A Debiased LASSO Estimator for Design Matrices with Non-zero Mean Elements

Ashish Zantye[1] **Ajit Rajwade**[1] **Radhendushka Srivastava**[2]

[1]Department of Computer Science and Engineering, IIT Bombay, India
[2]Department of Mathematics, IIT Bombay, India

Abstract

The LASSO estimator has been popular in diverse fields such as statistics, bio-informatics and compressed sensing for estimation of sparse signals from under-sampled measurements. As the LASSO does not provide confidence intervals for the estimates of the individual elements, techniques such as the debiased LASSO have been developed that provide such confidence intervals and also mitigate the inherent bias in the LASSO estimates. However the theoretical guarantees underlying the debiased LASSO have been developed for a design or measurement matrix whose elements have zero mean. On the other hand, many commonly used measurement matrices in sparse regression applications in optical imaging or pooled biological testing involve non-negative, particularly binary, matrices whose elements have a non-zero mean. In this paper, we propose a debiased LASSO technique based on computing differences between different row subsets of the underlying design matrix iteratively which effectively converts the design matrix into one with zero-mean elements. We show analytically that this results in lower variances for the individual elements of the debiased LASSO estimate as compared to the standard debiased LASSO. We also provide numerical results to support our theory.

1 INTRODUCTION

The LASSO estimator has been widely used in diverse fields for estimating a sparse signal vector $\boldsymbol{x}^* \in \mathbb{R}^p$ from $n \ll p$ measurements assembled in a vector $\boldsymbol{z}$ such that $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}^* + \boldsymbol{\eta}$, where $\boldsymbol{B}$ is a design matrix (also called sensing matrix) of size $n \times p$, and $\boldsymbol{\eta}$ is a zero-mean noise vector of size $n \times 1$, whose elements are independent and typically modeled to belong to $\mathcal{N}(0, \sigma^2)$. That is, for all $i \in [n]$ where $[n] := \{1, 2, ..., n\}$, we have $z_i = \boldsymbol{b}_{i.}\boldsymbol{x}^* + \eta_i$ where $\boldsymbol{b}_{i.}$ is the ith row of $\boldsymbol{B}$. The LASSO reconstruction proceeds by minimizing the cost function $J(\boldsymbol{x}; \lambda) := \|\boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$. If $s = \|\boldsymbol{x}^*\|_0$, if $\boldsymbol{B}$ obeys the so-called restricted eigenvalue condition (REC) Bickel et al. [2009], and if $\lambda \geq \|\boldsymbol{B}^T\boldsymbol{\eta}\|_\infty$ (i.e. $\lambda \sim \mathcal{O}(\sigma\sqrt{\frac{\log p}{n}})$ for i.i.d. Gaussian noise), then it is established that the LASSO estimate $\hat{\boldsymbol{x}}_\lambda$ obeys upper bounds of the form $\|\hat{\boldsymbol{x}}_\lambda - \boldsymbol{x}^*\|_2 \leq \mathcal{O}\left(\sigma\sqrt{\frac{s \log p}{n}}\right)$ with high probability Hastie et al. [2015]. It has been shown that zero-mean sub-Gaussian random matrices obey the REC with high probability given a sufficiently large number of rows Zhou [2009], Raskutti et al. [2010].

The debiased LASSO: However, the LASSO is known to provide biased estimates, that is $E(\hat{\boldsymbol{x}}_\lambda) \neq \boldsymbol{x}^*$ where the expectation is taken over different noise instances. Furthermore, the LASSO provides no mechanism regarding the accuracy of *individual* elements of $\hat{\boldsymbol{x}}_\lambda$, or any confidence intervals for its elements. In works such as Javanmard and Montanari [2014a], Zhang and Zhang [2014], van de Geer et al. [2014], attempts have to made to reduce this bias by employing a so-called 'debiasing' mechanism. These works replace the LASSO estimate $\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}$ by a 'debiased' estimate $\hat{\boldsymbol{x}}_{\boldsymbol{d}}$ expressed mathematically as follows:

$$\hat{\boldsymbol{x}}_{\boldsymbol{d}} = \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}} + \frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top(\boldsymbol{z} - \boldsymbol{B}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}). \tag{1}$$

Here $\boldsymbol{M}$ is an approximate inverse of $\hat{\boldsymbol{\Sigma}} := \boldsymbol{B}^\top\boldsymbol{B}/n$, in a precise sense defined in (3). The second term on the RHS of (1) can be interpreted to be a kind of adjustment to $\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}$ with a weighted sum of the entries of the residual vector $(\boldsymbol{z} - \boldsymbol{B}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}})$. Substituting $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}^* + \boldsymbol{\eta}$ into (1) and treating $\frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top\boldsymbol{B}$ as approximately equal to the identity matrix, yields:

$$\hat{\boldsymbol{x}}_{\boldsymbol{d}} = \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}} + \frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top(\boldsymbol{B}\boldsymbol{x}^* + \boldsymbol{\eta} - \boldsymbol{B}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) \approx \boldsymbol{x}^* + \frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top\boldsymbol{\eta}, \tag{2}$$

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:7977–7991.

which is referred to as a debiased estimate, because $E(\hat{\boldsymbol{x}}_{\boldsymbol{d}}) \approx \boldsymbol{x}^*$.

Javanmard and Montanari [2014a] provides an algorithm to construct the matrix $\boldsymbol{M}$ with a view to minimizing the variance of every element of $\frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top\boldsymbol{\eta}$ while simultaneously decreasing the bias. In order to obtain $\boldsymbol{M}$, the following convex optimization problem is solved:

$$\begin{aligned} &\forall j \in [p], \text{minimize}_{\boldsymbol{m}_{\cdot j}} \;\; \boldsymbol{m}_{\cdot j}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{m}_{\cdot j} \\ &\text{subject to} \;\; \|\hat{\boldsymbol{\Sigma}}\boldsymbol{m}_{\cdot j} - \boldsymbol{e}_{\cdot j}\|_\infty \leq \mu, \end{aligned} \tag{3}$$

where $\boldsymbol{m}_{\cdot j}$ is the jth column of $\boldsymbol{M}$, $\mu > 0$ is a slackness parameter that determines the degree of approximation of the approximate inverse and that also controls the bias. The constraint on $\boldsymbol{M}$ is designed to allow for small-valued or negligible asymptotic bias of $\hat{\boldsymbol{x}}_{\boldsymbol{d}}$ given appropriate requirements that $\boldsymbol{\Sigma} := E(\hat{\boldsymbol{\Sigma}})$, $n,\ p, s$ and μ must satisfy.

Solving for $\boldsymbol{M}$ is computationally expensive, hence some works such as Javanmard and Montanari [2014b] use $\boldsymbol{M} = \boldsymbol{I}$. Likewise, the work in Banerjee et al. [2026] reparameterizes the optimization problem given in (3) using a weights matrix $\boldsymbol{W}^\top := \boldsymbol{M}\boldsymbol{B}^\top$ and provides an exact closed form solution for $\boldsymbol{W}$ as weighted columns of $\boldsymbol{B}$. This reparameterization leads to a fast debiasing of the LASSO estimator, retaining the original properties of the debiased LASSO for design matrices whose rows have uncorrelated or weakly correlated elements. In this setup, the debiased LASSO estimate given in (1) reduces to

$$\hat{\boldsymbol{x}}_{\boldsymbol{d}} = \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}} + \frac{1}{n}\boldsymbol{W}^\top(\boldsymbol{z} - \boldsymbol{B}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}), \tag{4}$$

where

$$\boldsymbol{W} := (\boldsymbol{w}_{\cdot 1}|\boldsymbol{w}_{\cdot 2}|...|\boldsymbol{w}_{\cdot p}); \boldsymbol{w}_{\cdot j} = n(1-\mu)\boldsymbol{b}_{\cdot j}/\|\boldsymbol{b}_{\cdot j}\|_2^2. \tag{5}$$

A detailed derivation of the above result is presented in Banerjee et al. [2026].

The debiased estimate $\hat{\boldsymbol{x}}_{\boldsymbol{d}}$ can be decomposed as:

$$\hat{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}^* = \frac{1}{n}\boldsymbol{M}\boldsymbol{B}^\top\boldsymbol{\eta} + (\boldsymbol{M}\hat{\boldsymbol{\Sigma}} - \boldsymbol{I}_{\boldsymbol{n}})(\boldsymbol{x}^* - \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}). \tag{6}$$

which can written as

$$\sqrt{n}(\hat{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}^*) = \frac{1}{\sqrt{n}}\boldsymbol{W}^\top\boldsymbol{\eta} + \sqrt{n}(\frac{1}{n}\boldsymbol{W}^\top\boldsymbol{B} - \boldsymbol{I}_{\boldsymbol{n}})(\boldsymbol{x}^* - \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}). \tag{7}$$

by substituting $\boldsymbol{M}\boldsymbol{B}^\top$ with $\boldsymbol{W}^\top$. The second term on the right hand side of (7) is the bias of $\hat{\boldsymbol{x}}_{\boldsymbol{d}}$ which is shown to be negligible if $n \geq \mathcal{O}(s^2 \log^2 p)$ where $s = \|\boldsymbol{x}^*\|_0$ Javanmard and Montanari [2014a], whereas the variance-covariance matrix of the first term is $\boldsymbol{W}^\top\boldsymbol{W}\sigma^2/n$. In other words, this means that for every $j \in [p]$, $\hat{x}_d(j)$ has a distribution which is approximately $\mathcal{N}(x_j^*, \sigma^2[\boldsymbol{W}^\top\boldsymbol{W}]_{jj}/n)$. This can be used to produce confidence intervals for x_j^*, and the cost function in (3) is aimed to minimize the variance of this estimate since $\boldsymbol{m}_{\cdot j}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{m}_{\cdot j} = [\boldsymbol{W}^\top\boldsymbol{W}]_{jj}$.

The non-zero mean problem: The debiasing theory from Javanmard and Montanari [2014a], van de Geer et al. [2014], Zhang and Zhang [2014], as well as later developments in the debiasing literature such as Celentano et al. [2023], Celentano and Montanari [2024], van de Geer [2019], require that the elements of $\boldsymbol{B}$ obey the zero-mean property. However in applications such as compressive optical imaging Hunt et al. [2018], Li and Raskutti [2018] and pooled biological testing Ghosh et al. [2021], the underlying measurement matrix is often designed to be random Bernoulli, which does not satisfy the mean-zero condition. For example, the celebrated Rice Single Pixel Camera for compressive optical imaging Duarte et al. [2008], Gibson et al. [2020] uses matrices with entries chosen to be 0 or 1 with 0.5 probability, implemented via an array of mirrors that modulate the incident signal with a binary pattern. Furthermore, random Bernoulli matrices with different probability values are also quite common in pooled testing based on compressed sensing Ghosh et al. [2021].

In this paper, we propose a technique to perform LASSO debiasing for matrices that do not obey the zero-mean condition. For this, we appropriately scale and center $\boldsymbol{B}$ to obtain the centered matrix $\boldsymbol{A}$ by computing differences between different randomly chosen row-subsets of $\boldsymbol{B}$. Correspondingly, this converts the original measurements $\boldsymbol{z}$ into 'centered' measurements $\boldsymbol{y}$. The final debiased estimate of $\boldsymbol{x}^*$ can be computed by averaging of different debiased estimates, each obtained by applying the debiased LASSO using a particular set of row-subsets. We prove that the final debiased estimate thus derived has lower element-wise variances when multiple subsets are used, as compared to that produced by the original technique in Javanmard and Montanari [2014a]. We support our theory with numerical results regarding the reconstruction accuracy, bias and variance of the debiased estimates.

2 DEBIASING LASSO USING CENTERING OPERATION

Suppose that the rows of the sensing matrix $\boldsymbol{B}$ are independent and identically distributed. The different elements within a row may or may not be independent of each other. Let us consider the case that the elements of $\boldsymbol{B}$, that is b_{ij} for $i \in [n] := \{1, 2, ..., n\}$ and $j \in [p]$, have mean θ. By differencing two rows of $\boldsymbol{B}$, one can transform the sensing matrix into a zero-mean form. Here, we will present this centering mechanism and the properties of the associated debiased LASSO estimator.

2.1 PROPOSED APPROACH: DIFFERENCING BETWEEN ROWS OF B

Let π_k be a random permutation of the set $[n]$ and let $\pi_k(i)$ be its i^{th} element. Let $m := \lfloor n/2 \rfloor$. We consider the following centering operation:

$$\boldsymbol{a}_{i\cdot}^{(\pi_k)} := \boldsymbol{b}_{\pi_k(i)\cdot} - \boldsymbol{b}_{\pi_k(m+i)\cdot}, \ \forall i \in [m] \quad (8)$$

$$y_i^{(\pi_k)} := z_{\pi_k(i)} - z_{\pi_k(m+i)}, \quad \forall i \in [m], \quad (9)$$

where $\boldsymbol{b}_{l\cdot}$ and z_l denote the l^{th} row of $\boldsymbol{B}$ and the l^{th} element of $\boldsymbol{z}$, respectively. These together with the forward model for $\boldsymbol{z}$ produce the following:

$$\boldsymbol{y}^{(\pi_k)} = \boldsymbol{A}^{(\pi_k)}\boldsymbol{x}^* + \boldsymbol{\epsilon}^{(\pi_k)}, \quad (10)$$

where $\boldsymbol{y}^{(\pi_k)} := [y_1^{(\pi_k)}, y_2^{(\pi_k)}, \ldots, y_m^{(\pi_k)}]^\top$, $\boldsymbol{A}^{(\pi_k)}$ is the centered sensing matrix with $\boldsymbol{a}_{i\cdot}^{(\pi_k)}$ as the i^{th} row created as above from (8). Further, the i^{th} element of the transformed noise vector $\boldsymbol{\epsilon}^{(\pi_k)}$ is given as follows:

$$\epsilon_i^{(\pi_k)} := \eta_{\pi_k(i)} - \eta_{\pi_k(m+i)}, \quad \forall i \in [m]. \quad (11)$$

As π_k is a permutation of $[n]$ and the elements of $\boldsymbol{\eta}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$, the elements of $\boldsymbol{\epsilon}^{(\pi_k)}$ are i.i.d. $\mathcal{N}(0, 2\sigma^2)$. By using (10) and (4), a debiased LASSO estimate of $\boldsymbol{x}^*$ is given by

$$\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)} = \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}{}^{(\pi_k)} + \frac{1}{m}\boldsymbol{W}^{(\pi_k)\top}\left(\boldsymbol{y}^{(\pi_k)} - \boldsymbol{A}^{(\pi_k)}\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}{}^{(\pi_k)}\right), \quad (12)$$

where the debiasing matrix $\boldsymbol{W}^{(\pi_k)}$ is constructed as follows Banerjee et al. [2026], Javanmard and Montanari [2014b]:

$$\boldsymbol{W}^{(\pi_k)} := (\boldsymbol{w}_{\cdot 1}|\boldsymbol{w}_{\cdot 2}|...|\boldsymbol{w}_{\cdot p}),$$
$$\boldsymbol{w}_{\cdot j} := m(1-\mu)\boldsymbol{a}_{\cdot j}^{(\pi_k)}/\|\boldsymbol{a}_{\cdot j}^{(\pi_k)}\|_2^2, \quad (13)$$

where $\boldsymbol{a}_{\cdot j}^{(\pi_k)}$ stands for the j^{th} column of $\boldsymbol{A}^{(\pi_k)}$. In (12), the corresponding LASSO estimate $\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}{}^{(\pi_k)}$ of $\boldsymbol{x}^*$ is obtained as follows:

$$\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}{}^{(\pi_k)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{y}^{(\pi_k)} - \boldsymbol{A}^{(\pi_k)}\boldsymbol{x}\right\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \quad (14)$$

The debiased LASSO estimate $\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}$ is based on the m-dimensional measurement vector $\boldsymbol{y}^{(\pi_k)}$, that is contaminated by noise $\boldsymbol{\epsilon}^{(\pi_k)}$ whose elements have twice the variance of those of $\boldsymbol{\eta}$. Therefore, it may appear that the centering operation reduces the number of measurements, and that too with a higher noise variance. We now point out that if *many* different (say K) permutations $\{\pi_k\}_{k=1}^K$ are used, then the centering operations *collectively* utilize *all* the available measurements in $\boldsymbol{z}$ in a different way and actually do *not* discard any measurements. In particular, the random permutation vector π_k introduces an additional source of variation in the debiased LASSO estimate $\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}$. This additional variation is minimized by aggregating the estimates $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$ over several independent random permutations via an averaging operation.

Let $\{\pi_k\}_{k=1}^K$ be K independent random permutations of the set $[n]$. Let $\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}$, for $k \in [K]$, be the debiased LASSO estimate of $\boldsymbol{x}^*$ obtained from (12) via the measurement vector $\boldsymbol{y}^{(\pi_k)}$ corresponding to the matrix $\boldsymbol{A}^{\pi(k)}$ for permutation π_k. Since the rows of $\boldsymbol{B}$ are i.i.d., then the rows of $\boldsymbol{A}^{(\pi_k)}$ are also i.i.d. In such a case, note that $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$ are identically distributed but not independent. Moreover, the estimates $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$ constitute a sequence of *exchangeable* random vectors since the permutations $\{\pi_k\}_{k=1}^K$ are i.i.d.; that is the joint distribution of the different $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$ does not change if their positions are permuted. The average of exchangeable random variables is known to have a smaller variance than that of the individual variables, in spite of a non-zero correlation among them (see Lemma SL.2 of the Appendix). Therefore, we propose to estimate $\boldsymbol{x}^*$ by the average of these debiased LASSO estimates, i.e.,

$$\bar{\boldsymbol{x}}_{\boldsymbol{d}} = \frac{1}{K}\sum_{k=1}^K \hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}. \quad (15)$$

If $K = 1$, that is only one permutation is used, then our presented technique reduces to the original technique from Javanmard and Montanari [2014a], albeit after a centering operation performed on $\boldsymbol{B}, \boldsymbol{z}$ to yield $\boldsymbol{A}, \boldsymbol{y}$.

We now have the following theoretical result, which is proved in Appendix S.5.2.

Lemma 1 *Let the rows of the sensing matrix $\boldsymbol{B}$ be independent and identically distributed. Let $\boldsymbol{A}^{(\pi_k)}$ be constructed from $\boldsymbol{B}$ as in* (8)*, and let the debiasing matrix $\boldsymbol{W}^{(\pi_k)}$ be constructed from $\boldsymbol{A}^{(\pi_k)}$ via* (13)*.*
(a) Then the averaged debiased LASSO *estimate $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ from* (15) *can be decomposed as follows, where $\hat{\boldsymbol{x}}_{\lambda}^{(\pi_k)}$ is the* LASSO *estimate from $\boldsymbol{y}^{(\pi_k)}, \boldsymbol{A}^{(\pi_k)}$:*

$$\sqrt{m}(\bar{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}^*) = \frac{1}{K}\sum_{k=1}^K \underbrace{\frac{1}{\sqrt{m}}\boldsymbol{W}^{(\pi_k)\top}\boldsymbol{\epsilon}^{(\pi_k)}}_{\boldsymbol{h}_k} + \frac{1}{K}\sum_{k=1}^K \boldsymbol{\Delta}_m^{(\pi_k)}, \quad (16)$$

where

$$\boldsymbol{\Delta}_m^{(\pi_k)} := \sqrt{m}\left(\frac{1}{m}\boldsymbol{W}^{(\pi_k)\top}\boldsymbol{A}^{(\pi_k)} - \boldsymbol{I}_p\right)\left(\boldsymbol{x}^* - \hat{\boldsymbol{x}}_{\lambda}^{(\pi_k)}\right), \quad (17)$$

and $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix.
(b) Further, $\{\boldsymbol{h}_k\}_{k\geq 1}^K$ (defined above) are exchangeable random vectors.
(c) Moreover, let $\boldsymbol{\Sigma}_{\boldsymbol{A}}$ be the covariance matrix of the rows of $\boldsymbol{A}^{(\pi_k)}$. If the least singular value of $\boldsymbol{\Sigma}_{\boldsymbol{A}}$ is greater than zero, if its maximum singular value is bounded, if $\boldsymbol{A}^{(\pi_k)}\boldsymbol{\Sigma}_{\boldsymbol{A}}^{-1/2}$ has independent sub-Gaussian rows, and if $m \geq \mathcal{O}(s\log p)$

with $s := \|\boldsymbol{x}^*\|_0$, then we have:

$$\frac{1}{K}\sum_{k=1}^{K}\|\boldsymbol{\Delta}_m^{(\pi_k)}\|_\infty = O_P\left(\frac{s\log p}{\sqrt{m}}\right).\blacksquare$$

Remarks on Lemma 1:

1. In (16), $\sqrt{m}(\bar{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}^*)$ is decomposed into two parts. The first term is purely Gaussian since $\{\boldsymbol{\epsilon}^{(\pi_k)}\}_{k=1}^K$ are jointly Gaussian (via the Gaussianity of $\boldsymbol{\eta}$). The second part corresponds to the bias term, which turns out to be small as long as $m \geq \mathcal{O}(s^2\log^2 p)$ using [Javanmard and Montanari, 2014a, Theorem 8].
2. Part (c) of this lemma states that the absolute bias is bounded by a term which converges to 0 as m (or n) increases. The first column of Fig. 3 shows empirically that the spread of the bias term (across different coordinates in $[1, ..., p]$) reduces as K increases.

We now present an important distributional property of $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ via the following theorem, which is proved in Appendix. S.5.3.

Theorem 1 *Let the conditions of Lemma 1 hold. For $j \in [p]$, the distribution of the j^{th} coordinate of $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ converges as follows:*

$$P\left(\frac{\sqrt{m}\left(\bar{x}_{d,j} - x_j^*\right)}{\tau_j} \leq t\right) \longrightarrow \Phi(t) \text{ as } m \to \infty \quad (18)$$

with $\Phi(\cdot)$ denoting the cumulative distribution function (CDF) of standard normal random variable, $t \in \mathbb{R}$, and

$$\tau_j^2 := \frac{1}{K^2}\sum_{k=1}^{K}\frac{2\sigma^2}{m}\boldsymbol{w}_{j\cdot}^{(\pi_k)\top}\boldsymbol{w}_{j\cdot}^{(\pi_k)} + \frac{2}{K^2}\sum_{k=1}^{K}\sum_{k'<k}\frac{1}{m}\boldsymbol{w}_{j\cdot}^{(\pi_k)\top}\boldsymbol{\mathcal{C}}^{(\pi_k,\pi_{k'})}\boldsymbol{w}_{j\cdot}^{(\pi_{k'})} \quad (19)$$

where $\boldsymbol{w}_{j\cdot}^{(\pi_k)\top}$ and $\boldsymbol{w}_{j\cdot}^{(\pi_{k'})\top}$ respectively denote the j^{th} row of $\boldsymbol{W}^{(\pi_k)\top}$ and $\boldsymbol{W}^{(\pi_{k'})\top}$ (and hence $\boldsymbol{w}_{j\cdot}^{(\pi_k)}, \boldsymbol{w}_{j\cdot}^{(\pi_{k'})}$ are $m \times 1$ column vectors for $j \in [p]$). Let $\boldsymbol{\mathcal{C}}^{(\pi_k,\pi_{k'})}$ be the $m \times m$ covariance matrix between $\boldsymbol{\epsilon}^{(\pi_k)}$ and $\boldsymbol{\epsilon}^{(\pi_{k'})}$ for $k \neq k'$. For $l, l' \in [m]$, its $(l,l')^{th}$ element $c_{l,l'} := \sigma^2\tilde{c}_{l,l'}$ is given as follows, where & and || refer to the logical AND and OR operators respectively:

$$\tilde{c}_{l,l'} = \begin{cases} 2 & \text{if } \pi_k(l) = \pi_{k'}(l') \,\&\, \pi_k(m+l) = \pi_{k'}(m+l') \\ -2 & \text{if } \pi_k(l) = \pi_{k'}(m+l') \,\&\, \pi_k(m+l) = \pi_{k'}(l') \\ 1 & \text{if } (\pi_k(l) = \pi_{k'}(l') \,\&\, \pi_k(m+l) \neq \pi_{k'}(m+l')) \\ & \quad || \,(\pi_k(l) \neq \pi_{k'}(l') \,\&\, \pi_k(m+l) = \pi_{k'}(m+l')) \\ -1 & \text{if } (\pi_k(l) = \pi_{k'}(m+l') \,\&\, \pi_k(m+l) \neq \pi_{k'}(l')) \\ & \quad || \,(\pi_k(l) \neq \pi_{k'}(m+l') \,\&\, \pi_k(m+l) = \pi_{k'}(l')) \\ 0 & \text{otherwise.} \end{cases}$$

Remarks on Theorem 1:

1. Referring to $\boldsymbol{h}_k$ defined in Lemma 1, this theorem implies that $\sqrt{m}(\bar{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}^*) \overset{\mathcal{L}}{\approx} \frac{1}{K}\sum_{k=1}^K \boldsymbol{h}_k$, i.e. the two quantities have approximately the same distribution.
2. Furthermore, since $\{\boldsymbol{h}_k\}_{k=1}^K$ are exchangeable, we have

$$\text{Var}(\sqrt{m}(x_{d,j} - x_j^*)) \approx \text{Var}(\frac{1}{K}\sum_{k=1}^{K}h_{k,j}) = \underbrace{\frac{1}{K^2}\sum_{k=1}^{K}\text{Var}(h_{k,j})}_{T_1} + \underbrace{\frac{2}{K^2}\sum_{k=1}^{K}\sum_{k'<k}\text{Cov}(h_{k,j}, h_{k',j})}_{T_2}.$$

   The expressions for T_1 and T_2 are respectively given by the first and second terms in the formula for τ_j^2 on the RHS of (19).
3. Theorem 1 helps derive a statistical test of significance level $\alpha \in (0,1)$ for the hypothesis $\mathsf{H}_0 : x_j^* = 0$ vs. $\mathsf{H}_1 : x_j^* \neq 0$ for each $j \in [p]$. This is expressed as follows:

$$\text{Reject } \mathsf{H}_0 \text{ in favour of } \mathsf{H}_1 \text{ if } \frac{\sqrt{m}|\bar{x}_d(j)|}{\tau_j} \geq \xi_{\alpha/2}, \quad (20)$$

   where $\xi_{\alpha/2}$ is the upper $(\frac{\alpha}{2})^{\text{th}}$ quantile of the standard normal distribution.

2.2 A BASELINE: DIFFERENCING WITH A REFERENCE ROW OF ALL ONES

We now present a baseline technique which consists of estimating the sum total signal value, i.e. $\sum_{i=1}^p x_i^*$, by deliberately setting some rows of the sensing matrix $\boldsymbol{B}$ to be all ones. Such a technique has been implemented in some works on debiasing such as Abdalla and Kur [2025], albeit in the context of Poisson measurement noise. Without loss of generality, suppose the last $r < n$ rows of the sensing matrix $\boldsymbol{B}$ consist of all ones, i.e., $\boldsymbol{b}_{(n-j+1)\cdot} \triangleq \boldsymbol{1}_p^\top$, for $j \in [r]$ with $\boldsymbol{1}_p$ denoting the $1 \times p$ vector of all ones. We refer to these rows as the 'reference rows' for the centering operation. We consider the following centering mechanism using the reference rows:

$$\boldsymbol{a}_{i\cdot} := \boldsymbol{b}_{i\cdot} - \theta\boldsymbol{1}_p^\top, \ \forall i \in [n-r] \quad (21)$$

$$y_i := z_i - \theta\bar{z}_r, \quad \forall i \in [n-r], \quad (22)$$

where $\bar{z}_r = \frac{1}{r}\sum_{j=1}^r z_{n-j+1}$. With this centering operation and using the forward model for $\boldsymbol{z}$, we obtain the following transformed linear model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{\epsilon}, \quad (23)$$

where $\boldsymbol{y} := [y_1, y_2, \ldots, y_{n-r}]^\top$, $\boldsymbol{A}$ is the corresponding $(n-r) \times p$ sensing matrix with i^{th} row $\boldsymbol{a}_{i\cdot}$ defined as in

(21), and transformed noise vector $\boldsymbol{\epsilon} := [\epsilon_1, \epsilon_2, \ldots, \epsilon_{n-r}]^\top$ for which

$$\epsilon_i = \eta_i - \frac{\theta}{r}\sum_{j=1}^{r}\eta_{n-j+1}, \text{ for } i \in [n-r]. \tag{24}$$

In the linear model (23), the noise vector $\boldsymbol{\epsilon}$ clearly has correlated entries due to the factors $\{\eta_{n-j+1}\}_{j=1}^{r}$ that are common across all $\{\epsilon_i\}_{i=1}^{n-r}$. Hence $\boldsymbol{\epsilon}$ is a $(n-r)$-dimensional zero-mean Gaussian vector with covariance matrix $\sigma^2\Sigma_{n-r}$ where

$$\Sigma_{n-r} = \boldsymbol{I}_{(n-r)} + \frac{\theta^2}{r}\mathbf{1}_{(n-r)}\mathbf{1}_{(n-r)}^\top, \tag{25}$$

where $\boldsymbol{I}_{n-r}$ is the identity matrix of size $(n-r)\times(n-r)$. The debiasing theory in Javanmard and Montanari [2014a], Zhang and Zhang [2014], van de Geer et al. [2014] is applicable to independent/uncorrelated noise, whereas the elements of $\boldsymbol{\epsilon}$ are correlated. To address the issue of correlation between the elements of $\boldsymbol{\epsilon}$, we whiten the model (23) as follows:

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{A}}\boldsymbol{x}^* + \tilde{\boldsymbol{\epsilon}}, \tag{26}$$

where $\tilde{\boldsymbol{y}} = \Sigma_{n-r}^{-1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{A}} = \Sigma_{n-r}^{-1/2}\boldsymbol{A}$ and $\tilde{\boldsymbol{\epsilon}} = \Sigma_{n-r}^{-1/2}\boldsymbol{\epsilon}$. Note that the elements of $\tilde{\boldsymbol{\epsilon}}$ are i.i.d. zero-mean Gaussian random variables with variance σ^2. The LASSO estimate of $\boldsymbol{x}^*$ from the $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{A}}$ is given by

$$\begin{aligned}\tilde{\boldsymbol{x}}_{\boldsymbol{\lambda}} &= \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{A}}\boldsymbol{x}\right\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \\ &= \arg\min_{\boldsymbol{x}} \frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^\top\Sigma_{n-r}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + \lambda\|\boldsymbol{x}\|_1.\end{aligned}$$

We debias the LASSO estimate $\tilde{\boldsymbol{x}}_{\boldsymbol{\lambda}}$ using (4) as follows.

$$\tilde{\boldsymbol{x}}_{\boldsymbol{d}} = \tilde{\boldsymbol{x}}_{\boldsymbol{\lambda}} + \frac{1}{n-r}\tilde{\boldsymbol{W}}^\top(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}}_{\boldsymbol{\lambda}}), \tag{27}$$

where

$$\begin{aligned}\tilde{\boldsymbol{W}} &:= (\tilde{\boldsymbol{w}}_{\cdot 1}|\tilde{\boldsymbol{w}}_{\cdot 2}|\ldots|\tilde{\boldsymbol{w}}_{\cdot p}) \\ \tilde{\boldsymbol{w}}_{\cdot j} &:= (n-r)(1-\mu)\tilde{\boldsymbol{a}}_{\cdot j}/\|\tilde{\boldsymbol{a}}_{\cdot j}\|_2^2,\end{aligned} \tag{28}$$

with $\tilde{\boldsymbol{a}}_{\cdot j}$ denoting the j^{th} column of $\tilde{\boldsymbol{A}}$.

One may reject the null hypothesis $\mathsf{H}_0 : x_j^* = 0$ in favor if $\mathsf{H}_1 : x_j^* \neq 0$ if $\dfrac{\sqrt{n-r}|\tilde{x}_{d,j}|}{\sigma\sqrt{\frac{1}{n-r}\tilde{\boldsymbol{w}}_{j\cdot}^\top\tilde{\boldsymbol{w}}_{j\cdot}}} \geq \xi_{\alpha/2}$.

Note that in this baseline, the application of the debiasing theory is only an approximation. This is because the whitening operation in (26) is derived from the statistics of the transformed noise in $\boldsymbol{\epsilon}$ and aims to decorrelate the elements of $\boldsymbol{\epsilon}$. However, it induces a correlation in the rows of $\tilde{\boldsymbol{A}}$ even though the rows of $\boldsymbol{A}$ are independent. On the other hand, the debiasing theory has been derived for matrices with independent (and identically distributed) rows–see [Javanmard and Montanari, 2014a, Theorem 8]. Due to which the application of debiasing theory for this baseline is only an approximation that ignores the dependence structure of the rows of $\tilde{\boldsymbol{A}}$. This issue does not affect our proposed multiple-differencing approach from Sec. 2.1. In numerical experiments, we will see that the multiple-differencing approach significantly outperforms the baseline presented in this section.

3 EXPERIMENTAL RESULTS

Data Generation Strategy: We now describe the data generation procedure used in our simulation study. We synthetically generated sparse signals $\boldsymbol{x}^* \in \mathbb{R}^p$, where the dimension was set to $p = 500$. The support size $s = \|\boldsymbol{x}^*\|_0$ was varied in the form $f_{sp}p$ where $f_{sp} \in (0,1)$ is the sparsity fraction. The non-zero entries of $\boldsymbol{x}^*$ were placed uniformly at random among the p indices. For each selected index j, a rate parameter β_j was independently drawn from a uniform distribution $\mathcal{U}(100, 10^6)$. The corresponding non-zero entries were then generated as independent Poisson random variables: $x_j^* \sim \text{Poisson}(\beta_j)$. These generated values were assigned to the selected support locations, while all remaining entries were set to zero. The measurement matrix $\boldsymbol{B} \in \mathbb{R}^{n\times p}$ was generated with i.i.d. entries drawn from a Bernoulli distribution $b_{ij} \sim \text{Bernoulli}(\theta)$ where $\theta \in (0,1)$. The noisy measurement vector $\boldsymbol{z} \in \mathbb{R}^n$ was generated in the form $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}^* + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I_n})$ where the noise standard deviation σ was chosen as $\sigma := f_\sigma \cdot \frac{1}{n}\sum_{i=1}^{n}|(\boldsymbol{B}\boldsymbol{x}^*)_i|$ with $0 < f_\sigma < 1$ controlling the noise level. The $\boldsymbol{z}$ and $\boldsymbol{B}$ of the original problem were transformed into $\boldsymbol{y}$ and $\boldsymbol{A}$ using the methods in Section 2 in order to have a zero-mean sensing matrix $\boldsymbol{A}$, before implementing LASSO debiasing.

Choice of Regularization Parameters: The regularization parameter λ was selected from a logarithmically spaced grid. $\log(\lambda) \in [1 : 0.25 : 15]$ i.e., $\lambda = \exp(t)$ for $t \in \{1, 1.25, \ldots, 15\}$. The optimal value of λ was determined using L-fold cross-validation with $L = 10$. In each fold, 90% of the measurements were used for reconstruction, while the remaining 10% were used for validation. Let $\boldsymbol{A}_r$ and $\boldsymbol{y}_r$ denote the reconstruction sub-matrix and measurement vector, respectively. Let $\boldsymbol{A}_{\text{cv}}$ and $\boldsymbol{y}_{\text{cv}}$ denote the held-out validation sub-matrix and measurement vector respectively. For each candidate λ, the estimator $\hat{\boldsymbol{x}}_\lambda$ was computed using $(\boldsymbol{A}_r, \boldsymbol{y}_r)$, and the validation error was evaluated as $\|\boldsymbol{y}_{\text{cv}} - \boldsymbol{A}_{\text{cv}}\hat{\boldsymbol{x}}_\lambda\|_2^2$. The cross-validation error was averaged across the 10 folds, and the value of λ that minimized the validation error was selected Ward [2009].

Experimental Setups: We first describe the following experimental setups to examine the variation in reconstruction performance w.r.t. one of the setup parameters, keeping others fixed:

- **Setup EA:** Varying $n \in [100 : 50 : 300]$ with $f_{sp} = 0.01$ and $f_\sigma = 0.05$.

- **Setup EB:** Varying $f_\sigma \in [0.05 : 0.05 : 0.5]$ with $n = 100$ and $f_{sp} = 0.01$.
- **Setup EC:** Varying $f_{sp} \in [0.002 : 0.002 : 0.01]$ with $n = 100$ and $f_\sigma = 0.05$.

The experiments in each setup were run 100 times across different instances of noise $\boldsymbol{\eta}$, keeping the same signal $\boldsymbol{x}^*$ (in Setups EA and EB) and the same sensing matrix $\boldsymbol{B}$ (in Setups EB and EC). In Setup EC, since the sparsity of the signal varies, the signal vector $\boldsymbol{x}^*$ also varies accordingly. Similarly, in Setup EA, as the n varies, the sensing matrix $\boldsymbol{B}$ varies in its row dimension. In all experiments we used $\theta = 0.2$ as the mean value for the elements of $\boldsymbol{B}$, though any other value in $(0, 1)$ can also be used.

Methods Compared: We henceforth refer to the proposed method of averaging the estimates from multiple differencing as MD (see Sec. 2.1), the baseline method of reference rows with all ones as RR (see Sec. 2.2) and the method of single differencing (i.e., MD with $K = 1$) as SD. As mentioned before Lemma 1, SD is a version of the debiasing method from Javanmard and Montanari [2014a] after a single centering operation. For RR, we report results with the number of reference rows r set to either 1 or 20. In general, we observed that $r > 1$ did not yield better results with RR–see Sec. S.6.1) of the appendix. For MD, we varied the number of permutations K from 1 up to 20.

Discussion of Results: As shown in Fig. 1 (top row), we see that for all three setups, the proposed MD technique with $K = 20$ outperforms RR and SD in terms of the RRMSE metric given by $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^*\|_2/\|\boldsymbol{x}^*\|_2$ where $\hat{\boldsymbol{x}}$ stands for the estimate of $\boldsymbol{x}^*$ (as obtained by the appropriate method). As K increases beyond a point, we notice that the RRMSE does not decrease much, as shown in Fig. 2. This is consistent with the intuition that, after sufficiently large K, the additional estimates become highly correlated and therefore contribute limited new information, leading to diminishing returns in accuracy. However in Fig. 2, we see a 30% RRMSE reduction for $K = 20$ compared to $K = 1$. We also experimented with other parameter regimes, where we obtained similar levels of RRMSE reduction in each case.

Hypothesis testing results: Here, we present experiments for hypothesis testing of the debiased signal estimates in order to determine from compressive measurements in $\boldsymbol{y}$ whether any jth element of the signal $\boldsymbol{x}^*$ is non-zero or not ($j \in [n]$). In a biological pooled testing application (where $\boldsymbol{B}$ is a binary pooling matrix and $\boldsymbol{z}$ is the result of the pooled tests), this will serve to determine whether or not the j^{th} sample is infected. Here we perform testing with a 95% confidence interval such that the $\xi_{\alpha/2} = 1.96$. Given the debiased signal estimate $\hat{\boldsymbol{x}}$, we create a binary vector $\hat{\boldsymbol{\phi}}$ with p elements such that for every $j \in [p]$, (*i*) we set $\hat{\phi}_j := 1$ if the null-hypothesis H_0 (see Remarks after Theorem 1) is rejected for x_j^*, and (*ii*) we set $\hat{\phi}_j := 0$ if the null-hypothesis is accepted for x_j^*. We consider the j^{th} element a false positive (FP) if $x_j^* = 0, \hat{\phi}_j \neq 0$; a false negative (FN) if $x_j^* \neq 0, \hat{\phi}_j = 0$; a true positive (TP) if $x_j^* \neq 0, \hat{\phi}_j \neq 0$; and a true negative (TN) if $x_j^* = 0, \hat{\phi}_j = 0$. We consider the following performance measures: sensitivity given by TP/(TP+FN), specificity given by TN/(TN+FP), accuracy given by (TP + TN)/(TP + TN + FP + FN) and precision given by TP/(TP+FP). As seen in Table 1, the MD method with $K = 20$ reduces the number of false positives as compared to RR with $r = 1$, RR with $r = 20$, and SD. This gets reflected in significant improvement in precision and F1-score–see Fig. S.5. There is also an improvement in accuracy and specificity. Accuracy and specificity values in the experimental setups (EA), (EB), (EC) defined earlier, are presented in Fig. 1, where again the superior performance of MD is clear.

Table 1: Performance measures for different methods; parameters $n = 100, p = 500, f_{sp} = 0.01, f_\sigma = 0.05, \theta = 0.2$

Method	Acc.	Prec.	Sens.	Spec.	F1-score
RR, $r = 20$	0.816	0.052	1.000	0.814	0.099
RR, $r = 1$	0.959	0.202	1.000	0.958	0.335
SD, $K = 1$	0.967	0.239	1.000	0.967	0.384
MD, $K = 20$	0.983	0.389	1.000	0.983	0.554

Bias and variance analysis: In this part, we examined the effect of multiple differencing on the bias of the debiased LASSO. For this, we computed the debiased LASSO estimates $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}^{(\pi_k)}\}_{k=1}^K$ and their corresponding bias vectors $\{\boldsymbol{\Delta}_m^{(\pi_k)}\}_{k=1}^K$ as per the formula from (17). The total bias in $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ is given by the vector $\bar{\boldsymbol{\Delta}}^{(K)} := \frac{1}{K}\sum_{k=1}^K \boldsymbol{\Delta}_m^{(\pi_k)}$. In Fig. 3 (left), a box-plot of the noise-normalized bias given by $\{\bar{\boldsymbol{\Delta}}_j^{(K)}/\sigma\}_{j=1}^p$ (that is, bias divided by the standard deviation of the noise) is plotted. This is repeated for every value of K from 1 to 20. The plot demonstrates that the range of the noise-normalized bias values becomes smaller with increase in the number of permutations K. In Fig. 3 (center), we also plotted the relative bias, i.e. coordinate-wise bias values normalized by the standard deviation of the corresponding coordinate of $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$, i.e. we prepared box-plots of $\{\bar{\boldsymbol{\Delta}}_j^{(K)}/\tau_j\}_{j=1}^p$ where τ_j is given in (19). Here again, we see a decreasing trend as K increases.

The right side of Fig. 3 shows box-plots of the noise-normalized standard deviation of the elements of $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$, i.e. it shows a box-plot of the values of $\{\tau_j/\sigma\}_{j=1}^p$ where τ_j is as computed from (19). These box-plots are across different values of the number of permutations K, i.e. $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ in (15) is computed for different values of K. The plots show that with increased K, there is a significant decrease in the variance of the debiased estimates.

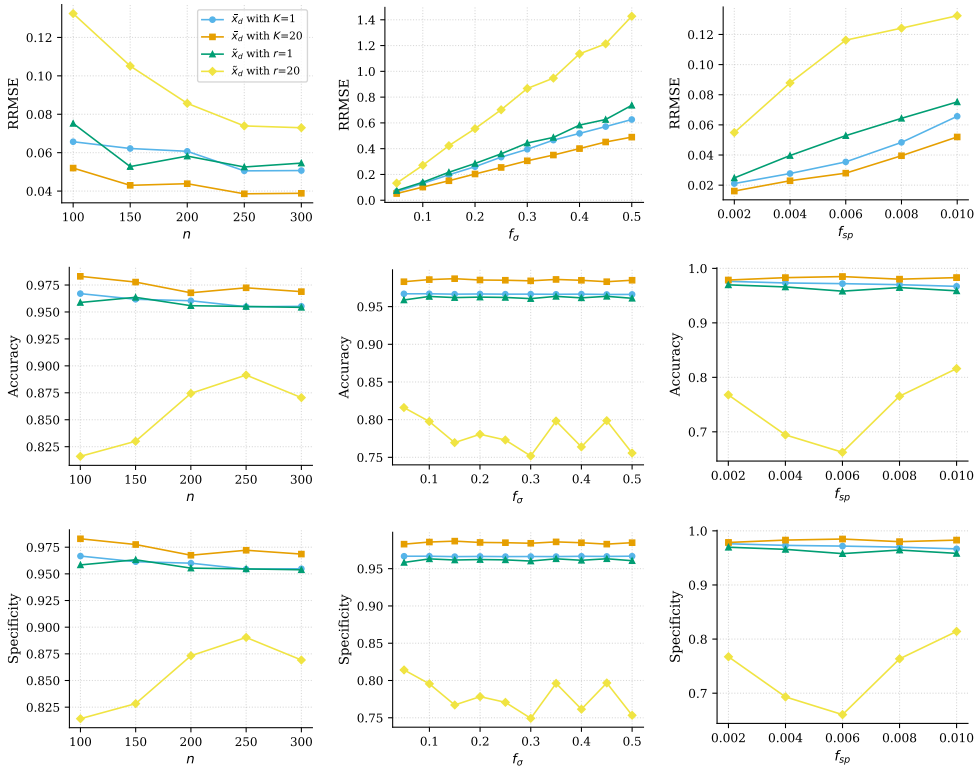

Figure 1: RRMSE (top row), accuracy (middle row), specificity (last row) as a function of n (Setup EA), f_σ (Setup EB) and f_{sp} (Setup EC), for the methods of multiple differencing (MD) with $K = 20$, single differencing (SD), and reference-rows (RR) with $r = 1, r = 20$. In any row, the setups EA, EB, EC are respectively in the left, center and right columns.

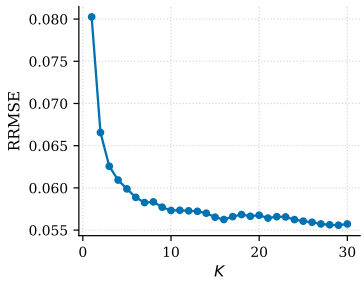

Figure 2: RRMSE across number of permutations K (see (15)) for multiple differencing (MD) with parameters $n = 100$, $p = 500$, $f_{sp} = 0.01$, $f_\sigma = 0.05, \theta = 0.2$.

4 CONCLUSION

In this paper, we have presented a technique of LASSO debiasing when the design matrix $\boldsymbol{B}$ (of size $n \times p$) has elements with a constant non-zero mean. Such a design matrix emerges in many applications such as compressive optical imaging and pooled biological testing. Our technique is based on the simple idea of creating a new matrix $\boldsymbol{A}$ (of size $m \times p, m = n/2$) whose rows are obtained by computing the difference between rows of the original matrix. Although this appears to reduce the number of measurements, we show that use of different permutations for performing the difference operations overcomes this limitation and allows all original measurements to be used effectively. This gives rise to different debiased signal estimates $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$, one for each permutation π_k. These individual estimates turn out to be exchangeable random variables. Based on this, we derive a debiased LASSO estimator by averaging the individual estimates. We empirically show that with multiple differencing (i.e. $K > 1$), the element-wise variances of the debiased LASSO estimates are significantly reduced. This paves the way for narrow confidence intervals. Future work will involve investigating practical applications of this estimator in actual pooled testing and compressive imaging. Another interesting avenue for future work is to examine other ways of aggregating the different estimates $\{\hat{\boldsymbol{x}}_{\boldsymbol{d}}{}^{(\pi_k)}\}_{k=1}^K$ apart from averaging.

Acknowledgements: AR would like to acknowledge support from ANRF MATRICS grant number ANRF/ARGM/2025/002375/MTR.

References

Pedro Abdalla and Gil Kur. Debiased LASSO under Poisson–Gauss model. *Information and Inference: A*

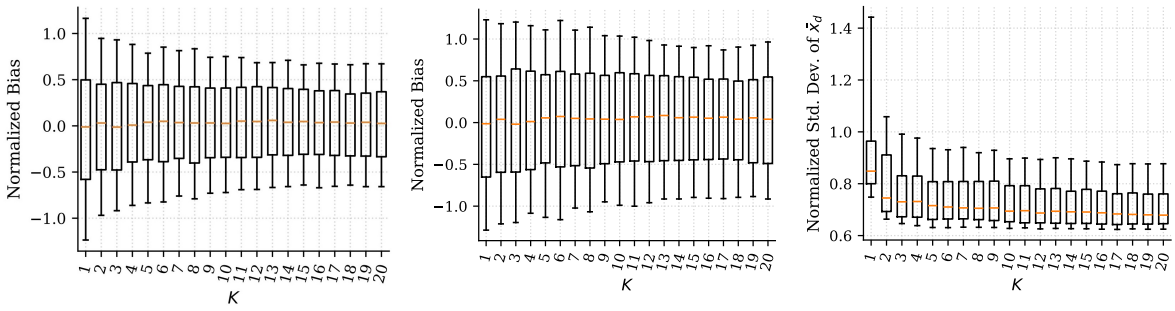

Figure 3: Box-plots of Noise-normalized coordinate-wise bias $\{\bar{\boldsymbol{\Delta}}_j^{(K)}/\sigma\}_{j=1}^p$ (left), coordinate-wise relative bias $\{\bar{\boldsymbol{\Delta}}_j^{(K)}/\tau_j\}_{j=1}^p$ (center) and noise-normalized coordinate-wise standard deviation (i.e. τ_j in (19) divided by noise σ) (right). All three plots are for the debiased LASSO estimate with multiple differencing, given by $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$. The box-plots show the median (orange line), the $25-$ and $75-$percentile values (given by the top and bottom lines of each box), and the $10-$ and $90-$percentile values (given by the whiskers).

Journal of the IMA, 14(2), 2025.

Shuvayan Banerjee, James Saunderson, Radhendushka Srivastava, and Ajit Rajwade. Fast debiasing of the LASSO estimator. *Transactions on Machine Learning Research*, 2026.

Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of LASSO and Dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

Michael Celentano and Andrea Montanari. Correlation adjusted debiased LASSO: debiasing the LASSO with inaccurate covariate model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5): 1455–1482, 2024.

Michael Celentano, Andrea Montanari, and Yuting Wei. The LASSO with general gaussian designs with applications to hypothesis testing. *Annals of Statistics*, 51(5):2194–2220, 2023.

Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

Sabyasachi Ghosh, Rishi Agarwal, Mohammad Ali Rehan, Shreya Pathak, Pratyush Agarwal, Yash Gupta, Sarthak Consul, Nimay Gupta, Ritesh Goenka, Ajit Rajwade, and Manoj Gopalkrishnan. A compressed sensing approach to pooled RT-PCR testing for COVID-19 detection. *IEEE Open Journal of Signal Processing*, 2:248–264, 2021.

Graham M. Gibson, Steven D. Johnson, and Miles J. Padgett. Single-pixel imaging 12 years on: a review. *Optics Express*, 28(19):28190–28208, 2020.

Trevor Hastie, Ryan Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The LASSO and Generalizations*. CRC Press, 2015.

Xin Jiang Hunt, Patricia Reynaud-Bouret, Vincent Rivoirard, Laure Sansonnet, and Rebecca Willett. A data-dependent weighted LASSO under Poisson noise. *IEEE Transactions on Information Theory*, 65(3):1589–1613, 2018.

Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 2014a.

Adel Javanmard and Andrea Montanari. Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 60(10):6522–6554, 2014b.

Yuan Li and Garvesh Raskutti. Minimax optimal convex methods for Poisson inverse problems under lq-ball sparsity. *IEEE Transactions on Information Theory*, 64(8): 5498–5512, 2018.

Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. Restricted eigenvalue properties for correlated Gaussian designs. *Journal of Machine Learning Research*, 11:2241–2259, 2010.

Sara van de Geer. On the asymptotic variance of the debiased LASSO. *Electronic Journal of Statistics*, 13(2): 2970–3008, 2019.

Sara van de Geer, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics*, 42(3):1166 – 1202, 2014.

Rachel Ward. Compressed sensing with cross validation. *IEEE Transactions on Information Theory*, 55(12):5773–5782, 2009.

Cun-Hui Zhang and Stephanie S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):217–242, 2014.

Shuheng Zhou. Restricted eigenvalue conditions on subgaussian random matrices. Technical Report UCB/EECS-2009-90, University of California, Berkeley, 2009. https://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-90.pdf.

Appendix: A Debiased LASSO Estimator for Design Matrices with Non-zero Mean Elements

Ashish Zantye[1] **Ajit Rajwade**[1] **Radhendushka Srivastava**[2]

[1]Department of Computer Science and Engineering, IIT Bombay, India
[2]Department of Mathematics, IIT Bombay, India

In this appendix, we first provide the proofs of the theoretical results of the main paper. This is done in Sec. S.5. We also provide additional experimental results in Sec. S.6. All section, figure or equation numbers in this appendix start with an 'S'. The numbers that refer to entities defined in the main paper use the usual number convention without a starting 'S'.

S.5 PROOFS OF THEORETICAL RESULTS

We begin with a general fact for a sequence of exchangeable random variables.

S.5.1 BASIC LEMMA ON EXCHANGEABILITY

Lemma SL.2 *Let $U_1, U_2, \ldots, U_K$ be a sequence of exchangeable random variables with $E(U_1^2) < \infty$. Let $\bar{U} := \frac{1}{K}\sum_{k=1}^{K} U_k$. Then*

$$\textit{Var}(\bar{U}) \leq \textit{Var}(U_k) \textit{ for all } k = 1, 2, \ldots, K. \blacksquare \tag{S.1}$$

Proof of Lemma SL.2: By using the exchangeability of U_k's, we have

$$\text{Var}(\bar{U}) = \frac{1}{K^2}\sum_{k=1}^{K}\text{Var}(U_k) + \frac{2}{K^2}\sum_{k=1}^{K}\sum_{k'<k}\text{Cov}(U_k, U_{k'}) = \frac{\sigma_u^2}{K} + \frac{K-1}{K}\zeta_u, \tag{S.2}$$

where $\sigma_u^2 = \text{Var}(U_k)$ and $\zeta_u = \text{Cov}(U_k, U_{k'})$ for any $k, k', k \neq k'$ due to exchangeability. By using the Cauchy-Schwarz inequality, we have $|\zeta_u| \leq \sigma_u^2$. Therefore, by using (S.2), we have

$$\text{Var}(\bar{U}) \leq \frac{\sigma_u^2}{K} + \frac{K-1}{K}|\zeta_u| \leq \sigma_u^2 = \text{Var}(U_k).$$

This completes the proof. ■

Remark: Note that the equality holds in (S.1) if $\text{Cov}(U_k, U_{k'}) = \text{Var}(U_k)$. In this case, $P(U_k = U_{k'}) = 1$. Therefore, if exchangeable random variables are not identically the same, the variance of the average is typically *smaller* than the individual variance.

S.5.2 PROOF OF LEMMA 1

Part (a): By using (10) and (12), we have

$$\sqrt{m}(\hat{\boldsymbol{x}}_{\boldsymbol{d}}^{(\boldsymbol{\pi_k})} - \boldsymbol{x}^*) = \frac{1}{\sqrt{m}}\boldsymbol{W}^{(\pi_k)\top}\boldsymbol{\epsilon}^{(\pi_k)} + \sqrt{m}\left(\frac{1}{m}\boldsymbol{W}^{(\pi_k)\top}\boldsymbol{A}^{(\pi_k)} - \boldsymbol{I}_p\right)\left(\boldsymbol{x}^* - \hat{\boldsymbol{x}}_{\lambda}^{(\pi_k)}\right). \tag{S.3}$$

We get (16) by averaging (S.3) over permutations $\{\pi_k\}_{k=1}^K$.
(b) By using (11) and (16), we have

$$\boldsymbol{h}_k = \frac{1}{\sqrt{m}} \boldsymbol{W}^{(\pi_k)\top} \boldsymbol{\epsilon}^{(\pi_k)} = \frac{1}{\sqrt{m}} \boldsymbol{W}^{(\pi_k)\top} \left(\boldsymbol{\eta}_1^{(\pi_k)} - \boldsymbol{\eta}_2^{(\pi_k)} \right), \tag{S.4}$$

where $\boldsymbol{\eta}_1^{(\pi_k)} := [\eta_{\pi_k(1)}, \eta_{\pi_k(2)}, \ldots, \eta_{\pi_k(m)}]^\top$ and $\boldsymbol{\eta}_2^{(\pi_k)} := [\eta_{\pi_k(m+1)}, \eta_{\pi_k(m+2)}, \ldots, \eta_{\pi_k(m+m)}]^\top$. Note that random vectors $\boldsymbol{\eta}_1^{(\pi_k)}$ and $\boldsymbol{\eta}_2^{(\pi_k)}$ are independent Gaussian vectors. Since $\{\boldsymbol{\epsilon}^{(\pi_k)}\}_{k=1}^K$ are obtained by differencing the elements of $\boldsymbol{\eta}$, the conditional distribution of mK-dimensional random vector $\boldsymbol{\epsilon}_K^\top := [\boldsymbol{\epsilon}^{(\pi_1)\top}, \boldsymbol{\epsilon}^{(\pi_2)\top}, \ldots, \boldsymbol{\epsilon}^{(\pi_K)\top}]$ given $\{\pi_k\}_{k=1}^K$ is a zero-mean mK-dimensional multivariate Gaussian distribution. Further, given sensing matrix $\boldsymbol{B}$ and permutations $\{\pi_k\}_{k=1}^K$, the elements of random vectors $\{\boldsymbol{h}_k\}_{k=1}^K$ are weighted linear combinations of jointly Gaussian vectors $\{\boldsymbol{\epsilon}^{(\pi_k)}\}_{k=1}^K$. Therefore, the conditional joint distribution of random vectors $\{\boldsymbol{h}_k\}_{k=1}^K$ is an mK-dimensional multivariate Gaussian distribution with mean $\mathbf{0}_{mK}$ and covariances as

$$\begin{aligned} \text{Var}(\boldsymbol{h}_k) &= \frac{1}{m} \boldsymbol{W}^{(\pi_k)\top} \text{Var}\left(\boldsymbol{\epsilon}^{(\pi_k)} \right) \boldsymbol{W}^{(\pi_k)} & \text{(S.5)} \\ \text{Cov}(\boldsymbol{h}_k, \boldsymbol{h}_{k'}) &= \frac{1}{m} \boldsymbol{W}^{(\pi_k)\top} \text{Cov}\left(\boldsymbol{\epsilon}^{(\pi_k)}, \boldsymbol{\epsilon}^{(\pi_{k'})} \right) \boldsymbol{W}^{(\pi_k)}. & \text{(S.6)} \end{aligned}$$

Let ψ be a permutation of the set $[K] := \{1, 2, ..., K\}$. Since the elements of the sensing matrix $\boldsymbol{B}$ are i.i.d. and the permutations $\{\pi_k\}_{k=1}^K$ are independent, the unconditional distribution of the elements of $\text{Var}(h_{\psi(k)})$ will be the same for all $k \in [K]$. Similarly, the unconditional distribution of elements of $\text{Cov}\left(h_{\psi(k)}, h_{\psi(k')}\right)$ will also be same where $k, k' \in [K]; k \neq k'$. This completes the proof.
(c) From [Javanmard and Montanari, 2014a, Theorem 8], we have the following given the conditions mentioned in Lemma 1:

$$\|\boldsymbol{\Delta}_m^{(\pi_k)}\|_\infty = O_P\left(\frac{s \log p}{\sqrt{m}} \right) \; \forall k = 1, 2, \ldots, K. \tag{S.7}$$

Note that

$$\left\| \frac{1}{K} \sum_{i=1}^K \boldsymbol{\Delta}_m^{(\pi_k)} \right\|_\infty \leq \max_{k \in \{1,2,\ldots,K\}} \|\boldsymbol{\Delta}_m^{(\pi_k)}\|_\infty = \|\boldsymbol{\Delta}_m^{(\pi_{k_0})}\|_\infty \quad \text{for some } k_0 \in \{1, 2, \ldots, K\}. \tag{S.8}$$

Therefore, by using (S.7) and (S.8), we have the result. ■

S.5.3 PROOF OF THEOREM 1

Suppose $m \geq \mathcal{O}(s^2 \log^2 p)$. Then, by using Lemma 1 and (16), we have

$$\sqrt{m}(\bar{x}_{d,j} - x_j^*) = \frac{1}{K} \sum_{i=1}^K \frac{1}{\sqrt{m}} \boldsymbol{w}_{j.}^{(\pi_k)\top} \boldsymbol{\epsilon}^{(\pi_k)} + o_P(1), \forall\, j = 1, 2, \ldots, p \tag{S.9}$$

where $\boldsymbol{w}_{j.}^{(\pi_k)\top}$ denote the j^{th} row of the $\boldsymbol{W}^{(\pi_k)\top}$. Note that $\boldsymbol{w}_{j.}^{(\pi_k)}$ is $m \times 1$ vector. Note that, given $\{\pi_k\}_{k=1}^K$, the transformed noise vectors $\{\boldsymbol{\epsilon}^{(\pi_k)}\}_{k=1}^K$ are jointly Gaussian. Thus, the conditional asymptotic distribution of $\sqrt{m}(\bar{x}_{d,j} - x_j^*)$ is a zero mean Gaussian distribution with variance

$$\begin{aligned} \tau_j^2 &= \text{Var}\left(\frac{1}{K} \sum_{i=1}^K \frac{1}{\sqrt{m}} \boldsymbol{w}_{j.}^{(\pi_k)\top} \boldsymbol{\epsilon}^{(\pi_k)} \right) \\ &= \frac{1}{K^2} \sum_{k=1}^K \frac{1}{m} \boldsymbol{w}_{j.}^{(\pi_k)\top} \text{Var}\left(\boldsymbol{\epsilon}^{(\pi_k)} \right) \boldsymbol{w}_{j.}^{(\pi_k)} + \frac{2}{K^2} \sum_{k=1}^K \sum_{k'<k} \frac{1}{m} \boldsymbol{w}_{j.}^{(\pi_k)\top} \text{Cov}\left(\boldsymbol{\epsilon}^{(\pi_k)}, \boldsymbol{\epsilon}^{(\pi_{k'})} \right) \boldsymbol{w}_{j.}^{(\pi_{k'})}. \end{aligned} \tag{S.10}$$

By using (11), we have

$$\boldsymbol{\epsilon}^{(\pi_k)\top} = [\eta_{\pi_k(1)} - \eta_{\pi_k(m+1)}, \eta_{\pi_k(2)} - \eta_{\pi_k(m+2)}, \ldots, \eta_{\pi_k(m)} - \eta_{\pi_k(m+m)}]^\top. \tag{S.11}$$

Since π_k is a permutation of $[n]$ and $\{\eta_i\}_{i=1}^n$ are i.i.d. $\mathcal{N}(0, \sigma^2)$, we have

$$\text{Var}\left(\boldsymbol{\epsilon}^{(\pi_k)} \right) = 2\sigma^2 \boldsymbol{I}_m. \tag{S.12}$$

By using (S.11), the $(l,l')^{\text{th}}$ element of the covariance matrix, $\text{Cov}\left(\boldsymbol{\epsilon}^{(\pi_k)}, \boldsymbol{\epsilon}^{(\pi_{k'})}\right) := \boldsymbol{\mathcal{C}}^{(\pi_k,\pi_{k'})}$, is given as follows:

$$\begin{aligned} c_{l,l'} &= \text{Cov}\left(\eta_{\pi_k(l)} - \eta_{\pi_k(m+l)}, \eta_{\pi_{k'}(l')} - \eta_{\pi_{k'}(m+l')}\right) \\ &= E(\eta_{\pi_k(l)}\eta_{\pi_{k'}(l')}) - E(\eta_{\pi_k(m+l)}\eta_{\pi_{k'}(l')}) - E(\eta_{\pi_k(l)}\eta_{\pi_{k'}(m+l')}) + E(\eta_{\pi_k(m+l)}\eta_{\pi_{k'}(m+l')}). \end{aligned} \tag{S.13}$$

Since $\{\eta_i\}_{i=1}^n$ are i.i.d. $\mathcal{N}(0,\sigma^2)$, each term on RHS of (S.13) is of the form $E(\eta_a\eta_b)$ where a and b are indices belonging to $[n]$. If $a=b$, then such a term equals σ^2, otherwise it equals 0. Further, note that π_k and $\pi_{k'}$ are two different permutations of $[n] := \{1,2,\ldots,n\}$, therefore $\pi_k(l) \neq \pi_k(m+l)$ and $\pi_{k'}(l') \neq \pi_{k'}(m+l')$. We enumerate all possibilities of the same indices of η for the terms on the RHS of (S.13) and have the following (where & and || denote the logical AND and OR operators respectively):

$$c_{l,l'} = \begin{cases} 2\sigma^2 & \text{if } \pi_k(l) = \pi_{k'}(l') \;\&\; \pi_k(m+l) = \pi_{k'}(m+l') \\ -2\sigma^2 & \text{if } \pi_k(l) = \pi_{k'}(m+l') \;\&\; \pi_k(m+l) = \pi_{k'}(l') \\ \sigma^2 & \text{if } \left(\pi_k(l) = \pi_{k'}(l') \;\&\; \pi_k(m+l) \neq \pi_{k'}(m+l')\right) \text{ || } \left(\pi_k(l) \neq \pi_{k'}(l') \;\&\; \pi_k(m+l) = \pi_{k'}(m+l')\right) \\ -\sigma^2 & \text{if } \left(\pi_k(l) = \pi_{k'}(m+l') \;\&\; \pi_k(m+l) \neq \pi_{k'}(l')\right) \text{ || } \left(\pi_k(l) \neq \pi_{k'}(m+l') \;\&\; \pi_k(m+l) = \pi_{k'}(l')\right) \\ 0 & \text{otherwise.} \end{cases} \tag{S.14}$$

To understand this formula, consider the first case where $\pi_k(l) = \pi_{k'}(l') \;\&\; \pi_k(m+l) = \pi_{k'}(m+l')$. In this case, we have:

$$\begin{aligned} c_{l,l'} &= E(\eta_{\pi_k(l)}\eta_{\pi_{k'}(l')}) - E(\eta_{\pi_k(m+l)}\eta_{\pi_{k'}(l')}) - E(\eta_{\pi_k(l)}\eta_{\pi_{k'}(m+l')}) + E(\eta_{\pi_k(m+l)}\eta_{\pi_{k'}(m+l')}) \\ &= E(\eta^2_{\pi_k(l)}) - E(\eta_{\pi_k(m+l)}\eta_{\pi_k(l)}) - E(\eta_{\pi_k(l)}\eta_{\pi_k(m+l)}) + E(\eta^2_{\pi_k(m+l)}) \\ &= \sigma^2 - 0 - 0 + \sigma^2 = 2\sigma^2. \end{aligned} \tag{S.15}$$

Likewise, the other cases can also be easily evaluated.

Now, by plugging the expressions from (S.12) and (S.14) in (S.10), we get the expression for τ_j^2 as in (19). Thus, if $m \geq \mathcal{O}(s^2\log^2 p)$ due to which the bias terms become very small, we have

$$\left.\frac{\sqrt{m}(\bar{x}_{d,j} - x_j^*)}{\tau_j}\right| \{\pi_k\}_{k=1}^K \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ as } m \to \infty, \tag{S.16}$$

where $\xrightarrow{\mathcal{L}}$ denotes the convergence in Law (or distribution). Since the limiting distribution does not depend upon the conditioning event, the limit of the unconditional distribution will coincide with the limit of the conditional distribution. This completes the proof. ■

S.6 ADDITIONAL EXPERIMENTAL RESULTS

S.6.1 IMPACT OF INCREASING ALL-ONES REFERENCE ROWS

Increasing the number of al1-ones reference rows reduces the noise correlation ((25)); however, this comes at the cost of degraded estimation accuracy. Specifically, as the number of reference rows r increases, the number of effective measurement rows decreases ($\tilde{m} = n - r$), resulting in higher RRMSE as seen in Fig. S.4. This trade-off suggests that the loss of informative measurements outweighs the benefit gained from reduced noise correlation. Furthermore, whitening the design matrix $\boldsymbol{A}$ and the response vector $\boldsymbol{y}$ to account for the known noise covariance via the Mahalanobis transformation (see Sec. 2.2) does not lead to a significant improvement in RRMSE. We employ the whitening procedure in the experiments for consistency, although its impact on RRMSE is marginal.

S.6.2 ADDITIONAL PERFORMANCE MEASURES: PRECISION & F1 SCORE

For the experimental setups (EA), (EB), (EC) defined in Sec. 3, we present additional results for precision and F1 Score of the results using the different methods. These results are presented in Fig. S.5.

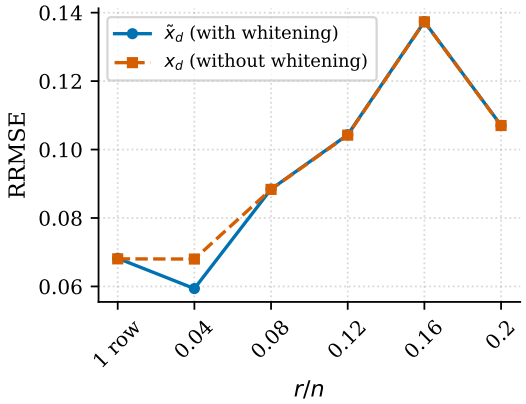

Figure S.4: RRMSE as a function of the fraction of total measurements which are all-ones reference rows (r/n) with parameters being $n = 100, p = 500, f_{sp} = 0.01, f_\sigma = 0.05, \theta = 0.2$ using a whitening transformation as in (i.e., as per (26) in Sec. 2.2) as well as without the whitening transformation (i.e., as per (23) in Sec. 2.2).

S.6.3 CORRELATION COEFFICIENTS OF DEBIASED ESTIMATES

In Fig. S.6, we plot a histogram the coordinate-wise correlation coefficient values $\{\rho_j\}_{j=1}^p$ for the entries of $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$. This is given by:

$$\rho_j := \frac{\frac{1}{K(K-1)} \sum_{k=1}^K \sum_{k'=1, k' \neq k}^K (\hat{x}_{d,j}^{(\pi_k)} - x_j^*)(\hat{x}_{d,j}^{(\pi_{k'})}) - x_j^*)}{s_{d,j}^2},$$

where

$$s_{d,j}^2 := \frac{\sum_{k=1}^K \left(\hat{x}_{d,j}^{(\pi_k)} - x_j^*\right)^2}{K}.$$

From Fig. S.6, we see that the correlation values are concentrated around 0 and indicates small probability for larger values of the correlation value.

S.7 EXPERIMENTS ON IMAGE RECONSTRUCTION

We ran an image reconstruction experiment given compressive measurements that follow the architecture of the Rice Single Pixel Camera Duarte et al. [2008], a well known compressive camera. Consider an image $\boldsymbol{x} \in \mathbb{R}^p$ that is sparse in some orthonormal basis $\boldsymbol{\Psi} \in \mathbb{R}^{p \times p}$ in the form $\boldsymbol{x} = \boldsymbol{\Psi\theta}$, yielding a sparse coefficient vector $\boldsymbol{\theta} \in \mathbb{R}^p$. Consider compressive measurements of the form $\boldsymbol{y} = \boldsymbol{Bx} + \boldsymbol{\eta} = \boldsymbol{B\Psi\theta} + \boldsymbol{\eta}$ where the elements of the noise vector $\boldsymbol{\eta} \in \mathbb{R}^n$ are drawn from $\mathcal{N}(0, \sigma^2)$ and the elements of $\boldsymbol{B}$ are drawn iid from Bernoulli(0.5). In our experiments, we set $\sigma := 0.001 \times \frac{1}{n} \sum_{i=1}^n \boldsymbol{b}^i \boldsymbol{x}$ where $\boldsymbol{b}^i$ is the i-th row of $\boldsymbol{B}$, i.e. σ was set to one percent of the average absolute value of the noiseless measurements. Note that $\boldsymbol{y} \in \mathbb{R}^n$ and $n < p$. The aim is to reconstruct $\boldsymbol{x}$ from $\boldsymbol{y}, \boldsymbol{B}$ using the multiple differencing (MD) method. Note that the MD technique converts the binary $\boldsymbol{B}$ matrix to a Rademacher matrix $\boldsymbol{A}$ with $m = n/2$ measurements. Image reconstruction results show an improvement when using $K = 20$ over $K = 1$ – see Fig S.7.

S.8 COMPUTATIONAL COST

Though our method requires K separate executions of the LASSO, these can be executed in parallel. We also report the time required in the simulation study presented below. The time for single iteration using LASSO in sklearn linear model package was 0.68 seconds for the selected optimal λ for the problem $p = 500$, $n = 100$, $\theta = 0.2$, $f_{sp} = 0.01$, $f_\sigma = 0.05$. Experiments were conducted on an HP Spectre x360 laptop with an Intel Core i7-1355U processor (10 cores) and 32 GB of RAM. All implementations were executed on a 64-bit architecture. Reported run-times were measured using high-resolution timers and averaged over multiple runs to ensure stability. Since the time per execution is small, multiple executions via $K > 1$ do not significantly build up on the time (besides being parallelizable).

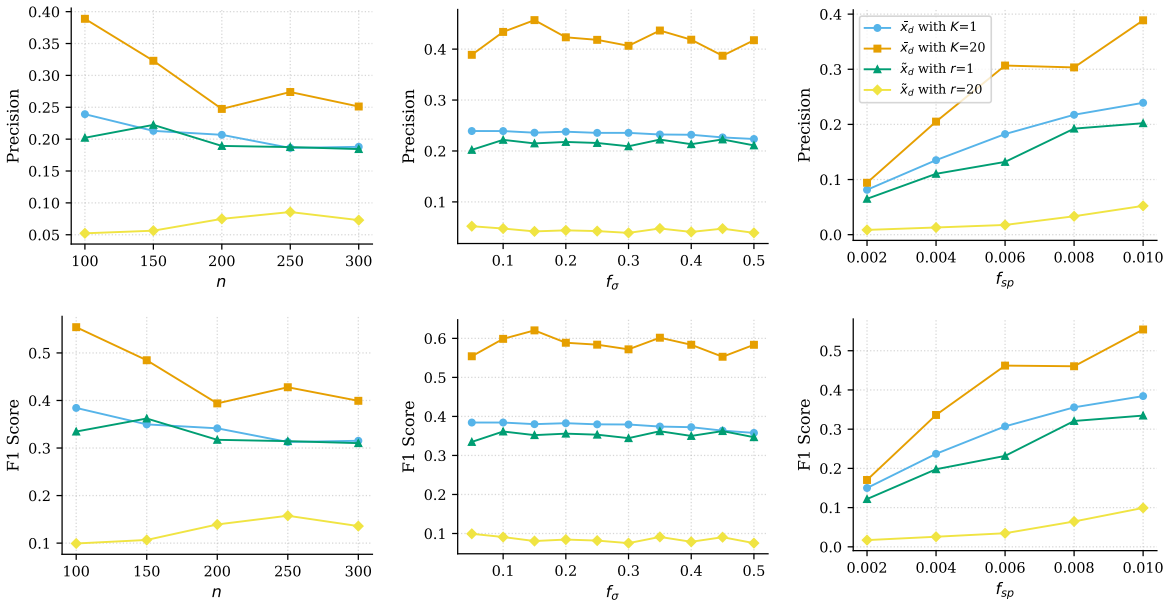

Figure S.5: Precision (first row) and F1 Score (second row) as a function of n (Setup EA), f_σ (Setup EB) and f_{sp} (Setup EC), for the methods of multiple differencing (MD) with $K = 20$, single differencing (SD) and reference-rows (RR) with $r = 1$ and $r = 20$. Other parameters are $n = 100$, $p = 500$, $f_{sp} = 0.01$, $f_\sigma = 0.05$, $\theta = 0.2$.

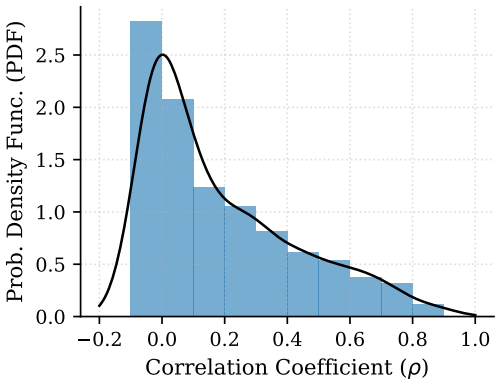

Figure S.6: Probability histogram of the correlation coefficient values $\{\rho\}_{j=1}^{p}$ of the entries of the debiased estimate $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ (using MD with $K = 20$), with parameters $n = 100, p = 500, f_{sp} = 0.01, f_\sigma = 0.05, \theta = 0.2$. The kernel density estimate of correlation values (Gaussian kernel, bandwidth=0.1) is shown in the black curve.

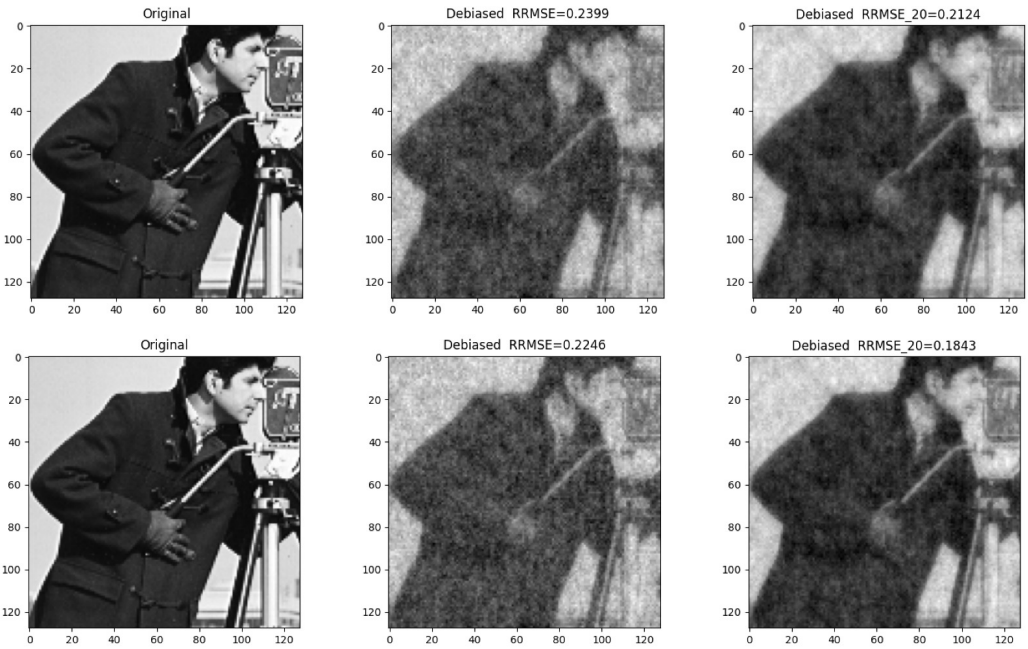

Figure S.7: Image reconstruction with multiple differencing technique from compressive measurements for an image of size 128×128, i.e. $p = 128^2$. In every row: ground truth (left), result with $K = 1$ (center), result with $K = 20$ (right). Top row: $n = 5000$ measurements, i.e. $m = 2500$; bottom row: $n = 10000$ measurements, i.e. $m = 5000$. Notice better quality reconstruction as well as lower RRMSE with $K = 20$ than with $K = 1$. RRMSE is computed as $\|\bar{\boldsymbol{x}}_{\boldsymbol{d}} - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2$ where $\boldsymbol{x}$ is the true image and $\bar{\boldsymbol{x}}_{\boldsymbol{d}}$ is its debiased estimate. In all cases, the measurements were corrupted with noise from $\mathcal{N}(0, \sigma^2)$ where σ was set to 0.1 percent of the average absolute value of the noiseless measurements.