SAMPLING STRATEGIES FOR COMPRESSIVE IMAGING UNDER STATISTICAL NOISE

Frederik Hoppe[1] *Felix Krahmer*[2,3,7] *Claudio Mayrink Verdun*[2,3] *Marion I. Menzel*[4,5,6] *Holger Rauhut*[1]

[1] Chair of Mathematics of Information Processing, RWTH Aachen University, Aachen, Germany
[2] Department of Mathematics, Technical University of Munich, Munich, Germany
[3] Munich Center for Machine Learning, Munich, Germany
[4] AImotion Bavaria, Technische Hochschule Ingolstadt, Ingolstadt, Germany
[5] Department of Physics, Technical University of Munich, Garching, Germany
[6] GE Healthcare, Munich, Germany
[7]Technical University of Munich, Munich Data Science Institute

ABSTRACT

Most of the compressive sensing literature in signal processing assumes that the noise present in the measurement has an adversarial nature, i.e., it is bounded in a certain norm. At the same time, the randomization introduced in the sampling scheme usually assumes an i.i.d. model where rows are sampled with replacement. In this case, if a sample is measured a second time, it does not add additional information. For many applications, where the statistical noise model is a more accurate one, this is not true anymore since a second noisy sample comes with an independent realization of the noise, so there is a fundamental difference between sampling with and without replacement. Therefore, a more careful analysis must be performed. In this short note, we illustrate how one can mathematically transition between these two noise models. This transition gives rise to a weighted LASSO reconstruction method for sampling without replacement, which numerically improves the solution of high-dimensional compressive imaging problems.

***Index Terms*—** LASSO, sparse regression, compressed sensing, statistical noise, non-uniform sampling

I. INTRODUCTION

High-dimensional sparse recovery problems have been subject to intensive study over the last decades [1]. The reason is that sparsity naturally arises in real signals. Consequently, there have been various analysis perspectives. Many works in the signal processing literature and adjacent areas of mathematics have addressed the challenges arising from inherent structure in the measurements by often considering a very general noise model. Indeed, usually the noise is considered to be bounded – the so-called adversarial noise model – and no statistical model is assumed [1]. At the same time, many works in the statistics literature have focused on the effects of statistical noise on the reconstruction accuracy, but mainly focused on generic measurement designs, e.g., with (sub-)Gaussian entries [2]. Nevertheless, the resulting theories are closely related. In particular, there is a relationship between the restricted isometry property (RIP), used in the former, and the restricted eigenvalue (RE) property, used for the analysis of the latter [3]. A subtle difference between random and deterministic noise, however, arises for measurement matrices sampled from a fixed measurement system. Such a scenario is naturally related to applications in magnetic resonance imaging (MRI), the first motivation for compressive sensing, where a suitable measurement system is the Fourier basis [4]. Consequently, such measurement systems have been studied in a number of works concerning theory and application [5], [6], [7], [8]. A challenge in the analysis of the solution of the high-dimensional sparse problem is that the underlying image is typically not sparse in the standard basis, but rather in a different dictionary, e.g., wavelets. As shown in [7], [9], this can be addressed by sampling with replacement from a distribution with larger density for small frequencies. Namely, in combination with an appropriate preconditioning matrix, this sampling strategy allows to overcome potential coherence issues – see Theorem 2 – which then implies recovery guarantees via the standard sparse recovery theory.

The large density for small frequencies has the effect that these frequencies are typically observed multiple times among the measurements. However, it has been argued that one should nevertheless sample without replacement and hence just ignore these multiple occurrences [1, Chapter 12]. The reason is that deterministic noise could just be the same for multiple observations of the same measurement, implying that repeated observations are not guaranteed to add information. This situation fundamentally changes for random noise models: Observing a measurement twice entails that each occurrence has independent noise, so it does make a difference how often a measurement is observed.

Contribution: In this note, we argue that an adjusted sampling without replacement may be the scheme of choice because it is equivalent to a sampling scheme with replacement and exhibits better numerical performance. We illustrate this phenomenon numerically with problems inspired by MRI reconstruction problems.

II. LASSO

Let $\Omega \subset [N]$ be an index set. For a measurement matrix $A_\Omega \in \mathbb{R}^{m\times N}$, whose rows $a_1^T, \dots, a_m^T$ are sampled from an underlying matrix $A \in \mathbb{R}^{N\times N}$, e.g., the discrete Fourier matrix, and a data vector $b = (b_1, \dots b_m) \in \mathbb{R}^m$, we are interested in the high dimensional regression model

$$b = A_\Omega x_0 + \varepsilon, \qquad N \gg m, \tag{1}$$

where $x_0 \in \mathbb{R}^N$ is s_0-sparse and the noise vector $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_{m\times m})$ is assumed to be standard Gaussian distributed with independent components ε_i and standard deviation $\sigma > 0$.

The main goal is to estimate $x_0 \in \mathbb{R}^N$ based on only few data b. An omnipresent unconstrained estimator in the literature is the LASSO [10], denoted by $\hat{x}$, which is the minimizer of

$$\min_{x\in\mathbb{R}^N} \frac{1}{2m}\|A_\Omega x - b\|_2^2 + \lambda\|x\|_1, \tag{2}$$

where $\lambda = \lambda(N, m, \sigma) \in \mathbb{R}$ is a tuning parameter that balances data fidelity and sparsity induced by the ℓ_1-norm. The following well-known oracle inequality establishes a bound on the reconstruction error:

Theorem 1. *[2, Theorem 7.13] Let $x_0 \in \mathbb{R}^N$ be s-sparse. Assume that the measurement matrix satisfies the restricted eigenvalue property over $\operatorname{supp}(x_0)$ with parameter $(\kappa, 3)$. Then, any solution of the LASSO (2) with regularization parameter lower bounded as $\lambda \geq 2\|\frac{A_\Omega^T\varepsilon}{m}\|_\infty$ satisfies the bound*

$$\|\hat{x} - x_0\|_2 \leq \frac{3}{\kappa}\sqrt{s}\lambda. \tag{3}$$

Since the RIP implies the RE condition, the whole theory for the LASSO can be translated for matrices that satisfy the RIP, which is a more common property in the signal processing literature. For a subsampled Fourier matrix $2\|\frac{A_\Omega^T\varepsilon}{m}\|_\infty \leq \frac{\sigma}{\sqrt{m}}(2 + \sqrt{12\log N}) := \lambda_0$, with high probability, see, e.g., [11, Lemma B.3.].

As mentioned in the introduction, when the underlying signal is not sparse in the canonical basis, but should rather be sparsified by the use of a different basis, the sampling strategy needs to compensate for that by using a variable density scheme together with a preconditioning operator. This reduces the local coherence of the overall measurement system – a notion studied in, e.g., [9], [7]. In particular, the following result on the number of required measurements, i.e., to fulfill the RIP, can be established.

Theorem 2. *[9, Theorem 1] Let $\Phi = \{\varphi_j\}_{j=1}^N$ and $\Psi = \{\psi_k\}_{k=1}^N$ be orthonormal bases of $\mathbb{C}^N$. Assume the local coherence of Φ with respect to Ψ is pointwise bounded by the function κ, that is $\sup_{1\leq k\leq N} |\langle\varphi_j, \psi_k\rangle| \leq \kappa_j$. Let $s \gtrsim \log(N)$, suppose*

$$m \gtrsim \delta^{-2}\|\kappa\|_2^2 s \log^3(s)\log(N), \tag{4}$$

and choose m (possibly not distinct) indices $j \in \Omega \subset [N]$ i.i.d. from the probability measure ν on $[N]$ given by $\nu(j) = \frac{\kappa_j^2}{\|\kappa\|_2^2}$. Consider the matrix $A \in \mathbb{C}^{m\times N}$ with entries

$$A_{j,k} = \langle\varphi_j, \psi_k\rangle, \quad j \in \Omega, k \in [N], \tag{5}$$

and consider the diagonal matrix $D = \operatorname{diag}(d) \in \mathbb{C}^N$ with $d_j = \|\kappa\|_2/\kappa_j$. Then with probability at least $1 - N^{-c\log^3(s)}$, the restricted isometry constant δ_s of the preconditioned matrix $\frac{1}{\sqrt{m}}DA$ satisfies $\delta_s \leq \delta$.

In the theorem above, the local coherence appears also in the sampling scheme and it dictates how the measurement matrix should be created by employing a non-uniform sampling strategy. For example, in the MRI setting, if the underlying image is sparsified with a wavelet transform [4], here denoted by T, then, estimates in the local coherence $\sup_{1\leq k\leq N} |\langle\varphi_j, \psi_k\rangle|$ [9] show that the sampling scheme, dictated by the probability measure $\nu(j) = \frac{\kappa_j^2}{\|\kappa\|_2^2}$, will be given by a (non-uniform) subsampled discrete Fourier matrix F_Ω, which is commonly employed in MRI [12]. In this case, the measurement operator A will be given by $A = F_\Omega T$. Theorem 2 says that such operator constitutes a bounded orthonormal system with respect to the uniform measure provided that the set of rows Ω is sampled with replacement. As a consequence, some rows are sampled several times and then discarded with high probability. The goal here is to show how to use this additional information, combined with a sampling without replacement strategy, in order to get a more precise LASSO solution in the case the noise is assumed to follow a statistical distribution.

III. SAMPLING WITH REPLACEMENT VS. SAMPLING WITHOUT REPLACEMENT

We consider two ways for selecting an index set $\Omega \subset [N]$ of size m independently at random with respect to a probability measure ν on $[N]$. This means that every sampling point $\omega_i \in \Omega$, $i \in [m]$, is selected with probability $\mathbb{P}(\omega_i = k) = \nu(k)$. The first way is sampling with replacement. In this scenario all sampling points $\omega_1, \dots, \omega_m$ are chosen independently at random with respect to ν allowing sampling points to be chosen more than once, i.e. $\omega_i = \omega_j$ is possible.

Sampling without replacement assures that the sampling points $\omega_1, \dots, \omega_m$ are different from each other. When we sample the point ω_i we compare if there is a $j \neq i$ with $\omega_i = \omega_j$. In this case we sample the point ω_i again until it differs from $\omega_1, \dots, \omega_{i-1}$.

Our analysis is based on a combination of these sampling methods. For that, we take another view on sampling with replacement. Rather than drawing a fixed number of samples – which could result in different numbers of distinct frequencies sampled depending on how often samples are repeated – one keeps drawing samples until a total of m distinct frequencies have been drawn. That is, the number of draws $m' \geq m$ is now a random variable. In addition to the m different frequencies ω_i observed, we also record for each of them the number c_i of how often it is sampled. Notably, the m frequencies ω_i now follow exactly the same distribution as if they were drawn without replacement. Consequently, it is also possible to convert a sample of m frequencies without replacement into a sample of m' frequencies with replacement by drawing the associated c_i from a (potentially complicated) distribution conditionally on the m frequencies that have been sampled. This argument has also been used in the compressed sensing literature to explain why for worst-case noise, recovery guarantees for frequencies with replacement imply guarantees for sampling without replacement [1, Chapter 12].

For random noise, however, this implication cannot be made as repeated samples in a random noise model are affected by independent noise realizations, so one cannot just take one measurement and repeat it c_i times. The main result of this note is that this issue can be overcome by an appropriate reweighting in the data fidelity term of the LASSO. More precisely, we consider the modified LASSO given by

$$\min_{x\in\mathbb{R}^N} \frac{1}{2m}\|\sqrt{c}\odot(A_\Omega x - b)\|_2^2 + \lambda\|x\|_1, \tag{6}$$

We refer to this last sampling strategy as reweighted sampling without replacement.

IV. MAIN RESULT

Now that we have introduced the different sampling approaches, we state our main result.

Theorem 3. *Let $A \in \mathbb{R}^{N\times N}$ and consider two random matrices created from its rows. Firstly, $A_{\Omega'} \in \mathbb{R}^{m'\times N}$ is created by sampling rows of A with replacement until m distinct rows have been selected, secondly, $A_\Omega \in \mathbb{R}^{m\times N}$ is created by sampling m rows from A without replacement, both according to a probability measure ν on $[N]$. Moreover, conditionally on Ω, let $c \in \mathbb{R}^m$ be a random vector of counts such that combined with the sampling without replacement model, it yields the sampling with replacement model – as described in Section III. Furthermore, let $\sqrt{c}$ be the pointwise square root of its entries. Let $b' = A_{\Omega'}x_0 + \epsilon$ and $b = A_\Omega x_0 + \varepsilon$, where $x_0 \in \mathbb{R}^N$ is s_0-sparse, $\epsilon \sim \mathcal{N}(0, \sigma^2\operatorname{diag}(c))$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_{n\times n})$. Then, the random variables*

$$\begin{aligned}&\arg\min_{x\in\mathbb{R}^N} \|\sqrt{c}\odot(A_\Omega x - b)\|_2^2 + \lambda\|x\|_1 \text{ and}\\ &\arg\min_{x\in\mathbb{R}^N} \|A_{\Omega'}x - b'\|_2^2 + \lambda\|x\|_1\end{aligned} \tag{7}$$

have the same probability distribution[1].

Proof. In the first step we assume that a row of A occurs twice in $A_{\Omega'}$ with independently and equally distributed noise components ϵ_i and ϵ_j. We denote the two samples of this row by a_i and a_j. Then,

$$\begin{aligned}&(\langle a_i, x^* - x_0\rangle - \epsilon_i)^2 + (\langle a_j, x^* - x_0\rangle - \epsilon_j)^2\\ =&(\langle a_i, x^* - x_0\rangle)^2 + (\langle a_j, x^* - x_0\rangle)^2\\ &-2(\langle a_i, x^* - x_0\rangle)^2\epsilon_i - 2(\langle a_j, x^* - x_0\rangle)^2\epsilon_j + \epsilon_j^2 + \epsilon_i^2.\\ =&2(\langle a_i, x^* - x_0\rangle)^2 - 2(\langle a_j, x^* - x_0\rangle)(\epsilon_j + \epsilon_i) + \epsilon_j^2 + \epsilon_i^2\\ =&\Big(\sqrt{2}\langle a_i, x^* - x_0\rangle) - \frac{1}{\sqrt{2}}(\epsilon_j + \epsilon_i)\Big)^2 + \frac{1}{\sqrt{2}}(\epsilon_j - \epsilon_i)^2.\end{aligned}$$

Similarly, if a row a_i is sampled $c_i \in \mathbb{N}$ times, by an induction argument, we have

$$\begin{aligned}&\sum_{j=1}^{c_i}(\langle a_i, x^* - x_0\rangle - \epsilon_j)^2\\ =&\Big(\sqrt{c_i}\langle a_i, x^* - x_0\rangle - \frac{1}{\sqrt{c_i}}\sum_{j=1}^{c_i}\epsilon_j\Big)^2 + \sum_{\substack{j,k=1\\ j<k}}^{c_i}\frac{1}{\sqrt{c_i}}(\epsilon_j - \epsilon_k)^2\\ =&\Big(\sqrt{c_i}\big[\langle a_i, x^* - x_0\rangle - \frac{1}{c_i}\sum_{j=1}^{c_i}\epsilon_j\big]\Big)^2 + \sum_{\substack{j,k=1\\ j<k}}^{c_i}\frac{1}{\sqrt{c_i}}(\epsilon_j - \epsilon_k)^2.\end{aligned}$$

The calculation above shows that

$$\sum_{\substack{j,k=1\\ j<k}}^{c_i}\frac{1}{\sqrt{c_i}}(\epsilon_j - \epsilon_k)^2 + \|\sqrt{c}\odot(A_\Omega x - b)\|_2^2 \tag{8}$$

$$\text{and}\quad \|A_{\Omega'}x - b'\|_2^2 \qquad \forall x \in \mathbb{R}^N$$

have the same probabiliy distribution. Since the first term in (8) is independent of x, the minimizer of the LASSO objective function does not change if this first term is ignored. Therefore, (7) holds. □

To put the theorem into perspective, assume that you have a given budget m of physical measurements that your MRI machine can perform in the allotted time. We now have the choice of either to sample with replacement, that is consider a sampling trajectory that passes through certain points multiple times – in this case, the corresponding measurements are observed multiple times with independent noise realizations – or to sample without replacement. In the former case, the existing theory applies, but due to repetitions, only a smaller number of distinct frequencies is observed. In the latter case, one maximizes the number of distinct frequencies, but the existing theory does not apply.

Equation (7) shows that this obstacle can be overcome by adding a reweighting in the data fidelity term of the LASSO. Namely, the solution of the resulting problem agrees in

[1]Here we assume that both problems have a unique solution which is the case for relevant classes of full row rank A_Ω, e.g., [13, Theorem 1].

distribution with the one of a problem under sampling with replacement, where the existing theory applies. Hence, the solution of the latter problem must be close to the true solution with high probability, which then implies that the same holds also for the solution of the reweighted problem without replacement.

Interestingly, the noise added to these measurements now has a different variance structure. It is given by $\varepsilon_i \sim \frac{1}{c_i}\sum_{j=1}^{c_i} \epsilon_j$ with variances

$$\mathrm{Var}(\epsilon_j) = \frac{1}{c_i}\sum_{j=1}^{c_i} \mathrm{Var}(\epsilon_j) = c_i \mathrm{Var}(\varepsilon) = c_i\sigma^2.$$

While the components of ε are i.i.d., the variances of the components of ϵ depend on c_i, i.e., $\mathrm{Var}(\epsilon_j) = c_i\sigma^2$.

V. NUMERICAL EXPERIMENTS

The numerical experiments are leaned on a simplified MRI setting. For simplicity, we consider the $1d$-case with the model $y = F_\Omega x_0 + \varepsilon$, where $F_\Omega \in \mathbb{C}^{3277\times 8192}$ is a subsampled Fourier matrix, where the rows are sampled according to the probability measure from Theorem 2, simulating the MRI acquisition process, $x_0 \in \mathbb{R}^{8192}$ representing the unknown image and $\varepsilon \sim \mathcal{CN}(0, \sigma^2 I_{m\times m})$ is a complex Gaussian noise vector. For the ground truth signal we assume a sparsity of $s_x = 4096$. The first and the last $\frac{s_x}{2}$ entries of the unknown signal x_0 are ones, the rest in between are zeros. We apply the adjoint Haar transform, such that $x_0 = T^* z_0$. The sparsity of the transformed signal is $s_z = 3$. The relative noise is $\frac{\|\varepsilon\|_2}{\|F_\Omega T^* z_0\|_2} \approx 6.3\%$.

We distinguish between sampling with replacement, sampling without replacement and reweighted sampling with replacement described in Section III with a total number of $m = 0.4N$ rows. In order to guarantee recovery from the LASSO we multiply our model by a diagonal normalization matrix D, which is defined in Theorem 2 and assures that with high probability the normalized measurement matrix $\frac{1}{\sqrt{m}} D F_\Omega H^*$ satisfies the RIP, i.e., $Dy = DF_\Omega H^* z_0 + D\varepsilon$. In the reweighted sampling without replacement regime we additionally weight the matrix with the counting vector $\sqrt{c}$.

We compute for different values of λ the quantities $\|\hat{z}(\lambda) - z_0\|_p$ with $p \in \{1, 2, \infty\}$. We average over 50 independent realisations of the sample scheme and the noise. Figure 1 shows the ℓ_2 error depending on multiples of $\lambda_0 = \frac{\sigma}{\sqrt{m}}(2 + \sqrt{12\log N})$ for sampling with replacement, sampling without replacement and reweighted sampling with replacement. The error of all three strategies decreases as λ grows until a certain minimum is reached around $\lambda = 10\lambda_0$ for the sampling with replacement and sampling without replacement and $\lambda = 35\lambda_0$ for the reweighted sampling strategy, before the error starts to increase again. The difference of the minima can be explained with the help of Theorem 1: Since in our combined method the entries of the measurement matrix and the noise are larger due to the counting vector c, a larger regularization parameter is used and hence leads to smaller ℓ_2 errors. In this sense, our method outperforms the other ones.

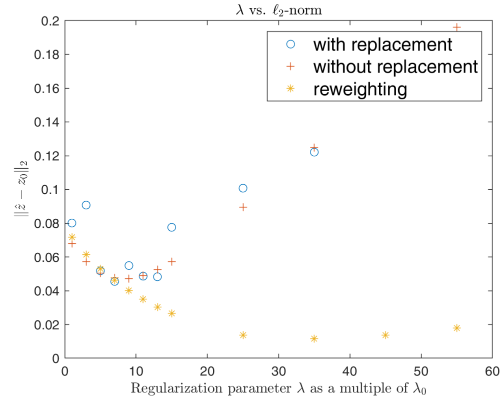

Fig. 1: The ℓ_2 error as a function of the regularization parameter.

The same behavior can be observed for the ℓ_1 and ℓ_∞ in Figure 2 and Figure 3. The spatial shift of the minima is due to the inequality $\|\hat{z} - z_0\|_1 \le \sqrt{|\operatorname{supp}(\hat{z} - z_0)|} \cdot \|\hat{z} - z_0\|_2$ and [14, Theorem 1], respectively.

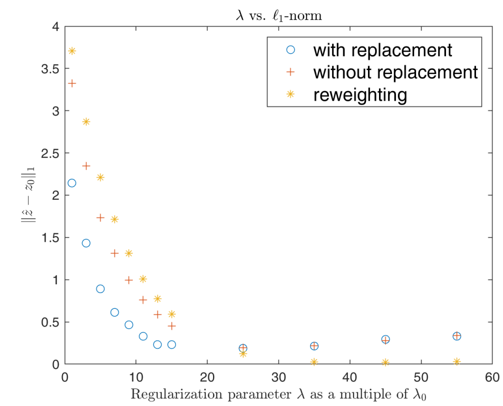

Fig. 2: The ℓ_1 error as a function of the regularization parameter.

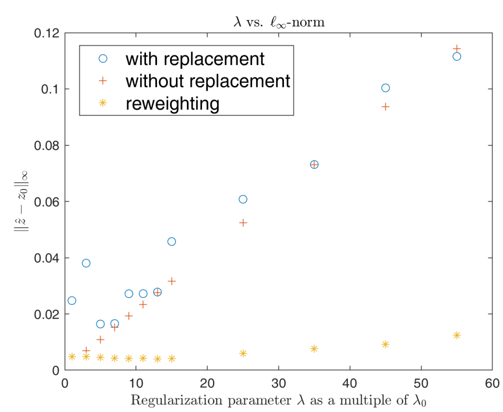

Fig. 3: The ℓ_∞ error as a function of the regularization parameter.

VI. CONCLUSION

In this note we analysed the compressive imaging with statistical noise. We developed a mathematical transition between measurement matrices constructed via sampling with replacement and those constructed via sampling without replacement that allows to apply the theory developed for the former model also in the latter case when a reweighted LASSO reconstruction method is used. The resulting reweighted reconstruction approach for sampling without replacement was shown to numerically outperform both the sampling with replacement strategy and the sampling without replacement strategy with no such reweighting.

VII. REFERENCES

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer New York, New York, NY, 2013.

[2] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48, Cambridge University Press, 2019.

[3] S. A. van de Geer and P. Bühlmann, "On the conditions used to prove oracle results for the LASSO," *Electronic Journal of Statistics*, vol. 3, pp. 1360–1392, 2009.

[4] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[5] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[6] I. Haviv and O. Regev, "The Restricted Isometry Property of Subsampled Fourier Matrices," in *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, Bo'az Klartag and Emanuel Milman, Eds., pp. 163–179. Springer International Publishing, Cham, 2017.

[7] B. Adcock, A. C. Hansen, and B. Roman, "A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements," *IEEE signal processing letters*, vol. 23, no. 5, pp. 732–736, 2016.

[8] S. Brugiapaglia, S. Dirksen, H. C. Jung, and H. Rauhut, "Sparse recovery in bounded Riesz systems with applications to numerical methods for PDEs," *Applied and Computational Harmonic Analysis*, vol. 53, pp. 231–269, 2021.

[9] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE transactions on image processing*, vol. 23, no. 2, pp. 612–622, 2013.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.

[11] F. Hoppe, F. Krahmer, C. Mayrink Verdun, M. I. Menzel, and H. Rauhut, "Uncertainty quantification for sparse Fourier recovery," *arXiv:2212.14864*, 2022.

[12] J. A. Fessler, "Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms," *IEEE signal processing magazine*, vol. 37, no. 1, pp. 33–40, 2020.

[13] H. Zhang, W. Yin, and L. Cheng, "Necessary and sufficient conditions of solution uniqueness in ℓ_1 minimization," *Journal of Optimization Theory and Applications*, vol. 164, no. 1, pp. 109–122, 2015.

[14] M. J. Wainwright, "Sharp thresholds for High-Dimensional and noisy sparsity recovery using ℓ_1-Constrained Quadratic Programming (Lasso)," *IEEE transactions on information theory*, vol. 55, no. 5, pp. 2183–2202, 2009.