in MC achieved through the proposed mixing with different numbers of training samples

Matrix	15	20	25	30
$\boldsymbol{A}$	> 0.99	> 0.99	> 0.99	> 0.99
$\boldsymbol{AQ}$	0.96	0.94	0.91	0.89

Table 17 records the reduction in MC of the expression matrix achieved by the proposed method. Notably, even in the challenging case of a 15×7129 matrix $\boldsymbol{A}$, the MC $\mu(\boldsymbol{A})$ can still be effectively reduced. As the number of training samples increases, the reduction becomes more pronounced.

H.3 Improvement in classification accuracy

Table 5 in the main text presents the classification accuracy obtained using two regression models. The results demonstrate that the proposed method substantially enhances classification accuracy compared with standard regression, with the improvement being especially remarkable when the number of training samples is small. Furthermore, the elastic net consistently outperforms LASSO, owing to its ability to enforce sparsity while also performing simultaneous shrinkage and selection. Its grouping tendency is useful when expression profiles are correlated, although predictive accuracy alone does not establish a biological mechanism.

Subset-based classification. As an additional classifier, Logistic ABESS (Zhu et al., 2022) selects a subset of predictors through a cardinality-constrained logistic fit. With $\sqrt{\epsilon} = 0.005$, its accuracies within the DogRot pipeline are 83.49%, 85.95%, 87.26%, and 89.43% for 15, 20, 25, and 30 training samples, respectively. These results extend the evaluation beyond penalized regression.

Gene-selection stability. We compare the sets of original genes selected across these repeated runs to assess the stability of the gene-selection pipeline. For two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, Jaccard similarity is $|\mathcal{S}_1 \cap \mathcal{S}_2|/|\mathcal{S}_1 \cup \mathcal{S}_2|$. Table 18 summarizes the results. LASSO remains essentially unchanged, while Elastic Net improves at three sample sizes. The absolute similarities are low, reflecting the difficulty of reproducible feature selection with thousands of correlated genes and few samples. As the dataset does not provide a ground-truth causal-gene support, this diagnostic measures stability rather than biological support accuracy.

Table 18: Jaccard similarity of selected-gene sets. Entries report baseline/DogRot. Strictly higher values within each pair are bold.

Training samples	15	20	25	30
LASSO	0.02/0.02	**0.04**/0.03	0.06/0.06	0.08/0.08
Elastic Net	0.03/**0.04**	0.06/0.06	0.08/**0.09**	0.11/**0.13**

H.4 Sensitivity to ϵ

Table 19 varies the mixing level at 30 training samples. Both classifiers reach their highest accuracy at $\sqrt{\epsilon} = 0.025$, the main experimental setting. Relative to no mixing, the gains are 3.92 percentage points for LASSO and 7.90 for Elastic Net. Stronger mixing eventually reduces accuracy below the baseline. This non-monotonic response is consistent with the decorrelation–leakage trade-off, and makes reconstruction or classification performance a necessary complement to MC when selecting the perturbation level.

Table 19: Gene classification sensitivity at 30 training samples. Accuracies are percentages. Bold and underlined values are best and second-best per classifier, with ties sharing a rank.

$\sqrt{\epsilon}$	LASSO	Elastic Net
0	83.21	86.72
0.005	84.55	89.51
0.010	<u>85.86</u>	<u>92.08</u>
0.025	**87.13**	**94.62**
0.050	<u>85.86</u>	91.85
0.075	83.65	82.37
0.100	72.96	79.46