Original Paper

Noise Variance Estimation Using Asymptotic Residual in Compressed Sensing

Ryo Hayakawa*

Institute of Engineering, Tokyo University of Agriculture and Technology, Japan

---

ABSTRACT

In compressed sensing, measurements are typically contaminated by additive noise, and therefore, information about the noise variance is often needed to design algorithms. In this paper, we propose a method for estimating the unknown noise variance in compressed sensing problems. The proposed method, called asymptotic residual matching (ARM), estimates the noise variance from a single measurement vector on the basis of the asymptotic result for the ℓ_1 optimization problem. Specifically, we derive the asymptotic residual corresponding to the ℓ_1 optimization and show that it depends on the noise variance. The proposed ARM approach obtains the estimate by comparing the asymptotic residual with the actual one, which can be obtained by empirical reconstruction without the information on the noise variance. For the proposed ARM, we also propose a method to choose a reasonable parameter based on the asymptotic residual. Simulation results show that the proposed noise variance estimation outperforms several conventional methods, especially when the problem size is small. We also show that, by using the proposed method, we can tune the regularization parameter of the ℓ_1 optimization to achieve good reconstruction performance, even when the noise variance is unknown.

---

1   Introduction

Compressed sensing [11–13, 25] has attracted much attention in the field of signal processing [17, 38, 45, 46]. One of the main purposes of compressed sensing is to solve underdetermined linear inverse problems of an unknown vector with a structure such as sparsity. Although the underdetermined problem has an infinite number of solutions in general, we can often reconstruct the unknown vector by using the sparsity as the prior knowledge appropriately. A similar idea can be applied to the reconstruction of other non-sparse structured vectors, e.g., discrete-valued vectors [2, 51], which often appear in wireless communication systems [16, 33, 59].

There are various algorithms proposed for compressed sensing. In greedy algorithms such as matching pursuit (MP) [47] and orthogonal matching pursuit (OMP) [55, 71], we update the support of the estimate of the unknown sparse vector in an iterative manner. Another approach based on message passing, e.g., approximated belief propagation (BP) [40] and approximate message passing (AMP) [5, 26], utilizes a Bayesian framework for the reconstruction of the structured vector. Such message passing-based methods can achieve good reconstruction performance with low complexity for large-scale problems. For small-scale problems with a few hundred unknown variables, however, their performance degrades and the algorithms may even diverge.

Various convex optimization-based approaches have also been studied in the literature on compressed sensing. The most popular convex optimization problem for compressed sensing is ℓ_1 optimization (a.k.a. least absolute shrinkage and selection operator (LASSO) [70]), where the ℓ_1 norm is used as the regularizer to promote the sparsity. The iterative shrinkage thresholding algorithm (ISTA) [19, 22, 29] and the fast iterative shrinkage thresholding algorithm (FISTA) [7] can solve the ℓ_1 optimization problem with feasible computational complexity. Another promising algorithm is the alternating direction method of multipliers (ADMM) [10, 18, 27, 31], which can be applied to a wider class of optimization problems than ISTA and FISTA. Such optimization-based approaches can also be applied to the reconstruction of other non-sparse structured vectors [4, 30, 34, 37]. Unlike the message passing-based methods, the convex optimization-based algorithms converge to the solution of the corresponding optimization problem even for small-scale reconstruction problems.

The measurement vector in compressed sensing is usually contaminated by additive noise in practice. In the design of the algorithms for compressed sensing, information on the noise variance is often required to obtain good

reconstruction performance. In optimization-based approaches, for example, the objective function and/or the constraint in the problem usually include some parameters to be fixed in advance. Since the appropriate value of the parameter depends on the noise variance in general, its information is essential to tune the parameter of the optimization problem, with few exceptions such as square-root LASSO [9, 60]. If the noise variance and the distribution of the unknown vector are known, we can obtain the optimal regularization parameter in terms of the asymptotic mean squared error (MSE) under several assumptions by using some analytical results [5, 26, 65]. Hence, when the noise variance is unknown, we need to estimate it from the measurement vector before the reconstruction of the unknown vector.

Although several estimation methods for the noise variance have been proposed in the context of linear regression in statistics [24, 41, 58], some of them mainly consider non-structured vectors and do not exploit the sparsity of the unknown vector. On the other hand, some sparsity-based methods such as [63, 76] cannot be extended to the case with other non-sparse structured vectors in a trivial manner. A possible exception is AMP-LASSO [6], which is based on the asymptotic analysis of the MSE of LASSO [49, 50]. The estimate of the noise variance by AMP-LASSO is consistent and can be simply calculated using the reconstructed vector by LASSO with a fixed regularization parameter. For small-scale problems, however, we need to choose an appropriate value of the regularization parameter to obtain a good estimate of the noise variance. For more details of related work, see Section 3.

In this paper, we propose a novel estimation method for the noise variance on the basis of the asymptotic analysis for the ℓ_1 optimization. The proposed approach, referred to as asymptotic residual matching (ARM), uses the fact that the residual of the estimate obtained by the ℓ_1 optimization can be well predicted under some assumptions when the problem size is sufficiently large. By using the convex Gaussian min-max theorem (CGMT) [65, 69] and a similar procedure to [48], we derive the asymptotic residual in the large system limit, where the problem size goes to infinity. The asymptotic residual depends on the noise variance, whereas the empirical residual can be computed without using the noise variance because we just need to solve the ℓ_1 optimization problem. We can thus estimate the noise variance by choosing the value whose corresponding asymptotic residual is the closest to the empirical residual. Hence, the proposed noise variance estimation firstly solves the ℓ_1 optimization problem with a fixed regularization parameter and then computes the empirical residual of the reconstructed vector. After that, we obtain the noise variance whose corresponding asymptotic residual matches the empirical residual.

As is the case with other methods such as AMP-LASSO, the estimation performance of the proposed method depends on the value of the regularization parameter. We thus propose a parameter initialization method for the proposed ARM on the basis of the asymptotic residual. The proposed method enables

us to choose a reasonable value of the regularization parameter without the computation of the solution of the ℓ_1 optimization problem. To further improve the estimation performance, we also propose the iterative approach, where we iterate the estimation of the noise variance and the update of the regularization parameter. Hence, unlike the conventional methods, the proposed method can estimate the noise variance without the manual tuning of the regularization parameter. Another advantage of the proposed ARM is that we can easily extend it to the case with other non-sparse structured vectors if the distribution is known. In this paper, we consider the reconstruction of binary vector as an example, which appears in some communication systems such as multiple-input multiple-output (MIMO) signal detection [15, 74].

Simulation results demonstrate that the proposed method can achieve good estimation performance even when the problem size is small. By using the estimate of the noise variance for the choice of the regularization parameter, we can obtain good reconstruction performance in compressed sensing even when the noise variance is unknown.

The rest of the paper is organized as follows. We describe the problem considered in this paper in Section 2 and related work in Section 3. We then provide the analytical results for the residual of the ℓ_1 optimization in Section 4. In Section 5, we explain the proposed noise variance estimation method based on the analytical result. In Section 6, we discuss the extension of the proposed method and show the example for binary vector reconstruction. We demonstrate several simulation results to show the validity of the proposed method in Section 7. Finally, Section 8 presents some conclusions.

In this paper, we use the following notations. We denote the transpose by $(\cdot)^\top$ and the identity matrix by $\boldsymbol{I}$. For a vector $\boldsymbol{a} = [a_1 \cdots a_N]^\top \in \mathbb{R}^N$, the ℓ_1 norm and the ℓ_2 norm are given by $\|\boldsymbol{a}\|_1 = \sum_{n=1}^{N} |a_n|$ and $\|\boldsymbol{a}\|_2 = \sqrt{\sum_{n=1}^{N} a_n^2}$, respectively. We denote the number of nonzero elements of $\boldsymbol{a}$ by $\|\boldsymbol{a}\|_0$. $\operatorname{sign}(\cdot)$ denotes the sign function. For a lower semicontinuous convex function $\zeta : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, we define the proximity operator and the Moreau envelope as

$$\operatorname{prox}_\zeta(\boldsymbol{a}) = \arg\min_{\boldsymbol{u} \in \mathbb{R}^N} \left\{ \zeta(\boldsymbol{u}) + \frac{1}{2} \|\boldsymbol{u} - \boldsymbol{a}\|_2^2 \right\}, \tag{1}$$

$$\operatorname{env}_\zeta(\boldsymbol{a}) = \min_{\boldsymbol{u} \in \mathbb{R}^N} \left\{ \zeta(\boldsymbol{u}) + \frac{1}{2} \|\boldsymbol{u} - \boldsymbol{a}\|_2^2 \right\}, \tag{2}$$

respectively. The probability density function (PDF) and the cumulative distribution function (CDF) of the standard Gaussian distribution are denoted as $p_{\mathrm{G}}(\cdot)$ and $P_{\mathrm{G}}(\cdot)$, respectively. When the PDF of the random variable X is given by p_{X}, we denote $X \sim p_{\mathrm{X}}$. When a sequence of random variables $\{\Theta_n\}$ $(n = 1, 2, \dots)$ converges in probability to Θ, we denote $\Theta_n \xrightarrow{\mathrm{P}} \Theta$ as $n \to \infty$ or $\operatorname{plim}_{n\to\infty} \Theta_n = \Theta$.

2 Noise Variance Estimation in Compressed Sensing

A standard problem in compressed sensing is the reconstruction of an N dimensional sparse vector $\boldsymbol{x} = [x_1 \; \cdots \; x_N]^\top \in \mathbb{R}^N$ from its linear measurements given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \in \mathbb{R}^M, \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a known measurement matrix and $\boldsymbol{v} \in \mathbb{R}^M$ is an additive noise vector. We denote the measurement ratio by $\Delta = M/N$. In the scenario of compressed sensing, we focus on the underdetermined case with $\Delta < 1$ and utilize the sparsity of $\boldsymbol{x}$ as the prior knowledge for the reconstruction.

One of the most famous convex optimization problems for compressed sensing is the ℓ_1 optimization given by

$$\hat{\boldsymbol{x}}(\lambda) = \arg\min_{\boldsymbol{s} \in \mathbb{R}^N} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2 + \lambda f(\boldsymbol{s}) \right\}, \tag{4}$$

where $f(\boldsymbol{s}) = \|\boldsymbol{s}\|_1$ is the ℓ_1 regularizer to promote the sparsity of the estimate $\hat{\boldsymbol{x}}(\lambda)$ of the unknown vector $\boldsymbol{x}$. The regularization parameter $\lambda \, (> 0)$ controls the balance between the data fidelity term $\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2$ and the ℓ_1 regularization term $\lambda f(\boldsymbol{s})$. Since the ℓ_1 optimization is the convex optimization problem, the sequence converging to the global optimum can be obtained by several convex optimization algorithms [7, 10, 18, 19].

In this paper, we assume that the noise variance σ_{v}^2 is unknown, and tackle the problem of estimating σ_{v}^2 from the single measurement $\boldsymbol{y}$ and the corresponding measurement matrix $\boldsymbol{A}$. The knowledge of the noise variance σ_{v}^2 is important to design the algorithms for compressed sensing. For the optimization problems in (4), for example, the reconstruction performance largely depends on the parameter λ and its appropriate value is different depending on the noise variance. In fact, by using the AMP framework or the CGMT framework, the asymptotically optimal parameter minimizing MSE can be obtained under some assumptions when the noise variance is known [49, 65]. Hence, the accurate estimate of the noise variance is significant to achieve good reconstruction performance in convex optimization-based compressed sensing. For other approaches, the information on the noise variance would also be helpful to design the algorithm.

3 Related Work

In statistics, several estimation methods for the noise variance have been discussed in the context of linear regression [23, 24, 39]. A method using the residual of the ridge regression has been proposed in [41], where simulation

results show that it outperforms some conventional approaches. The signal-to-noise ratio (SNR) estimation method in [62] is also based on the analysis of the ridge regularized least squares. Although it has good estimation performance when the number of measurements is sufficiently large, the performance degrades for underdetermined problems like compressed sensing. Moreover, the above methods mainly focus on the non-structured unknown vectors, and hence they do not take advantage of the sparsity in the estimation.

Some sparsity-aware methods have also been proposed for the noise variance estimation, e.g., scaled LASSO [63] and refitted cross-validation [28]. In [58], the authors have compared the performance of several estimators and have concluded that a promising estimator is given by

$$\hat{\sigma}_{\mathrm{v}}^2 = \frac{1}{M - \|\hat{\boldsymbol{x}}(\lambda)\|_0} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2. \tag{5}$$

For the estimator in (5), however, the regularization parameter λ significantly affects the estimation performance and the parameter should be carefully selected. Although the cross-validation technique can be used for the choice of λ, it increases the computational cost of the estimation. Even if we use several approximation techniques [32, 52, 57, 61], we need to obtain the estimate $\hat{\boldsymbol{x}}(\lambda)$ for various values of λ to choose its appropriate value. Moreover, these sparsity-aware methods only consider the sparse unknown vector, and the extension of other non-sparse structured vectors is not trivial.

Another LASSO-based method has also been proposed in [6] on the basis of the analysis of the AMP algorithm [5, 26]. The estimate by AMP-LASSO can be written as

$$\hat{\sigma}_{\mathrm{v}}^2 = \Delta\hat{\tau}^2 - \hat{R}(\hat{\tau}), \tag{6}$$

where $\hat{\tau} = \sqrt{N} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2 / (M - \|\hat{\boldsymbol{x}}(\lambda)\|_0)$ and

$$\hat{R}(\tau) = \tau^2 \left(\frac{2 \|\hat{\boldsymbol{x}}(\lambda)\|_0}{N} - 1 \right) + \frac{N \left\| \boldsymbol{A}^\top (\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)) \right\|_2^2}{(M - \|\hat{\boldsymbol{x}}(\lambda)\|_0)^2} \tag{7}$$

(under Assumption 1 below). The estimate in (6) is consistent and hence the noise variance is well predicted in large-scale problems. Simulation results in [6] show that the estimation performance of AMP-LASSO is better than several conventional methods such as [28, 63]. For small-scale problems, however, we need to choose an appropriate value of the regularization parameter in LASSO to obtain a good estimate of the noise variance. Moreover, the extension to the structure other than sparsity is not discussed explicitly.

Non-asymptotic and asymptotic analyses have been discussed for the residual $\|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2$ of the LASSO problem in [48]. Moreover, the tuning

method for the regularization parameter λ has been proposed on the basis of the analysis. However, the regularizer other than the ℓ_1 regularizer has not been considered explicitly in the paper. Furthermore, for the tuning of the regularization parameter, we need to solve the optimization problem for many values of λ, which increases the computational cost.

Although we focus on the case with measurement matrices $\boldsymbol{A}$ composed of independent and identically distributed (i.i.d.) Gaussian elements in this paper, the performance analyses for non-i.i.d. cases have been discussed in several papers, e.g., [8, 14]. Especially, in [8], the out-of-sample error has been analyzed for general regularizers, including the ℓ_1 regularizer and the elastic-net regularizer. The noise variance estimation has also been considered in [8], and the estimate is equivalent to that of AMP-LASSO in the case with i.i.d. Gaussian distribution.

4 Asymptotic Residual for ℓ_1 Optimization

In this section, by using the CGMT framework [65, 69], we provide an asymptotic result for the ℓ_1 optimization in (4), which will be used in the proposed noise variance estimation in Section 5. Although a part of the result can be derived from the general CGMT-based analysis in [65], we here derive the explicit formula required in the proposed method. We characterize the asymptotic property of the residual $\|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2$ in the following. It should be noted that a similar analysis has been discussed in [48] for the LASSO problem, though we use more general notation for the regularizer in this paper and actually consider a different problem in Section 6.

In the analysis, we use the following assumption.

Assumption 1. *The unknown vector $\boldsymbol{x}$ is composed of i.i.d. random variables with a distribution $p_{\mathrm{X}}(x)$ which have some mean and variance. The measurement matrix $\boldsymbol{A}$ is composed of i.i.d. Gaussian random variables with zero mean and variance $1/N$. The noise vector $\boldsymbol{v}$ is also Gaussian with mean $\boldsymbol{0}$ and covariance matrix $\sigma_{\mathrm{v}}^2\boldsymbol{I}$.*

In Assumption 1, we assume the Gaussian measurement matrix because it is required to apply CGMT in a rigorous manner. However, the universality [6, 53, 54] of random matrices suggests that the analytical result also holds for other i.i.d. measurement matrix. In fact, the simulation result in [3] shows that the result of the CGMT-based analysis is valid even for the measurement matrix from Bernoulli or Laplace distribution. Hence, it would be possible to utilize our theoretical results for such cases in practice.

By using the CGMT framework [65], we provide the asymptotic property of the residual $\|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2$ in the following. It should be noted that the

standard CGMT-based analysis gives the asymptotic error performance such as MSE, which is different from the residual analyzed here. In the theorem, we consider the large system limit $N, M \to \infty$ with the fixed ratio $\Delta = M/N$, which we simply denote as $N \to \infty$ in this paper.

Theorem 1. *We assume that Assumption 1 is satisfied. We also assume that the optimization problem $\min_{\alpha>0} \max_{\beta\geq 0} F(\alpha, \beta)$ has a unique optimizer (α^*, β^*)[1], where*

$$F(\alpha, \beta) = \frac{\alpha\beta\sqrt{\Delta}}{2} + \frac{\sigma_{\mathrm{v}}^2\beta\sqrt{\Delta}}{2\alpha} - \frac{1}{2}\beta^2 - \frac{\alpha\beta}{2\sqrt{\Delta}}$$
$$+ \frac{\beta\sqrt{\Delta}}{\alpha}\mathrm{E}\left[\mathrm{env}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(X + \frac{\alpha}{\sqrt{\Delta}}G\right)\right] \tag{8}$$

and $X \sim p_{\mathrm{X}}, G \sim p_{\mathrm{G}}$. Then, the asymptotic value of the objective function in (4) and the residual for the optimizer $\hat{\boldsymbol{x}}(\lambda)$ are given by

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\left(\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2 + \lambda f(\hat{\boldsymbol{x}}(\lambda))\right) = F(\alpha^*, \beta^*), \tag{9}$$

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2 = (\beta^*)^2, \tag{10}$$

respectively.

Proof. See Section A.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

To compute α^* and β^* in Theorem 1, we need to optimize the function $F(\alpha, \beta)$ in (8). Fortunately, for some distribution $p_{\mathrm{X}}(x)$, we can write the expectation $\mathrm{E}\left[\mathrm{env}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(X + \frac{\alpha}{\sqrt{\Delta}}G\right)\right]$ in (8) with an explicit formula. For example, when the distribution of the unknown vector is given by the Bernoulli-Gaussian distribution as

$$p_{\mathrm{X}}(x) = p_0\delta_0(x) + (1 - p_0)p_{\mathrm{G}}(x), \tag{11}$$

the expectation can be easily computed with the PDF and CDF of the standard Gaussian distribution, where $\delta_0(\cdot)$ denotes the Dirac delta function and $p_0 \in (0, 1)$. For details of the derivation, see Section A.2. For the Bernoulli distribution given by

$$p_{\mathrm{X}}(x) = p_0\delta_0(x) + (1 - p_0)\delta_0(x - 1), \tag{12}$$

---

[1]The uniqueness can be proven under some conditions. For example, if the set of minimizers of the problem over α is bounded, the uniqueness of α^* is guaranteed. However, it would be difficult to eliminate the assumption completely in the general case. For detailed discussions, see [65, Remark 19].
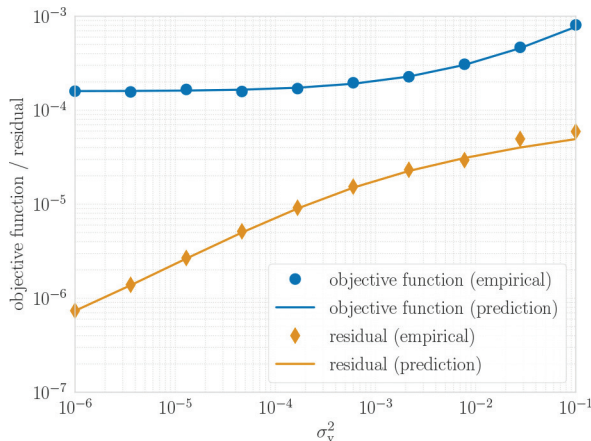
Figure 1: Objective function and residual for the optimizer ($N = 100$, $M = 90$, $p_0 = 0.8$, $\lambda = 0.001$, $p_{\mathrm{X}}(x)$: Bernoulli-Gaussian distribution).

we can also obtain the explicit form of the expectation in a similar way. In such case, we can easily optimize $F(\alpha, \beta)$ by line search techniques such as ternary search and golden-section search [44]. When the exact computation of the expectation is difficult, we can approximate it by the Monte Carlo method with many realizations of X and G.

From Theorem 1, we can predict the optimal value of the objective function and the residual in the empirical reconstruction for compressed sensing problems. Figure 1 shows the comparison between the empirical values and their prediction, where $N = 100$ and $M = 90$. The distribution of the unknown vector $\boldsymbol{x}$ is Bernoulli-Gaussian in (11) with $p_0 = 0.8$. In the figure, 'empirical' means the empirical value of the objective function $\frac{1}{N} \left(\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2 + \lambda f\left(\hat{\boldsymbol{x}}(\lambda)\right) \right)$ and the residual $\frac{1}{N} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2$, where $\hat{\boldsymbol{x}}(\lambda)$ is obtained by the ℓ_1 optimization in (4) with $\lambda = 0.001$. The empirical results are averaged over 100 independent trials. For the reconstruction, we use the LASSO solver of scikit-learn [56]. In Figure 1, we also plot the asymptotic value obtained from Theorem 1 as 'prediction'. We can see that the empirical value agrees well with the theoretical prediction for both the objective function and the residual.

5 Proposed Noise Variance Estimation

In this section, we propose an algorithm for the estimation of the noise variance σ_{v}^2 on the basis of the asymptotic analysis in Section 4.

5.1 Asymptotic Residual Matching

The proposed method uses the fact that the residual $\mathrm{Res}\,(\hat{\boldsymbol{x}}(\lambda)) \coloneqq \frac{1}{N} \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda)\|_2^2$ can be approximated by $(\beta^*)^2$ from (10) when N and M are sufficiently large. Since the function $F(\alpha, \beta)$ to be optimized depends on the regularization parameter λ, the noise variance σ_{v}^2, and the probability of zero p_0, the value of the optimal β^* can be considered as a function of $(\lambda, \sigma_{\mathrm{v}}^2, p_0)$. To explicitly show the dependency, we denote β^* as $\beta^*(\lambda, \sigma_{\mathrm{v}}^2, p_0)$ hereafter. On the other hand, we can calculate the empirical estimate $\hat{\boldsymbol{x}}(\lambda)$ and the corresponding residual $\mathrm{Res}\,(\hat{\boldsymbol{x}}(\lambda))$ from (4) without using σ_{v}^2 in the reconstruction. We can thus estimate the noise variance by choosing σ^2 which minimizes the difference $\left|\beta^*(\lambda, \sigma^2, \hat{p}_0)^2 - \mathrm{Res}\,(\hat{\boldsymbol{x}}(\lambda))\right|$, where $\hat{p}_0$ is the estimate of the probability p_0. Hence, the proposed estimate of the noise variance is given by

$$\hat{\sigma}_{\mathrm{v}}^2 = \arg\min_{\sigma^2 > 0} \left|\beta^*(\lambda, \sigma^2, \hat{p}_0)^2 - \mathrm{Res}\,(\hat{\boldsymbol{x}}(\lambda))\right|. \tag{13}$$

In the proposed optimization problem (13), we need the estimate of the probability p_0 when p_0 is unknown. In this paper, we use the rough estimate given by $\hat{p}_0 = 1 - \|\boldsymbol{y}\|_2^2/M$ on the basis of [42, Eq. (10)], which means that $\frac{N}{M}\|\boldsymbol{y}\|_2^2$ is an estimate of $\|\boldsymbol{x}\|_2^2$ for the measurement matrix satisfying Assumption 1. For simplicity, we here assume that the second moment of the non-zero value of $X \sim p_X$ is 1 as in (11) and (12). The problem in (13) is a scalar optimization problem over σ^2, and hence the optimal value can be obtained by line search methods such as the ternary search and the golden-section search [44].

Remark 1 (Advantage of Using Residual of ℓ_1 Optimization)**.** The proposed estimation method uses the asymptotic result for the residual of the ℓ_1 optimization problem. Although we can use the asymptotic value of the objective function in (9) for the noise variance estimation, the performance would be worse in that case. This is because the line of the objective function is flat especially when the noise variance σ_{v}^2 is small as shown in Figure 1. The conventional SNR estimation method in [62] has the same problem because it utilizes the asymptotic result for the objective function of the ridge regularized least squares. In fact, the simulation results in [62] show that the estimation performance becomes worse when the linear system is underdetermined. Another reason for the performance degradation is that the reconstruction performance of the ridge regularized least squares severely degrades for underdetermined problems like compressed sensing. On the other hand, as shown in Figure 1, the residual of the ℓ_1 optimization decreases more rapidly than the objective function as the noise variance σ_{v}^2 decreases. We thus conclude that we should use not the objective function but the residual for the noise variance estimation.

5.2   Initialization of λ

Since the prediction of the residual from Theorem 1 is not exactly accurate for finite N, the estimation performance of the proposed optimization problem (13) depends on the parameter λ. Figure 2 shows the asymptotic residual $\beta^*(\lambda, \sigma^2, p_0)^2$ for different values of λ when $\Delta = 0.8$ and $p_0 = 0.9$. From the figure, we can see that the slope of the line depends on λ. In the case of Figure 2, it is difficult to distinguish the noise variance between 10^{-6} and 10^{-4} if we use the regularization parameter $\lambda = 0.1$ because the line of the asymptotic residual with $\lambda = 0.1$ is flat in the range. Moreover, if the empirical value of the residual is smaller than the line unfortunately, there might be no positive candidate of σ_v^2. From the above discussion, it would be better to use $\lambda = 0.005$ when the true noise variance σ_v^2 is small. On the other hand, when $\sigma_v^2 = 10^{-1}$, for example, the choice $\lambda = 0.1$ seems the best of the four in Figure 2 because it has the steepest slope around $\sigma_v^2 = 10^{-1}$. We need to choose an appropriate value of λ to achieve better estimation performance.
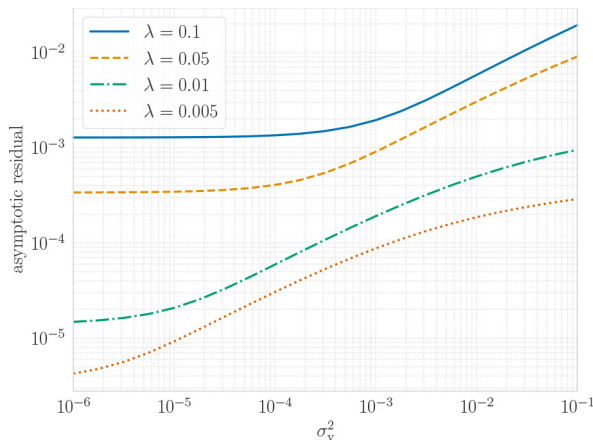

Figure 2: Asymptotic residual of the ℓ_1 optimization ($\Delta = 0.8$, $p_0 = 0.9$).

To tackle this problem, we propose an initialization method based on the max-min approach. We define the quantity

$$D(\lambda, \sigma^2, \hat{p}_0) = \frac{\beta^* \left(\lambda, (1 + \varepsilon)\sigma^2, \hat{p}_0 \right)^2}{\beta^* \left(\lambda, \sigma^2, \hat{p}_0 \right)^2}, \tag{14}$$

which represents how much β^* increases when the value of σ^2 increases to $(1+\varepsilon)\sigma^2$ ($\varepsilon > 0$). Since the scale of β^* is quite different for different λ and σ^2 as shown in Figure 2, we take the ratio of $\beta^* \left(\lambda, (1 + \varepsilon)\sigma^2, \hat{p}_0 \right)^2$ to $\beta^* \left(\lambda, \sigma^2, \hat{p}_0 \right)^2$. The larger $D(\lambda, \sigma^2, \hat{p}_0)$ is, the more rapidly $(\beta^*)^2$ increases along with the

increase of σ^2. Hence, from the discussion of the previous paragraph, λ should be chosen so that $D(\lambda, \sigma^2)$ becomes large. Since the noise variance σ^2 is unknown of course, we here adopt the max-min approach to obtain the proper regularization parameter as

$$\lambda_{\mathrm{prop}}(\hat{p}_0) = \underset{0 \leq \lambda \leq \lambda_{\max}}{\arg \max} \left\{ \min_{\sigma^2 \in \Sigma} D(\lambda, \sigma^2, \hat{p}_0) \right\}, \tag{15}$$

where $\lambda_{\max}$ restricts the range of λ and Σ denotes the set of the candidate values for the noise variance, e.g., $\Sigma = \left\{ 10^{-5}, 10^{-3}, 10^{-1} \right\}$. Note that we do not require that the true noise variance σ_{v}^2 is included in Σ. From (15), we can choose a reasonable regularization parameter λ in the sense that it maximizes $D(\lambda, \sigma^2, \hat{p}_0)$ for the worst $\sigma^2 \in \Sigma$, without the empirical reconstruction of $\boldsymbol{x}$.

5.3   Iterative Estimation

To improve the performance, we also propose an iterative approach as in Algorithm 1. We firstly compute the initial regularization parameter $\lambda_0 = \lambda_{\mathrm{prop}}(\hat{p}_0)$ with (15), and then iterate the updates of the estimated noise variance $\hat{\sigma}_{\mathrm{v},t}^2$ and the regularization parameter λ_t, where t denotes the iteration index. At the t-th iteration, the estimate of the noise variance $\hat{\sigma}_{\mathrm{v},t}^2$ is calculated by solving (13) with $\lambda = \lambda_{t-1}$. Using the estimate $\hat{\sigma}_{\mathrm{v},t}^2$, we update the regularization parameter as

$$\lambda_t = \underset{0 \leq \lambda \leq \lambda_{\max}}{\arg \max} D(\lambda, \hat{\sigma}_{\mathrm{v},t}^2, \hat{p}_0). \tag{16}$$

Algorithm 1 Proposed Asymptotic Residual Matching (ARM)

Input: measurement vector $\boldsymbol{y}$, measurement matrix $\boldsymbol{A}$, estimated probability $\hat{p}_0$

Output: estimated noise variance $\hat{\sigma}_{\mathrm{v},T}^2$

1: $\lambda_0 = \underset{0 \leq \lambda \leq \lambda_{\max}}{\arg \max} \left\{ \min_{\sigma^2 \in \Sigma} D(\lambda, \sigma^2, \hat{p}_0) \right\}$

2: **for** $t = 1$ to T **do**

3:      Solve (4) with $\lambda = \lambda_{t-1}$ and obtain $\hat{\boldsymbol{x}}(\lambda_{t-1})$.

4:      $\mathrm{Res}\left(\hat{\boldsymbol{x}}(\lambda_{t-1})\right) = \dfrac{1}{N} \left\| \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}(\lambda_{t-1}) \right\|_2^2$

5:      $\hat{\sigma}_{\mathrm{v},t}^2 = \underset{\sigma^2 > 0}{\arg \min} \left| \beta^* \left(\lambda_{t-1}, \sigma^2, \hat{p}_0\right)^2 - \mathrm{Res}\left(\hat{\boldsymbol{x}}(\lambda_{t-1})\right) \right|$

6:      $\lambda_t = \underset{0 \leq \lambda \leq \lambda_{\max}}{\arg \max} D\left(\lambda, \hat{\sigma}_{\mathrm{v},t}^2, \hat{p}_0\right)$

7: **end for**

---

to obatin a good regularization parameter for $\hat{\sigma}^2_{\mathrm{v},t}$. If the estimate $\hat{\sigma}^2_{\mathrm{v},t}$ is closer to the true value σ^2_{v} than $\hat{\sigma}^2_{\mathrm{v},t-1}$, the new parameter λ_t is expected to be better than the preivious parameter λ_{t-1}. After T iterations, the proposed ARM in Algorithm 1 outputs the final estimate of the noise variance $\hat{\sigma}^2_{\mathrm{v},T}$.

6   Extension to Other Structured Vectors

Although we have focused on the reconstruction of the sparse vector in the previous sections, the proposed approach using the asymptotic residual can also be applied to the reconstruction of other non-sparse structured vectors. For example, the noise variance estimation with the proposed ARM approach can be utilized in the reconstruction of discrete-valued vectors because the CGMT-based analysis has been applied to the problem [36, 67]. In this section, we mainly describe the noise variance estimation for the binary vector reconstruction as the simplest example.

In the binary vector reconstruction, we estimate the unknown binary vector $\boldsymbol{x}_{\mathrm{b}} \in \{1, -1\}^N$ from its linear measurements $\boldsymbol{y}_{\mathrm{b}} = \boldsymbol{A}\boldsymbol{x}_{\mathrm{b}} + \boldsymbol{v} \in \mathbb{R}^M$. In this seciton, we consider the unknown vector $\boldsymbol{x}_{\mathrm{b}}$ with the distribution

$$p_{\mathrm{X,b}}(x) = \frac{1}{2}\delta_0(x + 1) + \frac{1}{2}\delta_0(x - 1). \tag{17}$$

Such problem often appears in several communication systems, such as the MIMO signal detection [15, 74] and the multiuser detection [72]. As in the sparse vector reconstruction discussed in the previous sections, we require the information on the noise variance to obtain better performance with various methods [20, 21, 33, 59] for the binary vector reconstruction.

A simple approach for the binary vector reconstruction is the box relaxation method [64, 67, 75], which solves the optimization problem

$$\hat{\boldsymbol{x}}_{\mathrm{b}} = \arg\min_{\boldsymbol{s}\in[-1,1]^N} \left\{ \frac{1}{2}\left\| \boldsymbol{y}_{\mathrm{b}} - \boldsymbol{A}\boldsymbol{s} \right\|^2_2 \right\}. \tag{18}$$

Using the indicator function

$$f_{\mathrm{b}}(\boldsymbol{s}) = \begin{cases} 0 & (\boldsymbol{s} \in [-1, 1]^N) \\ \infty & (\boldsymbol{s} \notin [-1, 1]^N) \end{cases}, \tag{19}$$

we can rewrite the box relaxation problem in (18) as

$$\hat{\boldsymbol{x}}_{\mathrm{b}} = \arg\min_{\boldsymbol{s}\in\mathbb{R}^N} \left\{ \frac{1}{2}\left\| \boldsymbol{y}_{\mathrm{b}} - \boldsymbol{A}\boldsymbol{s} \right\|^2_2 + f_{\mathrm{b}}(\boldsymbol{s}) \right\}. \tag{20}$$

The asymptotic analysis in Theorem 1 can be applied to the optimization problem in (20). The following corollary shows the result of the analysis, which can be proven in the same way as Theorem 1.

Corollary 1. *We assume that Assumption 1 is satisfied. We also assume that the optimization problem $\min_{\alpha>0} \max_{\beta\geq0} F_\mathrm{b}(\alpha,\beta)$ has a unique optimizer $(\alpha_\mathrm{b}^*, \beta_\mathrm{b}^*)$, where*

$$F_\mathrm{b}(\alpha,\beta) = \frac{\alpha\beta\sqrt{\Delta}}{2} + \frac{\sigma_\mathrm{v}^2\beta\sqrt{\Delta}}{2\alpha} - \frac{1}{2}\beta^2 - \frac{\alpha\beta}{2\sqrt{\Delta}}$$
$$+ \frac{\beta\sqrt{\Delta}}{\alpha}\mathrm{E}\left[\mathrm{env}_{\frac{\alpha}{\beta\sqrt{\Delta}}f_\mathrm{b}}\left(X_\mathrm{b} + \frac{\alpha}{\sqrt{\Delta}}G\right)\right] \tag{21}$$

and $X_\mathrm{b} \sim p_{\mathrm{X,b}}, G \sim p_\mathrm{G}$. Then, the asymptotic value of the objective function in (18) and the residual for the optimizer $\hat{\boldsymbol{x}}_\mathrm{b}$ are given by

$$\mathrm{plim}_{N\to\infty} \frac{1}{N}\left(\frac{1}{2}\|\boldsymbol{y}_\mathrm{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_\mathrm{b}\|_2^2\right) = F_\mathrm{b}\left(\alpha_\mathrm{b}^*, \beta_\mathrm{b}^*\right), \tag{22}$$

$$\mathrm{plim}_{N\to\infty} \frac{1}{N}\|\boldsymbol{y}_\mathrm{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_\mathrm{b}\|_2^2 = (\beta_\mathrm{b}^*)^2, \tag{23}$$

respectively.

Thus, we can estimate the noise variance in the binary vector reconstruction by using the proposed ARM. It should be noted that we do not require the tuning of the regularization parameter in this case because the optimization problem in (18) does not contain any regularization parameter. Hence, the estimate of the noise variance can be obtained by the non-iterative approach as shown in Algorithm 2. Note that the information of the noise variance is useful for other reconstruction methods such as [20, 21, 33, 59], though the box relaxation does not include any parameter to be tuned.

Algorithm 2 Proposed ARM for binary vector reconstruction

Input: measurement vector $\boldsymbol{y}_\mathrm{b}$, measurement matrix $\boldsymbol{A}$
Output: estimated noise variance $\hat{\sigma}_\mathrm{v}^2$
 1: Solve (18) and obtain $\hat{\boldsymbol{x}}_\mathrm{b}$.
 2: $\mathrm{Res}(\hat{\boldsymbol{x}}_\mathrm{b}) = \dfrac{1}{N}\|\boldsymbol{y}_\mathrm{b} - \boldsymbol{A}\hat{\boldsymbol{x}}_\mathrm{b}\|_2^2$
 3: $\hat{\sigma}_\mathrm{v}^2 = \underset{\sigma^2>0}{\arg\min}\ \left|\beta_\mathrm{b}^*\left(\sigma^2\right)^2 - \mathrm{Res}(\hat{\boldsymbol{x}}_\mathrm{b})\right|$

The left hand sides of (22) and (23) differ only by a factor of two because $f_\mathrm{b}(\hat{\boldsymbol{x}}_\mathrm{b}) = 0$. It might be interesting to investigate the relation between the right-hand sides of (22) and (23).

The proposed ARM can also be applied to the binary sparse vector reconstruction and more general discrete-valued vector reconstruction [3, 36, 51] by using Theorem 1 with the corresponding distribution p_X and the proper regularizer $f(\cdot)$.

7   Simulation Results

In this section, we show some simulation results to demonstrate the performance of the proposed noise variance estimation. In the simulations, we compare the following methods.

- *ARM*: the noise variance estimation with the proposed ARM in Algorithms 1 or 2. For the optimization of σ^2 in (13) and λ in (15), (16), we use the solver scipy.optimize.minimize_scalar for scalar minimization in scipy [73]. The set Σ in (15) is fixed as $\Sigma = \left\{ 10^{-5}, 10^{-3}, 10^{-1} \right\}$ and $\lambda_{\max} = 1$ in the simulations. The value of ε in (14) is set as $\varepsilon = 0.1$.

- *AMP-LASSO*: the estimation method using AMP-LASSO given by (6) [6]. Since the tuning of the regularization parameter for AMP-LASSO has not been discussed in the paper, we use the proposed value $\lambda_{\text{prop}}(\hat{p}_0)$ unless otherwise stated.

- *scaled residual*: the estimation method using the scaled residual given by (5).

- *ridge regularization-based method*: the conventional SNR estimation method in [62] based on the asymptotic analysis of ridge regularized least squares. The regularization parameters are the same as those in [62].

- *ML (oracle)*: the maximum likelihood (ML) approach when the true sparse vector $\boldsymbol{x}$ is known. The estimate of σ_{v}^2 is given by $\hat{\sigma}_{\text{v}}^2 = \frac{1}{M} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_2^2$. Note that $\boldsymbol{x}$ is unknown in the other methods.

In all methods, the noise variance is estimated in the range $[10^{-6}, 1]$. The measurement matrix $\boldsymbol{A}$ and the noise vector $\boldsymbol{v}$ satisfy Assumption 1 in the simulations.

7.1   Sparse Vector Reconstruction

We first examine the effect of the regularization parameter λ in the noise variance estimation for the sparse vector reconstruction. Figure 3 shows the estimate of the noise variance $\hat{\sigma}_{\text{v}}^2$ versus the regularization parameter λ for $N = 200$ and $\Delta = 0.7$. The distribution of the unknown vector is the Bernoulli-Gaussian distribution in (11) with $p_0 = 0.9$. The true noise
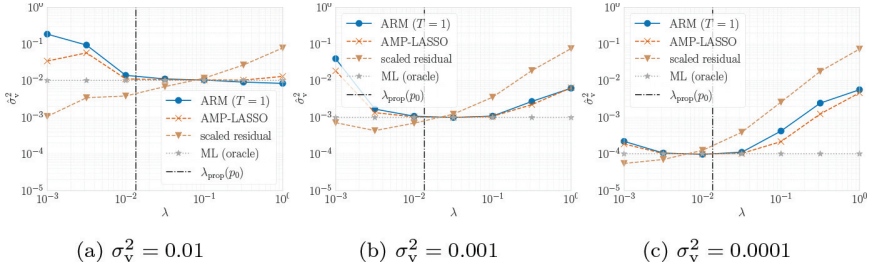
Figure 3: Estimated $\hat{\sigma}_v^2$ versus λ ($N = 200, \Delta = 0.7, p_0 = 0.9$, $p_X(x)$: Bernoulli-Gaussian distribution).
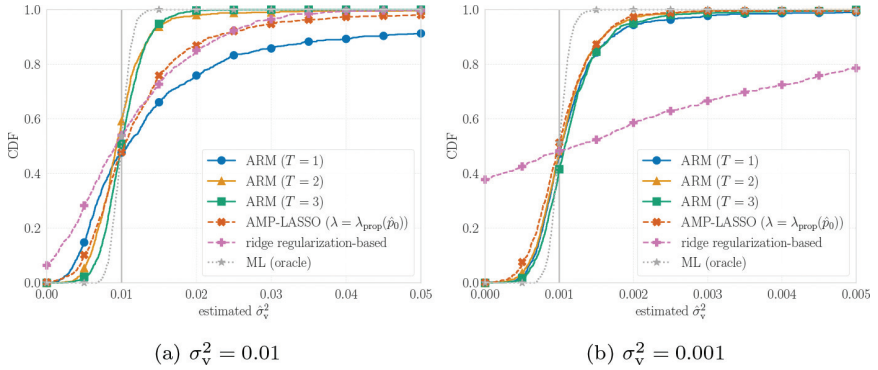

Figure 4: CDF of estimated $\hat{\sigma}_v^2$ ($N = 200$, $\Delta = 0.6$, $p_0 = 0.9$, $p_X(x)$: Bernoulli-Gaussian distribution).

variance is set as $\sigma_v^2 = 0.01, 0.001$, and 0.0001 in Figures 3(a), 3(b) and 3(c), respectively. To solve (4) in ARM, AMP-LASSO, and scaled residual, we use the LASSO solver of scikit-learn [56]. The estimated value is averaged over 100 independent trials. In the figures, the black vertical line shows the value of the proposed regularization parameter $\lambda_{\mathrm{prop}}(p_0)$ with the true probability p_0. Although the estimation performance depends on λ, the proposed regularization parameter can achieve good performance in both ARM and AMP-LASSO for $\sigma_v^2 = 0.01, 0.001$, and 0.0001. We can also see that the performance of the scaled residual is worse than the other methods.

Figure 4 shows the histogram of the empirical CDF of the estimated $\hat{\sigma}_v^2$, where $N = 200$, $\Delta = 0.6$, and $p_0 = 0.9$. The histogram is obtained from 1000 independent trials. Since the true noise variance is set to $\sigma_v^2 = 0.01$ in Figure 4(a), it is better that the CDF rapidly increases around $\sigma_v^2 = 0.01$. From the figure, we can see that the CDF of the proposed ARM with $T = 3$ increases around $\sigma_v^2 = 0.01$ more rapidly than AMP-LASSO and the ridge regularization-based method. This means that the proposed method obtains
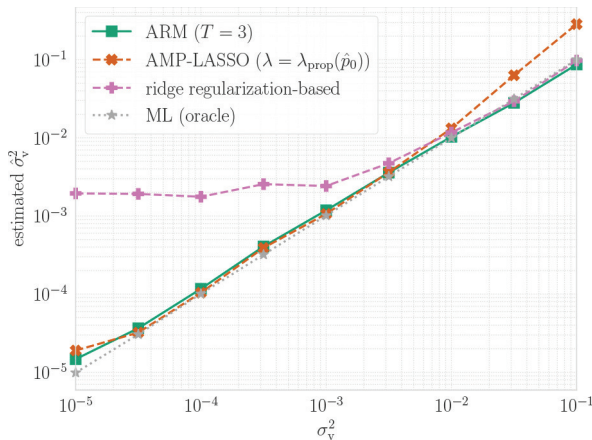
Figure 5: Estimated $\hat{\sigma}_{\mathrm{v}}^2$ versus σ_{v}^2 ($N = 200$, $\Delta = 0.6$, $p_0 = 0.9$, $p_{\mathrm{X}}(x)$: Bernoulli-Gaussian distribution).

the estimate near the true value with a higher probability. The figure also shows that the performance of the proposed ARM improves as the number of iterations T increases. Figure 4(b) shows the performance for $\sigma_{\mathrm{v}}^2 = 0.001$, where the proposed ARM and AMP-LASSO achieves similar performance. However, it should be noted that we use the proposed regularization parameter $\lambda_{\mathrm{prop}}(\hat{p}_0)$ for AMP-LASSO. The performance of AMP-LASSO degrades if we use an inappropriate parameter value as shown in Figure 3. We can see that the proposed ARM with $T = 2, 3$ achieves a similar performance for both $\sigma_{\mathrm{v}}^2 = 0.01$ and $\sigma_{\mathrm{v}}^2 = 0.001$, whereas the performance of AMP-LASSO and the ridge regularization-based method largely depends on the true value of σ_{v}^2.

We then evaluate the estimation performance for a wide range of noise variances. In Figure 5, we plot the estimate $\hat{\sigma}_{\mathrm{v}}^2$ versus its true value σ_{v}^2 when $N = 200$, $\Delta = 0.6$, and $p_0 = 0.9$. The performance is obtained by averaging the result of 100 independent trials. The figure shows that the proposed ARM with $T = 3$ can achieve good estimation performance close to 'ML (oracle)' for the whole range of σ_{v}^2 in the figure. On the other hand, the performance of AMP-LASSO and the ridge regularization-based method degrades for the large σ_{v}^2 and small σ_{v}^2, respectively.

Next, we demonstrate the reconstruction performance of the optimization problem in (4) with the proposed noise variance estimation. Figure 6 shows the CDF of the MSE $\frac{1}{N} \|\hat{x} - x\|_2^2$ ($\hat{x}$: estimate of x) obteind with 1000 independent trials, where $\Delta = 0.7$, $p_0 = 0.8$, and $\sigma_{\mathrm{v}}^2 = 0.001$. The dimension of the unknown vector is set as $N = 100$, 200, and 500 in Figures 6(a), 6(b), and 6(c), respectively. In the figures, 'LASSO with ARM' shows the performance of the optimization problem (4) with the parameter tuning by the proposed ARM.
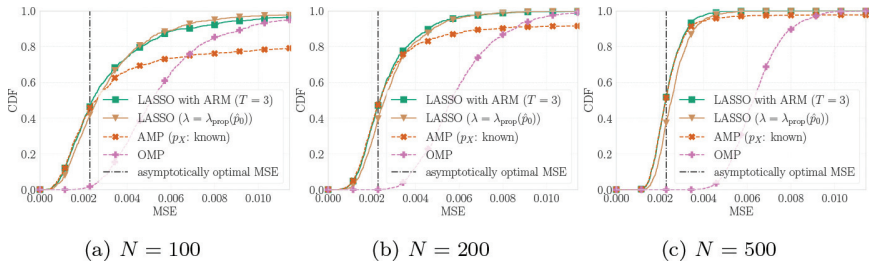
(a) $N = 100$ (b) $N = 200$ (c) $N = 500$

Figure 6: CDF of MSE ($\Delta = 0.7$, $p_0 = 0.8$, $\sigma_v^2 = 0.001$, $p_X(x)$: Bernoulli-Gaussian distribution).

Specifically, we first obtain the estimate of the noise variance $\hat{\sigma}_v^2$ with the proposed ARM, and then calculate the optimal value of λ in terms of asymptotic MSE by using the estimated $\hat{\sigma}_v^2$ and $\hat{p}_0$ via the CGMT framework. For comparison, we also plot the performance of LASSO with the proposed initial regularization parameter $\lambda = \lambda_{\mathrm{prop}}(\hat{p}_0)$ in (15) as 'LASSO ($\lambda = \lambda_{\mathrm{prop}}(\hat{p}_0)$)'. Moreover, we show the performance of the AMP algorithm with the optimal thresholding parameters [49] as 'AMP', for which the distribution of the unknown vector p_X is assumed to be perfectly known. In addition, 'OMP' denotes the performance of the OMP algorithm with the tolerance of 10^{-3}, which is implemented by using the solver of scikit-learn. In the figure, the vertical black line shows the asymptotically optimal MSE, which can be obtained by the CGMT or AMP framework. From the figure, we can see that LASSO outperforms the other methods especially when N is small. On the other hand, the CDF of the AMP algorithm is far from one when $N = 100$ and $N = 200$, which means that the AMP algorithm results in a large MSE or even diverges. This is because the large system limit is usually assumed in the AMP framework to obtain the low-complexity algorithm and the insightful analysis. Since the AMP algorithm achieves similar performance to LASSO when $N = 500$, it would be a suitable candidate for large-scale problems. The performance of the OMP algorithm is worse than the other methods, and hence somehow we need to choose an appropriate tolerance parameter. These results show that the proposed noise variance estimation enables us to obtain good reconstruction performance even when the true noise variance is unknown and the problem size is relatively small.

Figure 7 shows the CDF of the MSE obtained with 1000 independent trials, where $\Delta = 0.7$, $p_0 = 0.8$, and $\sigma_v^2 = 0.005$. We obsereve that the performance of LASSO with $\lambda = \lambda_{\mathrm{prop}}(\hat{p}_0)$ degrades compared to the case with Figure 6. On the other hand, LASSO with ARM can achieve good performance even in this case, which shows the effectiveness of the noise variance estimation for the parameter tuning.
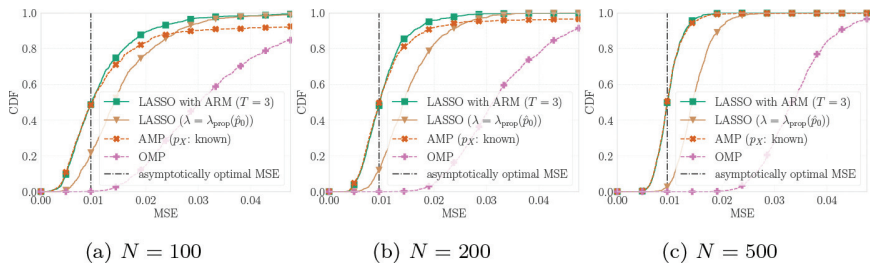
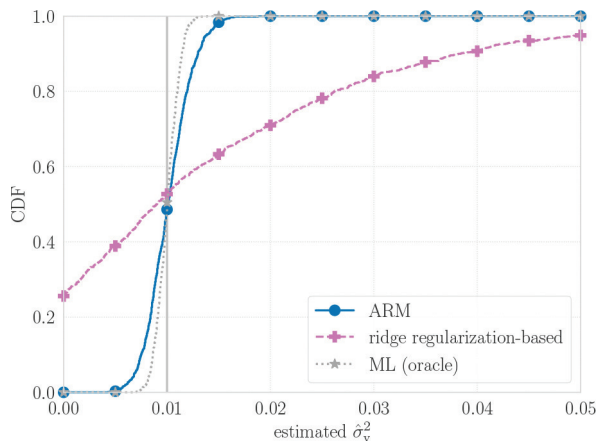Figure 7: CDF of MSE ($\Delta = 0.7$, $p_0 = 0.8$, $\sigma_v^2 = 0.005$, $p_X(x)$: Bernoulli-Gaussian distribution).

Figure 8: CDF of estimated $\hat{\sigma}_v^2$ ($N = 200$, $\Delta = 0.8$, $\sigma_v^2 = 0.01$, $p_{X,b}(x)$: Binary distribution).

7.2 Binary Vector Reconstruction

We then investigate the performance when the unknown vector is a binary vector with the distribution in (17). Figure 8 shows the histogram of the empirical CDF of the estimated $\hat{\sigma}_v^2$ when $N = 200$, $\Delta = 0.8$, and $\sigma_v^2 = 0.01$. In the simulation, we use ADMM [10, 18, 27, 31] to solve the optimization problem (18). Since the estimate by AMP-LASSO in (6) cannot be directly applied to the binary vector reconstruction, we compare the performance of the proposed method with the ridge-regularization based method [62]. As is the case with the Bernoulli-Gaussian distribution in Figure 4, the proposed ARM in Algorithm 2 achieves better performance than the ridge regularization-based method.

Finally, we evaluate the estimation performance versus the true noise variance σ_v^2. Figure 9 shows the performance when $N = 200$ and $\Delta = 0.7$. The performance is obtained by averaging the result of 100 independent trials.
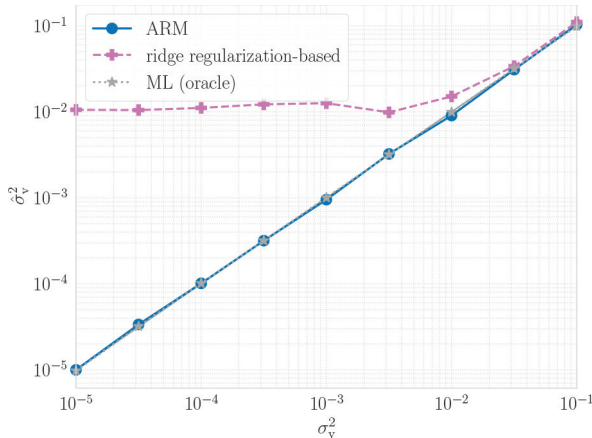
Figure 9: Estimated $\hat{\sigma}_{\mathrm{v}}^2$ versus σ_{v}^2 ($N = 200$, $\Delta = 0.7$, $p_{\mathrm{X,b}}(x)$: Binary distribution).

We observe that the proposed method achieves good estimation performance for a wide range of noise variances as is the case with Figure 5. We thus conclude that the proposed noise variance estimation is effective for the binary distribution $p_{\mathrm{X,b}}(x)$.

8   Conclusion and Future Work

In this paper, we have proposed the noise variance estimation algorithm for compressed sensing with the Gaussian measurement matrix. The proposed ARM algorithm utilizes the asymptotic property of the estimate obtained by the ℓ_1 optimization problem. Specifically, we estimate the noise variance by choosing the value whose corresponding asymptotic residual matches the empirical residual obtained by the actual reconstruction. The main advantages of the proposed approach can be summarized as follows:

- The proposed method can estimate a wide range of noise variances even in underdetermined problems.

- We can design the choice of the regularization parameter λ on the basis of the asymptotic results.

- The proposed idea using the asymptotic residual can be extended for the reconstruction of some non-sparse structured vectors other than sparse ones as shown in Section 6.

- The proposed methods can achieve good performance even when the problem size is relatively small.

Simulation results demonstrate that the proposed method can achieve better estimation performance than some conventional methods. Moreover, by using the estimate of the noise variance, we can choose an appropriate regularization parameter even when the noise variance is unknown. We have shown that the LASSO with the proposed noise variance estimation can achieve better performance than the AMP algorithm for small-scale problems.

Compared to AMP-LASSO in (6) and the scaled residual method in (5), the procedure of the proposed ARM is slightly complicated. For example, we need to estimate the probability p_0 of the unknown vector and solve some scalar optimization problems in the estimation. Although we have focused on the evaluation via computer simulations in this paper for the above reason, the proof of the consistency of the proposed method is important as a theoretical justification. Moreover, it would be an interesting research direction to apply the proposed idea for the choice of the regularization parameter λ to the AMP-based methods. Although we have focused on the compressed sensing problem from the perspective of signal processing in this paper, the application of the proposed approach to statistics would also be a fascinating topic. The extension of the proposed method to the case with unknown distribution p_X is also an important research direction. One possible approach is to iterate the estimation of σ^2 and $\boldsymbol{x}$ until the convergence, where we approximate the distribution p_X with the empirical distribution of the estimated vector $\hat{\boldsymbol{x}}$.

Extensions of the proposed approach to some variants of LASSO (e.g., constrained version) could be an interesting issue. It would be possible to apply the idea of the proposed approach to the case with other structured signals or other optimization problems because CGMT can be used for various optimization problems [43, 65–68]. Since CGMT has also been applied to an optimization problem in the complex-valued domain [1], the extension to the complex-valued case could also be an interesting research direction. The generalization beyond the setting of Assumpsion 1, e.g., partial Fourier measurements and non-i.i.d. measurement matrices [8, 14], would be also beneficial and left as an open problem. Application of conventional approaches as in [8] to binary vector reconstruction and the fair comparison with the proposed method would be important to reveal the advantage of each method.

A   Appendix

A.1   *Proof of Theorem 1*

In this section, we give the proof of Theorem 1. Although the procedure of the proof partly follows some CGMT-based analyses (e.g., [36, 65, 67]), we here show the sketch of the proof to derive the explicit formula in Theorem 1.

A.1.1   CGMT

We firstly summarize CGMT [65, 69] before the proof of Theorem 1. CGMT associates the following primary optimization (PO) and auxiliary optimization (AO).

$$(\text{PO}) : \Phi(\boldsymbol{G}) = \min_{\boldsymbol{w} \in \mathcal{S}_{\text{w}}} \max_{\boldsymbol{u} \in \mathcal{S}_{\text{u}}} \left\{ \boldsymbol{u}^{\top} \boldsymbol{G} \boldsymbol{w} + \xi(\boldsymbol{w}, \boldsymbol{u}) \right\} \tag{A1}$$

$$(\text{AO}) : \phi(\boldsymbol{g}, \boldsymbol{h}) = \min_{\boldsymbol{w} \in \mathcal{S}_{\text{w}}} \max_{\boldsymbol{u} \in \mathcal{S}_{\text{u}}} \left\{ \|\boldsymbol{w}\|_2 \, \boldsymbol{g}^{\top} \boldsymbol{u} - \|\boldsymbol{u}\|_2 \, \boldsymbol{h}^{\top} \boldsymbol{w} + \xi(\boldsymbol{w}, \boldsymbol{u}) \right\} \tag{A2}$$

Here, $\boldsymbol{G} \in \mathbb{R}^{K \times L}$, $\boldsymbol{g} \in \mathbb{R}^K$, and $\boldsymbol{h} \in \mathbb{R}^L$ are composed of i.i.d. standard Gaussian variables. The constraint sets $\mathcal{S}_{\text{w}} \subset \mathbb{R}^L$ and $\mathcal{S}_{\text{u}} \subset \mathbb{R}^K$ are assumed to be closed compact. The function $\xi(\cdot, \cdot)$ is a continuous convex-concave function on $\mathcal{S}_{\text{w}} \times \mathcal{S}_{\text{u}}$.

As in the following theorem, we can relate the optimal costs $\Phi(\boldsymbol{G})$, $\phi(\boldsymbol{g}, \boldsymbol{h})$ and the optimizer $\hat{\boldsymbol{w}}_{\Phi}(\boldsymbol{G})$ of (PO) (For more details, see [65, Theorem 3] and [67, Theorem IV.2]). Intuitively, the theorem enables us to analyze (AO) instead of (AO).

Theorem 2 (CGMT)**.**

1. *For all $\mu \in \mathbb{R}$ and $c > 0$, we have*

$$\Pr(|\Phi(\boldsymbol{G}) - \mu| > c) \leq 2 \Pr(|\phi(\boldsymbol{g}, \boldsymbol{h}) - \mu| \geq c). \tag{A3}$$

2. *Let $\mathcal{S}$ be a open set in $\mathcal{S}_{\text{w}}$ and $\mathcal{S}^{\text{c}} = \mathcal{S}_{\text{w}} \setminus \mathcal{S}$. Moreover, we denote the optimal cost of (AO) with the constraint $\boldsymbol{w} \in \mathcal{S}^{\text{c}}$ by $\phi_{\mathcal{S}^{\text{c}}}(\boldsymbol{g}, \boldsymbol{h})$. If there exists constants $\bar{\phi}$ and η (> 0) such that $\phi(\boldsymbol{g}, \boldsymbol{h}) \leq \bar{\phi} + \eta$ and $\phi_{\mathcal{S}^{\text{c}}}(\boldsymbol{g}, \boldsymbol{h}) \geq \bar{\phi} + 2\eta$ with probability approaching 1 as $L \to \infty$, we then have*

$$\lim_{L \to \infty} \Pr(\hat{\boldsymbol{w}}_{\Phi}(\boldsymbol{G}) \in \mathcal{S}) = 1, \tag{A4}$$

where $L \to \infty$ means that K and L go to infinity with a fixed ratio.

A.1.2   (PO) Problem

To obtain the result of Theorem 1 by using CGMT, we rewrite the ℓ_1 optimization problem (4) as (PO) problem. We firstly define the error vector $\boldsymbol{u} = \boldsymbol{s} - \boldsymbol{x}$ and rewrite (4) as

$$\Phi_N^* := \min_{\boldsymbol{u} \in \mathbb{R}^N} \frac{1}{N} \left\{ \frac{1}{2} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|_2^2 + \lambda f(\boldsymbol{x} + \boldsymbol{u}) \right\}, \tag{A5}$$

where the objective function is normalized by N. From [65, Lemma 5], we can introduce a compact set $\mathcal{S}_{\mathrm{u}} := \{\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leq C_{\mathrm{u}}\}$ with a constant C_{u} (> 0) as

$$\Phi_N^* = \min_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \frac{1}{N} \left\{ \frac{1}{2} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|_2^2 + \lambda f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A6}$$

Since we have

$$\frac{1}{2} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}\|_2^2 = \max_{\boldsymbol{w} \in \mathbb{R}^M} \left\{ \sqrt{N} \boldsymbol{w}^\top (\boldsymbol{A}\boldsymbol{u} - \boldsymbol{v}) - \frac{N}{2} \|\boldsymbol{w}\|_2^2 \right\}, \tag{A7}$$

the optimization problem can be represented as

$$\Phi_N^* = \min_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \max_{\boldsymbol{w} \in \mathbb{R}^M} \left\{ \frac{1}{N} \boldsymbol{w}^\top \left(\sqrt{N} \boldsymbol{A} \right) \boldsymbol{u} - \frac{1}{\sqrt{N}} \boldsymbol{v}^\top \boldsymbol{w} - \frac{1}{2} \|\boldsymbol{w}\|_2^2 + \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A8}$$

Moreover, by using [65, Lemma 6], we can introduce a sufficiently large constraint set $\mathcal{S}_{\mathrm{w}} := \{\boldsymbol{w} \mid \|\boldsymbol{w}\|_2 \leq C_{\mathrm{w}}\}$ $(C_{\mathrm{w}} > 0)$ which will not affect the optimization problem with high probability as

$$\Phi_N^* = \min_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \max_{\boldsymbol{w} \in \mathcal{S}_{\mathrm{w}}} \left\{ \frac{1}{N} \boldsymbol{w}^\top \left(\sqrt{N} \boldsymbol{A} \right) \boldsymbol{u} - \frac{1}{\sqrt{N}} \boldsymbol{v}^\top \boldsymbol{w} - \frac{1}{2} \|\boldsymbol{w}\|_2^2 + \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A9}$$

In the standard analysis based on CGMT, the minimization problem for the error vector $\boldsymbol{u}$ is analyzed. In our proof, however, we analyze the optimal value of $\boldsymbol{w}$ to obtain the result for the residual. We thus exchange the order of min-max from the minimax theorem and change the sign of the objective function to obtain

$$-\Phi_N^* = \min_{\boldsymbol{w} \in \mathcal{S}_{\mathrm{w}}} \max_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \left\{ \frac{1}{N} \boldsymbol{w}^\top \left(\sqrt{N} \boldsymbol{A} \right) \boldsymbol{u} + \frac{1}{\sqrt{N}} \boldsymbol{v}^\top \boldsymbol{w} + \frac{1}{2} \|\boldsymbol{w}\|_2^2 - \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}, \tag{A10}$$

where we can keep the sign of the first term $\frac{1}{N} \boldsymbol{w}^\top (\sqrt{N} \boldsymbol{A}) \boldsymbol{u}$ because the distribution of the matrix $\boldsymbol{A}$ is zero mean Gaussian and sign independent. The optimization problem (A10) is the form of (PO) normalized by N. Note that the optimal value of $\boldsymbol{w}$ can be written as

$$\hat{\boldsymbol{w}}_N^{(\mathrm{PO})} = \frac{1}{\sqrt{N}} \left(\boldsymbol{A} \hat{\boldsymbol{u}}_N^{(\mathrm{PO})} - \boldsymbol{v} \right) \tag{A11}$$

$$= \frac{1}{\sqrt{N}} \left(\boldsymbol{A} \hat{\boldsymbol{x}}(\lambda) - \boldsymbol{y} \right) \tag{A12}$$

from (A7), where $\hat{\boldsymbol{u}}_N^{(\mathrm{PO})} = \hat{\boldsymbol{x}}(\lambda) - \boldsymbol{x}$ is the optimal value of $\boldsymbol{u}$ in (PO).

A.1.3   (AO) Problem

We then analyze the corresponding (AO) problem. Since the procedure is similar to [36], we omit some details in the analysis. The (AO) problem corresponding to (A10) is given by

$$
-\phi_N^* := \min_{\boldsymbol{w} \in \mathcal{S}_{\mathrm{w}}} \max_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \left\{ \frac{1}{N} (\|\boldsymbol{w}\|_2 \, \boldsymbol{g}^\top \boldsymbol{u} - \|\boldsymbol{u}\|_2 \, \boldsymbol{h}^\top \boldsymbol{w}) \right.
$$
$$
\left. + \frac{1}{\sqrt{N}} \boldsymbol{v}^\top \boldsymbol{w} + \frac{1}{2} \|\boldsymbol{w}\|_2^2 - \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A13}
$$

Since the objective function in (A13) is not convex-concave, the order of min-max cannot be exchanged in general. As described in [65, Appendix A], however, we can flip the order in the asymptotic setting because the corresponding (PO) satisfies the condition for the min-max theorem. Hence, we exchange the order of min-max without detailed explanations hereafter. By exchanging the order of min-max and changing the sign of the objective function, we obtain

$$
\phi_N^* = \min_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \max_{\boldsymbol{w} \in \mathcal{S}_{\mathrm{w}}} \left\{ -\frac{1}{N} (\|\boldsymbol{w}\|_2 \, \boldsymbol{g}^\top \boldsymbol{u} - \|\boldsymbol{u}\|_2 \, \boldsymbol{h}^\top \boldsymbol{w}) \right.
$$
$$
\left. - \frac{1}{\sqrt{N}} \boldsymbol{v}^\top \boldsymbol{w} - \frac{1}{2} \|\boldsymbol{w}\|_2^2 + \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A14}
$$

Taking advantage of the fact that both $\boldsymbol{h}$ and $\boldsymbol{v}$ are Gaussian, we can rewrite $\frac{\|\boldsymbol{u}\|_2}{\sqrt{N}} \boldsymbol{h} - \boldsymbol{v}$ as $\sqrt{\frac{\|\boldsymbol{u}\|_2^2}{N} + \sigma_{\mathrm{v}}^2} \, \boldsymbol{h}$, where we use the slight abuse of notation $\boldsymbol{h}$ as i.i.d. standard Gaussian variables. Using this technique, we can set $\|\boldsymbol{w}\|_2 = \beta$ and obtain the equivalent optimization problem

$$
\phi_N^* = \min_{\boldsymbol{u} \in \mathcal{S}_{\mathrm{u}}} \max_{\beta \geq 0} \left\{ \sqrt{\frac{\|\boldsymbol{u}\|_2^2}{N} + \sigma_{\mathrm{v}}^2} \frac{\beta \, \|\boldsymbol{h}\|_2}{\sqrt{N}} - \frac{1}{N} \beta \boldsymbol{g}^\top \boldsymbol{u} - \frac{1}{2} \beta^2 + \frac{\lambda}{N} f(\boldsymbol{x} + \boldsymbol{u}) \right\}. \tag{A15}
$$

To further rewrite the optimization problem (A15), we use the following identity

$$
\chi = \min_{\alpha > 0} \left(\frac{\alpha}{2} + \frac{\chi^2}{2\alpha} \right) \tag{A16}
$$

for $\chi = \sqrt{\frac{\|\boldsymbol{u}\|_2^2}{N} + \sigma_{\mathrm{v}}^2}$ and obtain

$$
\min_{\alpha>0} \max_{\beta\geq 0} \left\{ \frac{\alpha\beta}{2} \frac{\|\boldsymbol{h}\|_2}{\sqrt{N}} + \frac{\sigma_{\mathrm{v}}^2\beta}{2\alpha} \frac{\|\boldsymbol{h}\|_2}{\sqrt{N}} - \frac{1}{2}\beta^2 \right.
$$
$$
\left. - \frac{1}{N} \frac{\alpha\beta \|\boldsymbol{g}\|_2^2}{2} \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2} + \frac{\beta \|\boldsymbol{h}\|_2}{\alpha} \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{n=1}^{N} \min_{u_n \in \mathbb{R}} J_n(u_n) \right\}, \quad (A17)
$$

where we define

$$
J_n(u_n) = \frac{1}{2}\left(u_n - \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}\alpha g_n \right)^2 + \frac{\alpha\lambda}{\beta} \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2} f(x_n + u_n). \quad (A18)
$$

Here, u_n and g_n are the n-th element of $\boldsymbol{u}$ and $\boldsymbol{g}$, respectively. Note that we have exchanged the order of min-max from (A15) with (A16) to (A17) by using the fact that the objective function is convex for $\alpha, \boldsymbol{u}$ and concave for β (For a similar and detailed discussion, see [35, Eq. (57)]). Since we have

$$
\min_{u_n \in \mathbb{R}} J_n(u_n) = \mathrm{env}_{\frac{\alpha\lambda}{\beta} \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}} f\left(x_n + \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}\alpha g_n \right), \quad (A19)
$$

the (AO) problem can be written as

$$
\phi_N^* = \min_{\alpha>0} \max_{\beta\geq 0} F_N(\alpha, \beta), \quad (A20)
$$

where

$$
F_N(\alpha, \beta) = \frac{\alpha\beta}{2} \frac{\|\boldsymbol{h}\|_2}{\sqrt{N}} + \frac{\sigma_{\mathrm{v}}^2\beta}{2\alpha} \frac{\|\boldsymbol{h}\|_2}{\sqrt{N}} - \frac{1}{2}\beta^2 - \frac{1}{N} \frac{\alpha\beta \|\boldsymbol{g}\|_2^2}{2} \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}
$$
$$
+ \frac{\beta \|\boldsymbol{h}\|_2}{\alpha} \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{n=1}^{N} \mathrm{env}_{\frac{\alpha\lambda}{\beta} \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}} f\left(x_n + \frac{\sqrt{N}}{\|\boldsymbol{h}\|_2}\alpha g_n \right). \quad (A21)
$$

We denote the optimal values of α and β in the (AO) problem by α_N^* and β_N^*, respectively.

A.1.4   Applying CGMT

By using the above analysis, we confirm (9) in Theorem 1. As $N \to \infty$, $F_N(\alpha, \beta)$ in (A21) converges pointwise to $F(\alpha, \beta)$ in (8). Letting $\phi^* = F(\alpha^*, \beta^*)$ be the optimal value of $F(\alpha, \beta)$, we can obtain $-\phi_N^* \xrightarrow{\mathrm{P}} -\phi^*$ and $(\alpha_N^*, \beta_N^*) \xrightarrow{\mathrm{P}} (\alpha^*, \beta^*)$ as $N \to \infty$ by a similar approach to the proof of [67,

Lemma IV. 1]. Hence, by setting $\mu = -\phi^*$ in (A3) of Theorem 2, we have $\lim_{N\to\infty} \Pr(|-\Phi_N^* - (-\phi^*)| > c) = 0$ for any $c > 0$, which means (9).

We can also demonstrate the convergence of the residual in (10) from the second statement in Theorem 2. We denote the optimal value of $\boldsymbol{w}$ in (A13) by $\hat{\boldsymbol{w}}_N^{(\mathrm{AO})}$ and define

$$\mathcal{S} = \left\{ \boldsymbol{z} \in \mathbb{R}^M \,\middle|\, \left| \|\boldsymbol{z}\|_2^2 - (\beta^*)^2 \right| < \varepsilon \right\}. \tag{A22}$$

We then have $\hat{\boldsymbol{w}}_N^{(\mathrm{AO})} \in \mathcal{S}$ with probability approaching 1 for any $\varepsilon \ (> 0)$ because $\left\| \hat{\boldsymbol{w}}_N^{(\mathrm{AO})} \right\|_2 = \beta_N^*$ from the definition of β and $\beta_N^* \xrightarrow{\mathrm{P}} \beta^*$. Considering the strong concavity of the objective function in (A15) over β, we can see that there exists $\eta \ (> 0)$ satisfying the condition in Theorem 2 with $\bar{\phi} = -\phi^*$. We thus have $\lim_{N\to\infty} \Pr\left(\hat{\boldsymbol{w}}_N^{(\mathrm{PO})} \in \mathcal{S} \right) = 1$, i.e.,

$$\mathrm{plim}_{N\to\infty} \left\| \hat{\boldsymbol{w}}_N^{(\mathrm{PO})} \right\|_2^2 = (\beta^*)^2. \tag{A23}$$

Combining (A12) and (A23) concludes the proof.

A.2   On Expectation in (8)

In this section, we derive the explicit formula of the expectation in (8) for the Bernoulli-Gaussian distribution in (11). The expectation in (8) can be written as

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{env}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}} f} \left(X + \frac{\alpha}{\sqrt{\Delta}} G \right) \right] \\
&= \frac{\alpha\lambda}{\beta\sqrt{\Delta}} \mathrm{E}\left[f\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}} f} \left(X + \frac{\alpha}{\sqrt{\Delta}} G \right) \right) \right] \\
&\quad + \frac{1}{2} \mathrm{E}\left[\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}} f} \left(X + \frac{\alpha}{\sqrt{\Delta}} G \right) - \left(X + \frac{\alpha}{\sqrt{\Delta}} G \right) \right)^2 \right].
\end{aligned} \tag{A24}$$

Since the proximity operator of $\gamma f(\cdot)$ $(\gamma > 0)$ is given by

$$\mathrm{prox}_{\gamma f}(q) = \mathrm{sign}(q) \max(|q| - \gamma, 0), \tag{A25}$$

the expectation in the first term of (A24) can be further rewritten as

$$
\mathrm{E}\left[f\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(X+\frac{\alpha}{\sqrt{\Delta}}G\right)\right)\right]
$$

$$
= p_0 \int_{-\infty}^{\infty}\left|\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(\frac{\alpha}{\sqrt{\Delta}}g\right)\right| p_G(g)dg
$$

$$
+ (1-p_0)\int_{-\infty}^{\infty}\left|\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}(z)\right| p_Z(z)dz \tag{A26}
$$

$$
= p_0 \cdot \frac{2\alpha}{\sqrt{\Delta}}\int_{\frac{\lambda}{\beta}}^{\infty}\left(g-\frac{\lambda}{\beta}\right)p_G(g)dg
$$

$$
+ (1-p_0)\cdot 2\int_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}}^{\infty}\left(z-\frac{\alpha\lambda}{\beta\sqrt{\Delta}}\right)p_Z(z)dz, \tag{A27}
$$

where $p_Z(z)$ is the PDF of the Gaussian distribution with zero mean and variance $1+\frac{\alpha^2}{\Delta}$. The expectation in the second term of (A24) can also be rewritten as

$$
\mathrm{E}\left[\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(X+\frac{\alpha}{\sqrt{\Delta}}G\right)-\left(X+\frac{\alpha}{\sqrt{\Delta}}G\right)\right)^2\right]
$$

$$
= p_0 \int_{-\infty}^{\infty}\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}\left(\frac{\alpha}{\sqrt{\Delta}}g\right)-\frac{\alpha}{\sqrt{\Delta}}g\right)^2 p_G(g)dg
$$

$$
+ (1-p_0)\int_{-\infty}^{\infty}\left(\mathrm{prox}_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}f}(z)-z\right)^2 p_Z(z)dz \tag{A28}
$$

$$
= p_0\left(\frac{2\alpha^2\lambda^2}{\beta^2\Delta}\int_{\frac{\lambda}{\beta}}^{\infty}p_G(g)dg+\frac{2\alpha^2}{\Delta}\int_0^{\frac{\lambda}{\beta}}g^2 p_G(g)dg\right)
$$

$$
+ (1-p_0)\left(\frac{2\alpha^2\lambda^2}{\beta^2\Delta}\int_{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}}^{\infty}p_Z(z)dz+2\int_0^{\frac{\alpha\lambda}{\beta\sqrt{\Delta}}}z^2 p_Z(z)dz\right). \tag{A29}
$$

We can compute the above integrals by using

$$
\int_a^b p_R(r)dr = P_G\left(\frac{b}{\sigma_r}\right)-P_G\left(\frac{a}{\sigma_r}\right), \tag{A30}
$$

$$
\int_a^b r p_R(r)dr = \sigma_r\left(-p_G\left(\frac{b}{\sigma_r}\right)+p_G\left(\frac{a}{\sigma_r}\right)\right), \tag{A31}
$$

$$
\int_a^b r^2 p_R(r)dr = \sigma_r^2\left(-\frac{b}{\sigma_r}p_G\left(\frac{b}{\sigma_r}\right)+\frac{a}{\sigma_r}p_G\left(\frac{a}{\sigma_r}\right)\right.
$$

$$
\left.+ P_G\left(\frac{b}{\sigma_r}\right)-P_G\left(\frac{a}{\sigma_r}\right)\right), \tag{A32}
$$

where $p_{\mathrm{R}}(r)$ is the PDF of the Gaussian distribution with zero mean and variance σ_{r}^2.

Biography

Ryo Hayakawa received the bachelor's degree in engineering, master's degree in informatics, and Ph.D. degree in informatics from Kyoto University, Kyoto, Japan, in 2015, 2017, and 2020, respectively. He was a Research Fellow (DC1) of the Japan Society for the Promotion of Science (JSPS) from 2017 to 2020, and an Assistant Professor at Osaka Univeristy from 2020 to 2023. From 2023, he is an Associate Professor at Institute of Engineering, Tokyo University of Agriculture and Technology. From 2023, he is an Associate Editor of IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. He received the 33rd Telecom System Technology Student Award, APSIPA ASC 2019 Best Special Session Paper Nomination Award, and the 16th IEEE Kansai Section Student Paper Award. His research interests include signal processing and mathematical optimization. He is a member of IEEE, IEICE, and ISCIE.

References

[1]   E. Abbasi, F. Salehi, and B. Hassibi, "Performance Analysis of Convex Data Detection in MIMO," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, 4554–8.

[2]   A. Aïssa-El-Bey, D. Pastor, S. M. A. Sbaï, and Y. Fadlallah, "Sparsity-Based Recovery of Finite Alphabet Solutions to Underdetermined Linear Systems," *IEEE Transactions on Information Theory*, 61(4), 2015, 2008–18.

[3]   I. B. Atitallah, C. Thrampoulidis, A. Kammoun, T. Y. Al-Naffouri, M. Alouini, and B. Hassibi, "The BOX-LASSO with Application to GSSK Modulation in Massive MIMO Systems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2017, 1082–6.

[4]   F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured Sparsity through Convex Optimization," *Statistical Science*, 27(4), 2012, 450–68.

[5]   M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing," *IEEE Transactions on Information Theory*, 57(2), 2011, 764–85.

[6]   M. Bayati, M. A. Erdogdu, and A. Montanari, "Estimating LASSO Risk and Noise Level," in *Proc. Advances in Neural Information Processing Systems*, 2013, 944–52.

[7]   A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, 2(1), 2009, 183–202.

[8]   P. C. Bellec, "Out-of-Sample Error Estimate for Robust M-estimators with Convex Penalty," March 2023.

[9]   A. Belloni, V. Chernozhukov, and L. Wang, "Square-Root LASSO: Pivotal Recovery of Sparse Signals via Conic Programming," *Biometrika*, 98(4), 2011, 791–806.

[10]   S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, 3(1), 2011, 1–122.

[11]   E. J. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, 52(2), 2006, 489–509.

[12]   E. J. Candès and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, 51(12), 2005, 4203–15.

[13]   E. J. Candès and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, 25(2), 2008, 21–30.

[14]   M. Celentano, A. Montanari, and Y. Wei, "The Lasso with General Gaussian Designs with Applications to Hypothesis Testing," August 2022.

[15]   A. Chockalingam and B. S. Rajan, *Large MIMO Systems*, Cambridge, U.K.: Cambridge University Press, 2014.

[16]   J. W. Choi and B. Shim, "Detection of Large-Scale Wireless Systems via Sparse Error Recovery," *IEEE Transactions on Signal Processing*, 65(22), 2017, 6038–52.

[17]   J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed Sensing for Wireless Communications: Useful Tips and Tricks," *IEEE Communications Surveys Tutorials*, 19(3), thirdquarter 2017, 1527–50.

[18]   P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Vol. 49, *Springer Optimization and Its Applications*, New York, NY: Springer New York, 2011, 185–212, ISBN: 978-1-4419-9569-8.

[19]   P. L. Combettes and V. R. Wajs, "Signal Recovery by Proximal Forward-Backward Splitting," *Multiscale Modeling & Simulation*, 4(4), 2005, 1168–200.

[20]   T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-Complexity near-Optimal Signal Detection in Underdetermined Large-MIMO Systems," in *Proc. National Conference on Communications (NCC)*, February 2012, 1–5.

[21] T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Random-Restart Reactive Tabu Search Algorithm for Detection in Large-MIMO Systems," *IEEE Communications Letters*, 14(12), 2010, 1107–9.

[22] I. Daubechies, M. Defrise, and C. D. Mol, "An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint," *Communications on Pure and Applied Mathematics*, 57(11), 2004, 1413–57.

[23] L. H. Dicker, "Variance Estimation in High-Dimensional Linear Models," *Biometrika*, 101(2), 2014, 269–84.

[24] L. H. Dicker and M. A. Erdogdu, "Maximum Likelihood for Variance Estimation in High-Dimensional Linear Models," in *Proc. Artificial Intelligence and Statistics*, PMLR, May 2016, 159–67.

[25] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, 52(4), 2006, 1289–306.

[26] D. L. Donoho, A. Maleki, and A. Montanari, "Message-Passing Algorithms for Compressed Sensing," *Proceedings of the National Academy of Sciences*, 106(45), 2009, 18914–9.

[27] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators," *Mathematical Programming*, 55(1), 1992, 293–318.

[28] J. Fan, S. Guo, and N. Hao, "Variance Estimation Using Refitted Cross-Validation in Ultrahigh Dimensional Regression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1), 2012, 37–65.

[29] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal of Selected Topics in Signal Processing*, 1(4), 2007, 586–97.

[30] S. M. Fosson and M. Abuabiah, "Recovery of Binary Sparse Signals from Compressed Linear Measurements via Polynomial Optimization," *IEEE Signal Processing Letters*, 26(7), 2019, 1070–4.

[31] D. Gabay and B. Mercier, "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation," *Computers & Mathematics with Applications*, 2(1), 1976, 17–40.

[32] R. Giordano, W. Stephenson, R. Liu, M. Jordan, and T. Broderick, "A Swiss Army Infinitesimal Jackknife," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, April 2019, 1139–47.

[33] R. Hayakawa and K. Hayashi, "Convex Optimization-Based Signal Detection for Massive Overloaded MIMO Systems," *IEEE Transactions on Wireless Communications*, 16(11), 2017, 7080–91.

[34] R. Hayakawa and K. Hayashi, "Reconstruction of Complex Discrete-Valued Vector via Convex Optimization with Sparse Regularizers," *IEEE Access*, 6, 2018, 66499–512.

[35] R. Hayakawa, "Asymptotic Performance Prediction for ADMM-based Compressed Sensing," *IEEE Transactions on Signal Processing*, 70, 2022, 5194–207.

[36] R. Hayakawa and K. Hayashi, "Asymptotic Performance of Discrete-Valued Vector Reconstruction via Box-Constrained Optimization with Sum of ℓ_1 Regularizers," *IEEE Transactions on Signal Processing*, 68, 2020, 4320–35.

[37] R. Hayakawa and K. Hayashi, "Discrete-Valued Vector Reconstruction by Optimization with Sum of Sparse Regularizers," in *Proc. 27th European Signal Processing Conference (EUSIPCO)*, September 2019, 1–5.

[38] K. Hayashi, M. Nagahara, and T. Tanaka, "A User's Guide to Compressed Sensing for Communications Systems," *IEICE Transactions on Communications*, E96-B(3), 2013, 685–712.

[39] L. Janson, R. F. Barber, and E. Candès, "EigenPrism: Inference for High Dimensional Signal-to-Noise Ratios," *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 79(4), 2017, 1037–65.

[40] Y. Kabashima, "A CDMA Multiuser Detection Algorithm on the Basis of Belief Propagation," *Journal of Physics A: Mathematical and General*, 36(43), 2003, 11111–21.

[41] X. Liu, S. Zheng, and X. Feng, "Estimation of Error Variance via Ridge Regression," *Biometrika*, 107(2), 2020, 481–8.

[42] M. E. Lopes, "Estimating Unknown Sparsity in Compressed Sensing," in *Proc. the 30th International Conference on International Conference on Machine Learning*, June 2013, 217–25.

[43] B. Loureiro, C. Gerbelot, H. Cui, S. Goldt, F. Krzakala, M. Mézard, and L. Zdeborová, "Learning Curves of Generic Features Maps for Realistic Datasets with a Teacher-Student Model," *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11), 2022, 114001.

[44] D. G. Luenberger and Y. Ye, "Basic Descent Methods," in *Linear and Nonlinear Programming*, *International Series in Operations Research & Management Science*, New York, NY: Springer US, 2008, 215–62, ISBN: 978-0-387-74503-9.

[45] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magnetic Resonance in Medicine*, 58(6), 2007, 1182–95.

[46] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed Sensing MRI," *IEEE Signal Processing Magazine*, 25(2), 2008, 72–82.

[47] S. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, 41(12), 1993, 3397–415.

[48] L. Miolane and A. Montanari, "The Distribution of the Lasso: Uniform Control over Sparse Balls and Adaptive Parameter Tuning," *The Annals of Statistics*, 49(4), 2021, 2313–35.

[49]   A. Mousavi, A. Maleki, and R. G. Baraniuk, "Consistent Parameter Estimation for LASSO and Approximate Message Passing," *The Annals of Statistics*, 46(1), 2018, 119–48.

[50]   A. Mousavi, A. Maleki, and R. G. Baraniuk, "Parameterless Optimal Approximate Message Passing," *arXiv:1311.0035*, 2013.

[51]   M. Nagahara, "Discrete Signal Reconstruction by Sum of Absolute Values," *IEEE Signal Processing Letters*, 22(10), 2015, 1575–9.

[52]   T. Obuchi and Y. Kabashima, "Cross Validation in LASSO and Its Acceleration," *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5), 2016, 053304.

[53]   S. Oymak and J. A. Tropp, "Universality Laws for Randomized Dimension Reduction, with Applications," *Information and Inference: A Journal of the IMA*, 7(3), 2018, 337–446.

[54]   A. Panahi and B. Hassibi, "A Universal Analysis of Large-Scale Regularized Least Squares Solutions," in *Proc. Advances in Neural Information Processing Systems*, 2017, 3381–90.

[55]   Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition," in *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, November 1993, 40–4.

[56]   F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of machine learning research*, 12, 2011, 2825–30.

[57]   K. R. Rad and A. Maleki, "A Scalable Estimate of the Out-of-sample Prediction Error via Approximate Leave-one-out Cross-validation," *Journal of the Royal Statistical Society Series B*, 82(4), 2020, 965–96.

[58]   S. Reid, R. Tibshirani, and J. Friedman, "A Study of Error Variance Estimation in LASSO Regression," *Statistica Sinica*, 26(1), 2016, 35–67.

[59]   H. Sasahara, K. Hayashi, and M. Nagahara, "Multiuser Detection Based on MAP Estimation with Sum-of-Absolute-Values Relaxation," *IEEE Transactions on Signal Processing*, 65(21), 2017, 5621–34.

[60]   X. Shen, L. Chen, Y. Gu, and H. C. So, "Square-Root Lasso with Nonconvex Regularization: An ADMM Approach," *IEEE Signal Processing Letters*, 23(7), 2016, 934–8.

[61]   W. Stephenson and T. Broderick, "Approximate Cross-Validation in High Dimensions with Guarantees," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, June 2020, 2424–34.

[62]   M. A. Suliman, A. M. Alrashdi, T. Ballal, and T. Y. Al-Naffouri, "SNR Estimation in Linear Systems with Gaussian Matrices," *IEEE Signal Processing Letters*, 24(12), 2017, 1867–71.

[63] T. Sun and C.-H. Zhang, "Scaled Sparse Linear Regression," *Biometrika*, 99(4), 2012, 879–98.

[64] P. H. Tan, L. K. Rasmussen, and T. J. Lim, "Constrained Maximum-Likelihood Detection in CDMA," *IEEE Transactions on Communications*, 49(1), 2001, 142–53.

[65] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Precise Error Analysis of Regularized M-Estimators in High Dimensions," *IEEE Transactions on Information Theory*, 64(8), 2018, 5592–628.

[66] C. Thrampoulidis and W. Xu, "The Performance of Box-Relaxation Decoding in Massive MIMO with Low-Resolution ADCs," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, June 2018, 821–5.

[67] C. Thrampoulidis, W. Xu, and B. Hassibi, "Symbol Error Rate Performance of Box-Relaxation Decoders in Massive MIMO," *IEEE Transactions on Signal Processing*, 66(13), 2018, 3377–92.

[68] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "LASSO with Non-Linear Measurements Is Equivalent to One with Linear Measurements," in *Proc. the 28th International Conference on Neural Information Processing Systems*, Montreal, Canada, 2015, 3420–8.

[69] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized Linear Regression: A Precise Analysis of the Estimation Error," in *Proc. Conference on Learning Theory*, June 2015, 1683–709.

[70] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 1996, 267–88.

[71] J. A. Tropp and A. C. Gilbert, "Signal Recovery from Random Measurements via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, 53(12), 2007, 4655–66.

[72] S. Verdú, *Multiuser Detection*, 1st, New York, NY, USA: Cambridge University Press, 1998, ISBN: 978-0-521-59373-1.

[73] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, 17, 2020, 261–72.

[74] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Communications Surveys Tutorials*, 17(4), Fourthquarter 2015, 1941–88.

[75]   A. Yener, R. D. Yates, and S. Ulukus, "CDMA Multiuser Detection: A Nonlinear Programming Approach," *IEEE Transactions on Communications*, 50(6), 2002, 1016–24.

[76]   G. Yu and J. Bien, "Estimating the Error Variance in a High-Dimensional Linear Model," *Biometrika*, 106(3), 2019, 533–46.