IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 11, 2019, accepted July 12, 2019, date of publication July 22, 2019, date of current version August 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930236

Bayesian Signal Recovery Under Measurement Matrix Uncertainty: Performance Analysis

ABOLFAZL RAZI
School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86001, USA
e-mail: abolfazl.razi@nau.edu

This material is based upon the work supported by the National Science Foundation under Grant No. 1755984. This work is also partially supported by the Arizona Board of Regents (ABOR) under Grant No. 1003329.

ABSTRACT Compressive sensing (CS) has gained a lot of attention in recent years due to its benefits in saving measurement time and storage cost in many applications including biomedical imaging, wireless communications, image reconstruction, remote sensing, and so on. The CS framework enables signal recovery from a small number of linear measurements with an acceptable fidelity taking advantage of signal sparsity in some potentially unknown domain. The core idea of different variants of CS methods is incorporating prior knowledge about the input signal (e.g., prior distribution or sparsity of signals) into the recovery algorithm to restrict the search space and enhance the signal recovery performance. However, the accuracy of signal reconstruction can be significantly compromised if the designed and implemented measurement matrices do not fully match. Often times, the measurement matrix mismatch is treated as an additional noise term in the recovery algorithm ignoring the fact that this mismatch is a learnable quantity which includes random but constant or slow-varying terms during the lifetime of the measurement system. In this paper, we consider this problem for a simple case of Gaussian prior with a sparsity-driven diagonal covariance matrix and find strict bounds on the deviation of the reconstructed signal from the optimal case of fully known measurement matrix based on the properties of the mismatch matrix. The obtained bounds are general, and hence can be used to assess the performance of learning algorithms designed for learning measurement matrix uncertainty and eliminating its effect from the signal recovery. We provide numerical results to illustrate this concept in real-world applications.

INDEX TERMS Compressed sensing, Bayesian methods, uncertainty analysis, signal reconstruction, adaptive estimation, dictionary learning.

I. INTRODUCTION

A linear measurement system is defined by a set of linear equations $\mathbf{y} = H\mathbf{x} + \mathbf{w}$, where $\mathbf{x}$ is the $n \times 1$ input vector, H is the $m \times n$ measurement matrix, and $\mathbf{w}$ is the measurement noise. This basic model and its extensions serve as the underlying framework of countless number of problems in different fields including estimation of dynamic systems [1], control theory [2], Kalman filtering [3], MIMO communication systems [4], channel estimation [5], target tracking and radar systems [6], classification, regression, and clustering models [7], and digital signal processing (e.g. FFT) [8], just to name a few. In most applications, the practical objective is formulated as an *inverse problem* with the goal of recovering an unknown input signal $\mathbf{x}$ with or without some sort of prior knowledge from a set of noisy linear measurements $\mathbf{y}$. When the system is under-determined ($m < n$), the lossless recovery of the input signal is not feasible in general, even if the measurement samples are not contaminated with noise ($\mathbf{w} = \mathbf{0}$). However, the performance of signal recovery can be significantly improved if one has some prior knowledge about the properties of the input signal in addition to the measurement samples $\mathbf{y}$.

An important class of under-determined linear measurement systems is called compressive sensing (CS) [9]. This problem is firstly considered in two seminal works by Candes [10] and Donoho [11] concerning signal recovery from a set of few measurements, where the input signal is sparse when represented in some potentially unknown space. For instance, one may recover a continuous signal with bandwidth BW

The associate editor coordinating the review of this manuscript and approving it for publication was Jun Shi.

from its samples taken at a rate much below the well-known Nyquist rate of 2BW, if the signal is sparse in time or frequency domains [12], [13].

Due to the efficacy of this approach in saving measurement resources as well as its obvious advantages in shrinking data storage size and communication load, this approach has gained a lot of attention from the research community in recent years. In fact, it already has found its way into many applications including image processing [14], [15], magnetic resonance imaging (MRI) [16], tomography [17], secure communication [18], cognitive radio networks [19], array antennas [20], seismic data acquisition [21], object recognition [22], and remote sensing [23].

Theoretically, if the input signal is representable as a k-sparse vector in some space with basis Φ (i.e., $\mathbf{x} = \Phi\mathbf{c}_k$, $|\mathbf{c}|_0 = k$), only $m = 2k$ error-free measurements are sufficient for a full recovery of the input signal, since we have $2k$ unknowns (k nonzero elements and their indexes). For a noisy measurement system $\mathbf{y} = H\mathbf{x} + \mathbf{w}$, it is known that finding the sparsest signal consistent with the measurement equations (i.e., minimize $|\mathbf{x}|_0$ s.t. $\mathbf{x} \in \mathcal{B}(y)$) is a NP-hard and non-convex problem [24]. Here, $\mathcal{B}(\mathbf{y})$ is the search space; for instance, it equals $\{z : |A\mathbf{z} - \mathbf{y}|_2 \leq \epsilon\}$ for an affordable estimation error of ϵ [25].

However, it has been shown that in order to obtain the sparsest solution of the system, one may solve an easiest problem of minimizing norm 1 (i.e., minimize $|\mathbf{x}|_1$ s.t. $\mathbf{x} \in \mathcal{B}(y)$). This problem can be posed as a linear programming for computationally feasible $\mathcal{B}(\mathbf{y}) = \{z : Az = y\}$ and as a convex optimization problem for $\mathcal{B}(y) = \{z : |A\mathbf{z} - \mathbf{y}|_2 \leq \epsilon\}$ [26]. Although this problem does not admit a close-form solution, but can be solved using standard methods developed for convex optimization and linear programming. Several computationally efficient algorithms including *iterative hard thresholding* (IHT) [27], *Dantzig selector* [28], *orthogonal matching pursuit* (OMP) [29], *stagewise* OMP [30], *compressive sampling matching pursuit* (CoSaMP) [31], *approximate message passing* (AMP) [32], and *belief propagation* [33] are proposed in the last decade to solve this problem and its variants with different sparsity models, noise models, and etc.

Sparsity is not the only type of information one may have about the input signal prior to taking measurements. For instance, knowing the prior distribution of the input vector is a very common assumption in classical estimation theory [34]. In *Kalman filtering*, a prior estimate of the signal at the current state is available by applying the state transition equations to the previous state [3]. Note that knowledge about the signal prior distribution is a broader concept since the sparsity of input signals can be enforced using sparsity-driven prior distributions in the context of sparse Bayesian learning (SBL). This approach is the core idea of regression and classification methods developed based on relevance vector machines (RVM) [35], [36]. For instance, a zero-mean Gaussian prior distribution with a diagonal covariance matrix is used for joint recovery of signals with a structured sparsity model in [37]. To impose element-wise and row-wise sparsity of the input matrix, they use SBL with a cost function that favors sparse diagonal covariance matrices. Likewise, a similar approach of penalizing the diagonal elements of the covariance matrix is used for sparse and low-rank matrix reconstruction in [38], and for sparse subspace clustering in [39]. In this paper, we will investigate the derived results for Bayesian signal reconstruction under measurement matrix uncertainty when the sparsity-imposing Gaussian priors are used.

With the recent advances in *dictionary learning* methods [40], finding approximate prior distributions of signals is widely used in a wide range of applications [41]. For instance, in image reconstruction from its compressed linear measurements, dictionary learning can be used to find the prior distribution of vectorized image patches. This method is used to model image tiles with *Gaussian mixture models* (GMM) to enhance the quality of reconstructed image from its compressed linear [42] or nonlinear [15] measurements. This problem is closely related to *statistical compressive sensing* (SCS) paradigm, where the goal is accurate reconstruction of a collection of signals with a shared statistical distribution [43]–[45]. Also, developing *compressive classification* methods to classify data based on its compressed measurements is another emerging use case of extracting information from the compressed representation of signals [42], [46].

A. RELATED WORK ON MATRIX UNCERTAINTY

An important issue that undermines the performance of recovering a signal from its compressed measurements is the uncertainty of the measurement matrix. The uncertainty can be due to the mismatch between the designed and implemented versions of the measurement matrix as well as the perturbations and loss-of-calibration errors in the elements of the measurement matrix. This problem has been recognized by researchers; and several studies have been devoted to alleviating the impacts of this issue for different sparse signal recovery algorithms [47]–[49].

It has been shown that the *Lasso* and *Dantzig selector* algorithms are extremely unstable in recovering sparse signals under the measurement matrix uncertainty, and a new algorithm called *matrix uncertainty selector* is proposed in [50] and [51]. In [52], the deviation of the solution of sparse signal recovery from noisy measurements (i.e., minimize $|\mathbf{x}|_1$ s.t.$|A\mathbf{x} - \mathbf{y}|_2 \leq \epsilon\}$) using *basis pursuit* (BP) algorithm is characterized, when the implemented measurement matrix H includes some perturbations with respect to the designed matrix A (i.e., $H = A + E$ and $\mathbf{y} = H\mathbf{x} + e$). They showed that if the original matrix A satisfies the *restricted isometry property* (RIP) of order k with parameter δ_k, then the perturbed matrix H also satisfies RIP with parameter $(1+\delta_k)\big(1+\epsilon_A\big)^2 - 1$, where $\epsilon_A = |E|_2/|A|_2$ is the relative perturbation. This implies that the distance of the solution of BP algorithm $\mathbf{z}^\star$ for a strictly $k - sparse$ signal $\mathbf{x}$ is bounded as $|\mathbf{z}^\star - \mathbf{x}|_2 \leq C_1(k_A\epsilon_A + \epsilon_y)|\mathbf{y}|_2$, where k_A is the condition number of A, C_1 is a constant, and $\epsilon_y = |e|_2/|y|_2$. In other words, the deviation of the solution from its optimal point is proportional to the condition number of A multiplied by the relative second norm of E to A. This problem is investigated for BP and CoSaMP algorithms in [53] and similar results are obtained. Another closely related problem arises when there is a mismatch between the assumed and the actual basis (dictionary) for which the input signal representation is sparse. It is shown that the norm one error of the best k-term approximation of signal reconstruction $|\mathbf{x} - \mathbf{x}_k|_1$ grows linearly with the signal dimension and the mismatch level between the assumed and actual basis for sparsity [54]. This analysis is extended for signal recovery in compressed sensing framework under the imprecise knowledge about the measurement matrix as well as the signal dictionary D (where the signal representation is sparse $\mathbf{x} = D\mathbf{c}$, $|\mathbf{c}|_0 \ll |\mathbf{x}|_0$) in [55] and necessary conditions for perturbations in A and D are obtained to ensure the robustness of q^{th} order norm minimization. An iterative signal recovery algorithm based on *generalized approximate message passing* (GAMP) is proposed in [47], where the perturbation term $\mathbf{e} = E\mathbf{x}$ in $\mathbf{y} = (A+E)\mathbf{x}+\mathbf{w} = A\mathbf{x}+\mathbf{e}+\mathbf{w}$ is treated as an additional additive noise term. An efficient algorithm based on *cognizant total least-squares* (TLS) is proposed in [56] for *perturbed compressive sampling* in order to jointly recover the signal $\mathbf{x}$ and the measurement perturbation matrix E from the compressed measurements. The problem of characterizing sparse signal recovery with sensing matrix perturbation is studied in [57] and bounds on the estimation error is found based on the Cramér-Rao bound and the Hammersley-Chapman-Robbins bound. The expressions are intricate and hard to interpret. Finally, the impact of measurement matrix perturbation on the performance of OMP is studied in [58], and tighter bounds than those proposed in [52] are derived for the estimation error. Most works consider either bounded perturbations in terms of the second norm of the perturbation matrix, or they treat it as a probabilistic problem, where the elements of the perturbation matrix are i.i.d random variables. A recently paper considers the signal recovery from compressive sensing using the *vector approximate message passing* (VAMP) algorithm when the perturbation of the measurement matrix is more structured [48].

Most of these methods consider point estimators and sparse signal recovery when the measurement matrix is perturbed or is partially known. Unfortunately, the more general problem of signal recovery with known prior distribution (which includes sparsity imposing priors as a special case) using *Bayes* methods is not well investigated when the measurement matrix is perturbed. Perhaps, the most closely related work is [59], where this problem is considered for a practical scenario of target tracking using the direction of arrivals, where the off-grid detection translates to the measurement matrix perturbation. They used a *Bayesian* method to estimate the signal but assumed a special case where only the diagonal elements of the measurement matrix are perturbed according to a uniform distribution. In this work, we consider the signal recovery using *Bayes* method, when the input prior is Gaussian and the measurement matrix $H = A+E$ is partially known (a known part equals to the design matrix A plus an unknown part E). In particular, we characterize the deviation of the recovered signal from the optimal solution in terms of the properties of the perturbation matrix E. This problem is important in the context of CS, as the MAP estimation and other related methods (e.g. RVM [35]) naturally tend to yield sparse signals when a proper sparsity-driven prior distribution is used [37]. Therefore, the obtained results can be used as a reference to evaluate the sensitivity of different sparse signal recovery algorithms to the uncertainty of the measurement matrix.

II. SIGNAL RECOVERY UNDER MEASUREMENT UNCERTAINTY

A. PROBLEM FORMULATION

We consider the following linear measurement framework

$$\begin{aligned} \mathbf{y}_i &= H\mathbf{x}_i + w_i, \quad i = 1, 2, \ldots, N \\ H &= A + E, \end{aligned} \tag{1}$$

where the designed measurement matrix is A but the implemented measurement matrix is H. The mismatch between the designed and implemented measurement matrices are modeled as a mismatch matrix $E = H - A$. This problem may arise in different scenarios. For instance, the measurement matrix can be perturbed over time, or the implementation may not match the design due to quantization errors of fixed-point implementations. Also, one may be able to partially observe the actual measurement matrix embedded in a system. In all cases, the actual measurements are taken by matrix H, but the signal recovery is performed using a slightly different postulated matrix $A = H - E$.

Following similar works (e.g. [52]), we assume that the mismatch matrix E has a bounded second norm and the goal is to characterize the degradation of signal recovery performance in terms of properties of E. We also assume that the input signal $\mathbf{x}$ is *normally* distributed ($\mathbf{x} \sim \mathcal{N}(\mu_x, \Sigma_x)$). Note that if E is fully known, then there is no performance loss; since we can use the actual matrix H in the recovery algorithm. Indeed, we have $I(\mathbf{x}; H\mathbf{x} + \mathbf{w}) = I(\mathbf{x}; A\mathbf{x} + E\mathbf{x} + \mathbf{w}) = I(\mathbf{x}; A\mathbf{x} + \mathbf{w})$.

In a system with multiple measurements, where we collect input-output pairs $\{(\mathbf{x}_i, y_i)\}_{i=1,2,3,\ldots}$ over time, one may develop an algorithm to gradually learn the mismatch matrix E over time and refine it from the newly taken measurements. In a *Bayesian* estimation framework with a Gaussian prior distribution for the input signal and an arbitrary distribution for E, the signal reconstruction is obtained by maximizing the posterior distribution which coincides the mean $\mathbb{E}_{\mathbf{x},E}(\mathbf{x}|\mathbf{y})$. Therefore, the averaged pairwise mean squared errors (MSE), namely $\mathbb{E}_{\mathbf{x}}[|\mathbf{x}_i - \mathbb{E}_{\mathbf{x},E}(\mathbf{x}_i|\mathbf{y}_i)|_2^2]$ lives within the following lower and upper bounds:

$$\mathrm{MSE}_i^{(l)} = \mathbb{E}_{\mathbf{x}}[|\mathbf{x}_i - \mathbb{E}_{\mathbf{x}}(\mathbf{x}_i|\mathbf{y}_i, E))|_2^2], \tag{2}$$

$$\mathrm{MSE}_i^{(u)} = \mathbb{E}_{\mathbf{x}}[|\mathbf{x}_i - \mathbb{E}_{\mathbf{x},E}(\mathbf{x}_i|\mathbf{y}_i))|_2^2], \tag{3}$$

where (2) and (3) respectively correspond to the two extreme cases of the fully known E and the unknown E. Using an optimal recovery algorithm, the average MSE departs from the upper bound to lower bound as we gain more information about E, if a proper learning method is utilized. In the following section, we characterize the signal recovery performance degradation in terms of the gap between the lower and upper bounds of the expected MSE.

B. RESULTS

For the two extreme cases of fully known and unknown E, the signal reconstruction is obtained respectively using the following standard equation for linear *Bayes* estimators [60]:

$$\begin{cases} \hat{\mathbf{x}}_1 = \mu_x + B_H(y - H\mu_x) \\ B_H = \Sigma_x H^T(H\Sigma_x H^T + \Sigma_w)^{-1} \end{cases} \quad \text{for known } E \tag{4}$$

$$\begin{cases} \hat{\mathbf{x}}_2 = \mu_x + B_A(y - A\mu_x) \\ B_A = \Sigma_x A^T(A\Sigma_x A^T + \Sigma_w)^{-1} \end{cases} \quad \text{for unknown } E \tag{5}$$

Before presenting the main results in terms of the gap between the lower and upper bounds, we present a set of lemmas.

Lemma 1: We can write $\mathbb{E}[|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1|_2^2]$ as a function of E as follows:

$$\mathbb{E}[|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1|_2^2] = tr(M\Sigma_y) + \mu_x^T E^T B_A^T B_A E \mu_x \tag{6}$$

where we have $M = (B_A - B_H)^T(B_A - B_H)$.

Proof: To characterize $\mathbb{E}[|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1|_2^2]$, we note

$$\begin{aligned} \hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1 &= B_A(\mathbf{y} - A\mu_x) - B_H(\mathbf{y} - H\mu_x) \\ &= B_E\mathbf{y} + (B_A A - B_H H)\mu_x \end{aligned} \tag{7}$$

where we define $B_E = B_A - B_H$. Therefore, we have

$$\begin{aligned} &E[|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1|_2^2] \\ &= E\Big[\big(B_E\mathbf{y} + (B_A A - B_H H)\mu_x\big)^T\big(B_E\mathbf{y} + (B_A A - B_H H)\mu_x\big)\Big] \\ &= \mathbb{E}[\mathbf{y}^T B_E^T B_E \mathbf{y}] + \mu_x^T(B_H A - B_A A)^T(B_H A - B_A A)\mu_x \\ &\quad + 2\mathbb{E}[\mathbf{y}^T]B_E^T(B_H A - B_A A) \\ &\overset{(a)}{=} \mathbb{E}[\mathbf{y}^T B_E^T B_E \mathbf{y}] + \mu_x^T(B_H A - B_A A)^T(B_H A - B_A A)\mu_x \\ &\quad + 2\mu_x^T H^T B_E^T(B_H A - B_A A)\mu_x \\ &\overset{(b)}{=} \text{tr}[M\Sigma_y] + \mu_x^T\Big[H^T M H + (B_H H - B_A A)^T(B_H H - B_A A) \\ &\quad - 2H^T(B_A - B_H)^T(B_A A - B_H H)\Big]\mu_x, \end{aligned} \tag{8}$$

where we used $\mu_y = \mathbb{E}[\mathbf{y}] = H\mu_x$ in (a) and used the identity $\mathbb{E}[\mathbf{y}^T A\mathbf{y}] = \text{tr}(A\Sigma_y) + \mu_y^T A\mu_y$ in (b). After substituting $H = A + E$ and some manipulations it simplifies to:

$$\mathbb{E}[|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1|_2^2] = \text{tr}(M\Sigma_y) + \mu_x^T E^T B_A^T B_A E \mu_x$$

□

Lemma 2: We have $tr(\Sigma_y) \le tr(\Sigma_w) + 2tr(\Sigma_x)[tr(AA^T) + tr(EE^T)]$.

Proof:

$$\begin{aligned} \text{tr}(\Sigma_y) &= \text{tr}(\Sigma_w + (A+E)\Sigma_x(A+E)^T) \\ &\overset{(a)}{\le} \text{tr}(\Sigma_w) + \text{tr}(\Sigma_x)\text{tr}((A+E)^T(A+E)) \\ &= \text{tr}(\Sigma_w) + \text{tr}(\Sigma_x)[\text{tr}(AA^T) + \text{tr}(EE^T) + 2\text{tr}(AE^T)] \\ &\overset{(b)}{\le} \text{tr}(\Sigma_w) + 2\text{tr}(\Sigma_x)[\text{tr}(AA^T) + \text{tr}(EE^T)], \end{aligned} \tag{9}$$

where (a) is due to the trace inequality $\text{tr}(AB) \le \text{tr}(A)\text{tr}(B)$ and (b) is due to the following identity:

$$\begin{aligned} 0 \le |A - E|_F &= \text{tr}\big((A-E)(A-E)^T\big) \\ &= \text{tr}(AA^T) + \text{tr}(EE^T) - 2\text{tr}(AE^T) \\ &\Rightarrow 2\text{tr}(AE^T) \le \text{tr}(AA^T) + \text{tr}(EE^T). \end{aligned} \tag{10}$$

Here, $|X|_F$ is the frobenius norm of matrix X. □

Lemma 3: For any matrices A, B, C, D with consistent sizes, we have:

$$|AB - CD|_2 \le \min\{|A-C|_2|B|_2 + |C|_2|B-D|_2,\ |A|_2|B-D|_2 + |A-C|_2|D|_2\} \tag{11}$$

Proof: It is an immediate result of applying the norm inequalities ($|AB| \le |A|.|B|$) and ($|A+B| \le |A| + |B|$) as follows:

$$\begin{aligned} |AB - CD|_2 &= |AB - CB + CB - CD|_2 \\ &\le |AB - CB|_2 + |CB - CD|_2 \\ &\le |(A-C)B|_2 + |C(B-D)|_2 \\ &\le |A-C|_2|B|_2 + |C|_2|B-D|_2 \end{aligned}$$

Likewise, we have $|AB - CD|_2 \le |A|_2|B-D|_2 + |A - C|_2|D|_2$, and the result follows. Note that $|.|$ always represents the second norm unless explicitly specified otherwise. □

Lemma 4: A classical result in linear algebra states that [61]

$$\frac{|(X+\delta X)^{-1} - X^{-1}|}{|X^{-1}|} \le \kappa(X)\frac{|\delta X|}{|X|} \tag{12}$$

and noting $\kappa(X) = |X|.|X^{-1}|$ we have:

$$|(X+\delta X)^{-1} - X^{-1}| \le |\delta X||X^{-1}|^2 \tag{13}$$

This inequality is called the Buer-Fike theorem.

Lemma 5: We have

$$|B_A - B_H|_2 \le |\Sigma_x||E||\Sigma_{AW}|\big(1 + 2|A|(|A| + |E|)|\Sigma_{AW}|\big)$$

with $\Sigma_{AW} = (A\Sigma_x A^T + \Sigma_w)^{-1}$ for B_A and B_H defined in (4) and (5).

Proof: We plug in the values of B_H and B_A defined in (4) and (5), and use the norm inequalities ($|AB| \le |A|.|B|$) and ($|A+B| \le |A| + |B|$) to obtain

$$\begin{aligned} &|B_A - B_H| \\ &= \big|\Sigma_x A^T(A\Sigma_x A^T + \Sigma_w)^{-1} \\ &\quad - \Sigma_x H^T(H\Sigma_x H^T + \Sigma_w)^{-1}\big| \\ &\overset{(a)}{\le} |\Sigma_x(A^T - H^T)|\big|(A\Sigma_x A^T + \Sigma_w)^{-1}\big| \end{aligned}$$

$$+|\Sigma_x H^T|\big|(A\Sigma_x A^T+\Sigma_w)^{-1}-(H\Sigma_x H^T+\Sigma_w)^{-1})\big|$$
$$\overset{(b)}{\le} |\Sigma_x||E^T||(A\Sigma_x A^T+\Sigma_w)^{-1}|$$
$$+|\Sigma_x||H^T|.|(A\Sigma_x A^T+\Sigma_w)^{-1}-(H\Sigma_x H^T+\Sigma_w)^{-1}|$$
$$\overset{(c)}{\le} |\Sigma_x||E||(A\Sigma_x A^T+\Sigma_w)^{-1}|$$
$$+|\Sigma_x||H|.|(A\Sigma_x A^T+\Sigma_w)^{-1}-(H\Sigma_x H^T+\Sigma_w)^{-1}|$$
$$\overset{(d)}{\le} |\Sigma_x||E||(A\Sigma_x A^T+\Sigma_w)^{-1}|$$
$$+|\Sigma_x||H|.|H\Sigma_x H^T-A\Sigma_x A^T||(A\Sigma_x A^T+\Sigma_w)^{-1}|^2$$
$$\overset{(e)}{\le} |\Sigma_x||E||(A\Sigma_x A^T+\Sigma_w)^{-1}|$$
$$+2|\Sigma_x||H|.|E||\Sigma_x||A||(A\Sigma_x A^T+\Sigma_w)^{-1}|^2$$
$$\overset{(f)}{\le} |\Sigma_x||E||\Sigma_{AW}|+2|\Sigma_x|^2|H|.|E|.|A||\Sigma_{AW}|^2$$
$$= |\Sigma_x||E||\Sigma_{AW}|\big(1+2|A|(|A|+|E|)|\Sigma_x|.|\Sigma_{AW}|\big), \quad (14)$$

where we used lemma 3 in (a), the norm inequality $|AB|\le|A|.|B|$ in (b), and the identities $E=H-A$ and $|E^T|=|E|$ in (c). Also, the inequality (d) is the application of Buer-Fike theorem (lemma 4) for $X=A\Sigma_x A^T+\Sigma_w$ and $\delta X=H\Sigma_x H^T-A\Sigma_x A^T$. The inequality (e) is due to the following identity:

$$|H\Sigma_x H^T-A\Sigma_x A^T|=|E\Sigma_x A^T+A\Sigma_x E^T|\le 2|E||\Sigma_x||A|. \quad (15)$$

Finally, we plugged in $\Sigma_{AW}=(A\Sigma_x A^T+\Sigma_w)^{-1}$ for notaion convenience in (f). □

Lemma 6: We have

$$\lambda_{max}(M)\le\Big[|\Sigma_x|.|E|.|\Sigma_{AW}|\big(1+2|A|(|A|+|E|)|\Sigma_{AW}|\big)\Big]^2. \quad (16)$$

Proof: Apply lemma 5 noting that $\lambda_{max}(M)=\lambda_{max}[(B_A-B_H)^T(B_A-B_H)]=|B_A-B_H|^2$. □

Now, it is turn to state the main results as follows:

Theorem 7: The performance degradation (in terms of mean squared errors) due to the unknown measurement matrix is bounded as follows:

$$dMSE=MSE(\mathbf{x}_2)-MSE(\mathbf{x}_1)=\mathbb{E}[|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1|^2]\le dMSE^{(u)}, \quad (17)$$

where

$$dMSE^{(u)}=|E|^2\Big[|\Sigma_x|^2.|\Sigma_{AW}|^2\big(1+2|A|(|A|+|E|)|\Sigma_{AW}|\big)^2 \times\Big(tr(\Sigma_w)+2tr(\Sigma_x)\big(tr(AA^T)+tr(EE^T)\big)\Big) +|B_A|^2|\mu_x|^2\Big]. \quad (18)$$

Proof: We follow the following chain of inequalities

$$dMSE = MSE(\mathbf{x})-MSE(\mathbf{x}_1)=\mathbb{E}[|\hat{\mathbf{x}}-\mathbf{x}|^2]-\mathbb{E}[|\hat{\mathbf{x}}_1-\mathbf{x}|^2]$$
$$=\mathbb{E}[|(\hat{\mathbf{x}}-\hat{\mathbf{x}}_1)+(\hat{\mathbf{x}}_1-\mathbf{x})|^2]-\mathbb{E}[|\hat{\mathbf{x}}_1-\mathbf{x}|^2]$$
$$\overset{(a)}{\le}(\mathbb{E}[|(\hat{\mathbf{x}}-\hat{\mathbf{x}}_1)|]+\mathbb{E}[|(\hat{\mathbf{x}}_1-\mathbf{x})|])^2-\mathbb{E}[|\hat{\mathbf{x}}_1-\mathbf{x}|^2]$$
$$\overset{(b)}{=}\mathbb{E}[|(\hat{\mathbf{x}}-\hat{\mathbf{x}}_1)|^2]+\mathbb{E}[|(\hat{\mathbf{x}}_1-\mathbf{x})|^2]-\mathbb{E}[|\hat{\mathbf{x}}_1-\mathbf{x}|^2]$$
$$=\mathbb{E}[|\hat{\mathbf{x}}-\hat{\mathbf{x}}_1|^2]$$
$$\overset{(c)}{=}\mathrm{tr}(M\Sigma_y)+\mu_x^T E^T B_A^T B_A E\mu_x$$
$$\overset{(d)}{\le}\mathrm{tr}(M\Sigma_y)+|E|^2|B_A|^2|\mu_x|^2$$
$$\overset{(e)}{\le}\sum_{i=1}^n\lambda_i(M)\lambda_i(\Sigma_y)+|E|^2.|B_A|^2|\mu_x|^2$$
$$\overset{(f)}{\le}\lambda_{\max}(M)\mathrm{tr}(\Sigma_y)+|E|^2.|B_A|^2|\mu_x|^2, \quad (19)$$

where we used the *triangle inequality* of the second norm, namely $|a+b|_2\le|a|_2+|b|_2\Rightarrow|a+b|_2^2\le(|a|_2+|b|_2)^2$ in (a) and the fact that the estimator is unbiased $\mathbb{E}[|\hat{\mathbf{x}}_1-\mathbf{x}|]=0$ in (b). Also,we used lemma 1 in (c), and the Cauchy-Schwarz inequality in (d), while (e) is due to the inequality $\mathrm{tr}(AB)\le\sum\lambda_i(A)\lambda_i(B)$ for PSD matrices A and B [62], and finally (f) is simply due to $\lambda_i(M)<\lambda_{\max}(M)$.

Now, we substitute the bounds we obtained in lemmas 6 and 2, respectively, for $\lambda_{\max}(M)$ and $\mathrm{tr}(\Sigma_y)$ in (19) to obtain:

$$dMSE\le\lambda_{\max}(M)\mathrm{tr}(\Sigma_y)+|E|_2^2.|B_A|_2^2|\mu_x|_2^2$$
$$\le\Big[|\Sigma_x|.|E|.|\Sigma_{AW}|\big(1+2|A|(|A|+|E|)|\Sigma_{AW}|\big)\Big]^2 \times\Big[\mathrm{tr}(\Sigma_w)+2\mathrm{tr}(\Sigma_x)\big(\mathrm{tr}(AA^T)+\mathrm{tr}(EE^T)\big)\Big] +|E|^2.|B_A|^2|\mu_x|^2$$
$$=|E|^2\Big[|\Sigma_x|^2.|\Sigma_{AW}|^2\big(1+2|A|(|A|+|E|)|\Sigma_{AW}|\big)^2 \times\Big(\mathrm{tr}(\Sigma_w)+2\mathrm{tr}(\Sigma_x)\big(\mathrm{tr}(AA^T)+\mathrm{tr}(EE^T)\big)\Big) +|B_A|^2|\mu_x|^2\Big] \quad (20)$$

□

C. APPLICATION OF THE RESULTS FOR SPARSE SIGNAL RECOVERY

In order to consider sparse signal recovery, we assume that a sparsity-driven Gaussian prior is used where the mean is zero $\mu_x=\mathbf{0}$ and the covariance matrix is diagonal $\Sigma_x=diag(\lambda_1,\lambda_2,\ldots,\lambda_n)$ with elements following power-law decay. More specifically, without loss of generality, we assume that λ_is are in descending order and we have $\lambda_i\le Ci^{-2r}$ for an arbitrary constant C and some $r>1$, which is corresponding to 1/r-compressible signals [63]. In this case, we have $\mathrm{tr}(\Sigma_x)=C\sum_{i=1}^n i^{-2r}$, which is known as p-series. To calculate it, we can simply use the Reimann sums approximation of integral to obtain $\sum_{i=1}^n i^{-2r}=1+\sum_{i=2}^n i^{-2r}<=1+\int_{i=1}^n x^{-2r}=1+\frac{1-(1/n)^{2r-1}}{2r-1}=\frac{2r-(1/n)^{2r-1}}{2r-1}\le\frac{2r}{2r-1}$, which implies $\mathrm{tr}(\Sigma_x)\le 2rC/(2r-1)$, as shown in Fig. 1. Likewise, we have $|\Sigma_x|=\prod_{i=1}^n\lambda_i\le C^n(n!)^{-2r}$. Therefore, the obtained bound in (20) reduces to:

$$dMSE\le|E|^2\Big[\lambda_{\max}^{2s}|\Sigma_{AW}|^2\big(1+2|A|(|A|+|E|)|\Sigma_{AW}|\big)^2 \times\Big(\mathrm{tr}(\Sigma_w)+2s\lambda_{\max}\big(\mathrm{tr}(AA^T)+\mathrm{tr}(EE^T)\big)\Big)\Big] \quad (21)$$

for sparse signal recovery.

D. INTERPRETATION OF THE RESULTS

We observe that the performance degradation is bounded by $|E|^2[k_1(k_2+|E|)^2(k_3+\mathrm{tr}(EE^T))+k_4]$, where k_is are

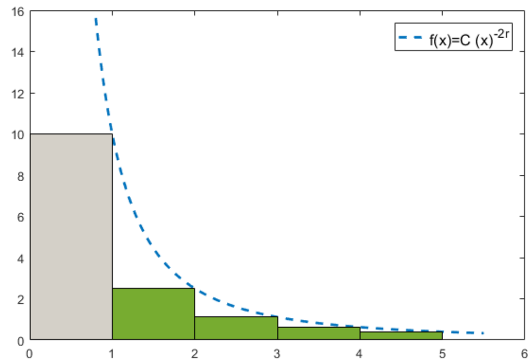


FIGURE 1. Reimann sum approximation of $f(x) = C(x)^{-2r}$ used to approximate $\mathrm{tr}(\Sigma_x)$. Here, we use $C = 10$ and $r = 1$.

independent of E. For small perturbations $|E| \to 0$, the dominant term is $\lambda_{max}(EE^T) = |E|^2$, so the upper bound on the performance degradation is directly proportional to $|E|^2$, which is consistent with the results reported in [52] for BP algorithm. Another implication of (19) is that the performance loss increases with the input signal energy (both $|\mu_x|^2$ and Σ_x), which reflects the fact that the term $E\mathbf{x}$ acts as an additional noise term with mean $E\mu_x$ and covariance $E\Sigma_x E^T$ for the unknown mismatch matrix E. The upper bound vanishes to zero when the signal energy approaches zero (i.e., $|\mu_x|, |\Sigma_x| \to 0$) which confirms the tightness of the derived upper bound. The interesting and somewhat unexpected fact is that performance degradation also increases with the measurement noise SNR $\mathbb{E}[|x|^2]/\Sigma_w|$, as shown in numerical results. This is perhaps due to the term $|\Sigma_{AW}| = |(A\Sigma_x A^T + \Sigma_w)^{-1}|$, which increases with the reduction in noise power Σ_w, as characterized in lemma 6.

E. CONNECTIONS TO MEASUREMENT MATRIX PERTURBATIONS

In this work, we assumed that the mismatch matrix E is an unknown but fixed matrix to model our uncertainty about the actual measurement matrix $H = A + E$. This problem is closely related to a different class of problems where the elements of the measurement matrix are subject to perturbation during each measurement cycle. In this case, the difference between the designed and actual matrices, $E = [E_{ij}]_{m\times n}$ is a random matrix that can differ from one measurement to another. One special case is when the columns of E (i.e., $E_1, E_2, \ldots, E_n$) are zero-mean random vectors with covariance matrix Σ_E. In this case, the columns of E can be considered data samples drawn from a population with the same distribution, and hence $\frac{1}{n}EE^T$ represents the sample covariance matrix of E_i. This matrix approaches the population covariance matrix Σ_E for $n/m \to \infty$. If E_is are zero mean Gaussian distributed random variables (i.e., $E_i \sim \mathcal{N}(\mathbf{0}, \Sigma_E)$), then $EE^T \sim \mathcal{W}_m(n, \Sigma_E)$, where $\mathcal{W}_m()$ represents the Wishart distribution. The joint distribution of the eigenvalues of this matrix is defined but are computationally hard to evaluate [64]. If we have $\Sigma_E = I_m$ and the dimensions of E grow sufficiently large $m, n \to \infty$ while maintaining a constant compression rate $\gamma = m/n$, then the joint distribution of eigenvalues follow the Marchenko-Pastur law. Furthermore, the largest and smallest eigenvalues of EE^T, namely $\hat{\lambda}_1 = \lambda_{\max}(EE^T)$, $\hat{\lambda}_m = \lambda_{\min}(EE^T)$ respectively converge (almost surely) to the right an left support points of the Marcenko-Pastur distribution defined as $b_+ = (1 + \sqrt{\lambda})^2$ and $b_- = (1 - \sqrt{\lambda})^2$ [65]. This means that for the case of the elements of E being i.i.d zero mean Gaussian distributed (i.e., $E_{ij} \sim \mathcal{N}(0, \sigma_2^E)$), we have $\hat{\lambda}_1 \xrightarrow{a.s.} m\sigma_E^2(1 + m/n)^2$. In other words, we have $|E|^2 < m\sigma_E^2(1 + m/n)^2$ and $\mathrm{tr}(EE^T) < m^2\sigma_E^2(1 + m/n)^2$ almost surely. Therefore, the derived equations are practically useful for this case as well. In other words, the results are practically useful for both cases of deterministic and probabilistic mismatch matrix E. The latter case models time-variant perturbations and loss-of-calibration errors.

III. NUMERICAL RESULTS

In this section, we illustrate the obtained bounds on the performance degradation of signal recovery from its compressed measurements when the measurement matrix is partially known to the decoder. Furthermore, we present numerical results to investigate the performance of signal reconstruction when the measurement matrix in the compression and reconstruction stages do not fully match.

In the presented simulations, we use the following parameters unless otherwise specified. $A = [A_{ij}]_{m\times n}$ is the measurement matrix whose elements are driven from zero-mean Gaussian distribution ($A_{ij} \sim \mathcal{N}(0, \sigma_A^2)$). For calculation convenience, we normalize the rows of A to 1 (equivalently we set $\sigma_A^2 = 1/n$). Similarly, we generate mismatch matrix $E = [E_{ij}]_{m\times n}$ using $E_{ij} \sim \mathcal{N}(0, \sigma_E^2)$, and then we obtain the actual measurement matrix as $H = A + E$. Note that we keep A and E fixed for $N = 10,000$ samples when obtaining numerical results. Then, to obtain smoother curves and avoid bias to realizations of A and E, we average the results over 1000 iterations with different realizations of A and E (external loop). The input vectors follow Gaussian distribution with mean $\mu_x \mathbf{1}_n$ and covariance matrix $\Sigma_x = \sigma_x^2\, I_n$ and likewise the measurement noise is $w_i \sim \mathcal{N}(0, \Sigma_w = \sigma_w^2\, I_m)$. We define the measurement signal to noise ratio as $SNR = (\mu_x^2 + \sigma_x^2)/\sigma_w^2$ (noting that A is row-normalized). Likewise, we define the mismatch SNR as $mmSNR = |A|/|E| = \lambda_{\max}(AA^T)/\lambda_{\max}(EE^T)$.

Fig. 2 depicts the obtained upper bound on the performance degradation of signal recovery in terms of MSE found in theorem 7 equation (18) with respect to the measurement matrix perturbation SNR (mmSNR) for an illustrative scenario with parameters $m = 8, n = 4$. Fig. 2a shows that the performance degradation decreases with mmSNR as expected. Also the performance gap surprisingly increases with the measurement noise SNR due to the term $|\Sigma_{AW}| = |(A\Sigma_x A^T + \Sigma_w)^{-1}|$ in (18). The performance loss increases almost by a factor of 10 for each 3 dB increment in the SNR. This means that the measurement matrix mismatch is more harmful at higher

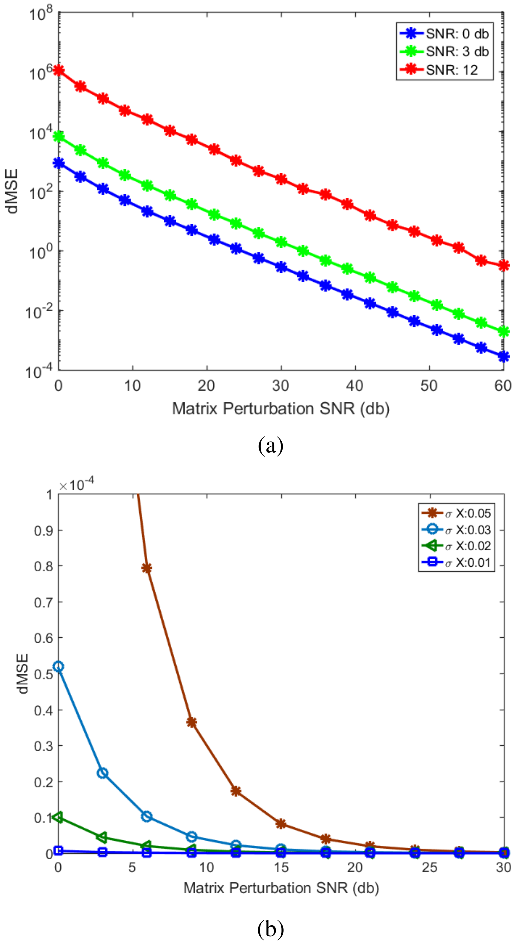


FIGURE 2. Presentation of upper bound on the signal recovery performance degradation (*dMSE*$^{(u)}$) versus mismatch SNR (*mmSNR*) averaged over 1000 realizations of *A* and *E* with parameters $m = 4$ and $n = 8$ for (a): Different SNR values, and (b): Different signal power (σ_X).

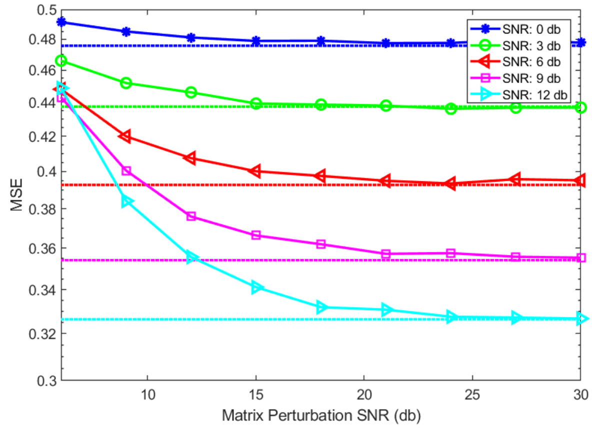


FIGURE 3. Performance loss in signal recovery due to measurement matrix mismatch. In this experiment, we use $N = 10,000$ randomly generated input vectors averaged over 1000 iterations with different realizations of *A* and *E*. The solid and dashed lines, respectively, represent the unknown and fully known mismatch matrix *E*.

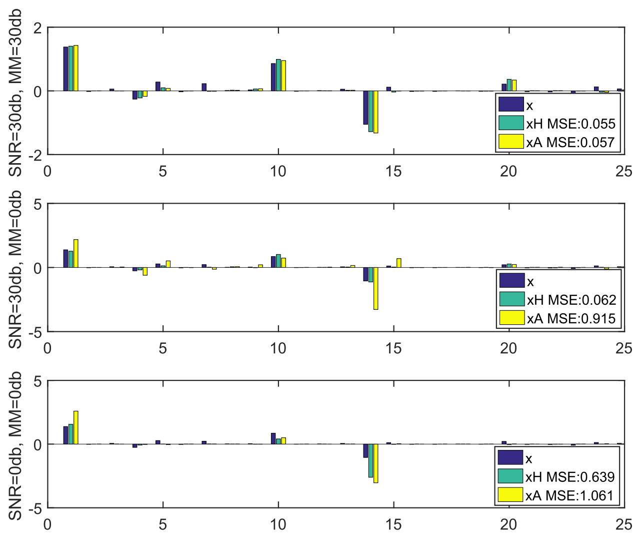
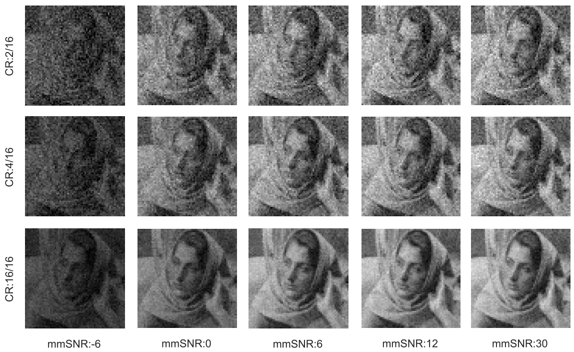


FIGURE 4. Performance loss due to the measurement matrix mismatch for sparse signal recovery for three scenarios. *xA* and *xH* respectively represented the reconstructed signal using *A* and *H*. The presented MSE errors are calculated by averaging over $N = 1,000$ randomly generated signals. One illustrative signal for each test scenario is presented.

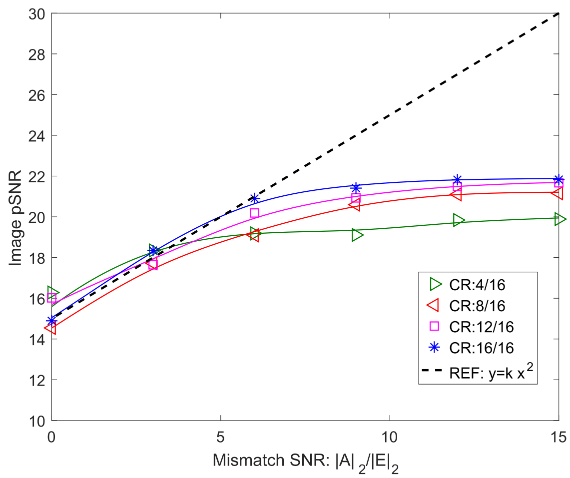
SNR values when the accurate signal recovery is possible for a fully known measurement matrix. Fig. 2b represents another interesting fact that the performance loss in the signal recovery increases with the power of input signal since $E\mathbf{x}$ acts as an additional noise term, as discussed in section II-D. Similar results are presented in Fig. 3 to more directly show the performance loss due to measurement matrix mismatch for different measurement SNR values.

Fig. 4 shows the performance of sparse signal recovery when a proper Gaussian distribution is used. Here we simulate 32-point time signals, where the signal vectors are derived from a multivariate zero-mean Gaussian distribution with diagonal covariance matrix whose elements follow power-law decay $\lambda_i = Ci^{-2r}$ with parameter $r = 1.5$. The compression rate is $CC = m/n = 8/32$. It is noticeable from the top figure that when both noise and mismatch SNRs are high, the algorithm recovers the sparse time signal with high accuracy from fewer linear measurements. The middle figure shows a scenario that the mismatch SNR is relatively low ($mmSNR = |A|/|E| = 0$) and hence the signal recovery using the postulated measurement matrix A does not perform well and the relative mean squared errors (MSE) increases from 0.062 to 0.915 for not using the actual measurement matrix H in the recovery step. The bottom figure illustrates a scenario when both additive and mismatch SNRs are relatively low. Even in this case, the recovery performance shows a significant drop for measurement matrix mismatch.

In order to investigate the performance loss in real scenarios, we provide numerical results in Fig. 5 using randomly-generated Gaussian-distributed input vectors. The results suggest that the signal recovery with an unknown mismatch matrix (solid line) departs from the ideal case of fully known mismatch matrix (dashed line), when the mismatch SNR goes below 20 dB. The performance gap is larger for higher SNR values since a more accurate signal



(a)



(b)

FIGURE 5. Image reconstruction from its compressed measurements using Bayesian inference and GMM prior. (a): Recovered image with different compression rates and mismatch SNR values. (b): pSNR of reconstructed image versus mismatch SNR.

reconstruction is feasible in the absence of measurement matrix mismatch.

Furthermore, we investigate the results for a practical case of reconstructing an image from its compressed version using Bayesian inference. In this simulation, the input images are grayscaled and then split into small tiles of 4×4 pixels. Then, the 4×4 image patches are vectorized (converted to a 16×1 input vector $\mathbf{x}_i$). In order to apply Bayesian inference method, we train a GMM model for the prior distribution of input signals using a large photo dataset, following our previous work in [15]. We take compressed measurements of image patches using $\mathbf{y}_i = (A + E)\mathbf{x}_i + \mathbf{w}_i$ in (1) and then reconstruct the image from its compressed version once with known and next with unknown E. In other words, once we used both the actual measurement matrix $H = A + E$ and the presumed measurement matrix A in the signal recovery stage. The results are shown in Fig. 2. The top figure (5a) presents the reconstructed images for different compression rates of $CR = m/n = 2/16,\ 4/16,\ 16/16$ and different mismatch SNR values $|A|/|E| = -6,\ 0, 6, 12,\ 30\ dB$. It is seen that the quality of reconstructed images is visually improved when taking more measurements ($m/n \to 1$) and increasing the mismatch SNR. To quantify the results, we also plot the obtained peak SNR (pSNR), as a popular image quality metric, versus the mismatch SNR in Fig. 5b. The results interestingly show that at lower mismatch SNR values, the performance loss in terms of pSNR is almost linearly proportional to $|E|^2/|A|^2$.

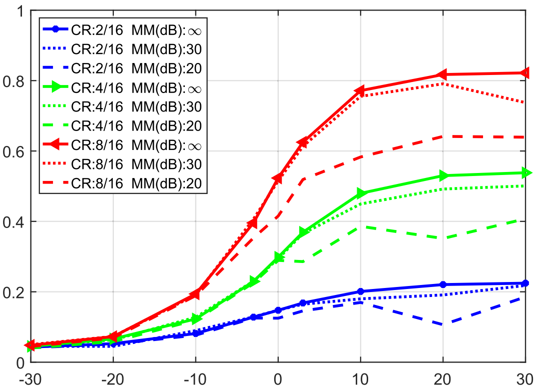


FIGURE 6. Bayesian classification of Letter dataset using compressed sampling under measurement matrix uncertainty. Here, $CR = m/n$ is the compression rate and MM is the mismatch SNR in terms of $|E|_2/|H|_2$ in dB scale.

Similar results are presented for Bayes classifier under measurement matrix mismatch in Fig. 6. Here, we use the method of classification by Bayesian inference following [66] and [15]. To this end, we train a different GMM distribution $f_c(\mathbf{x})$ for data samples $\mathbf{x}$ within each class $c \in \{1, 2, \ldots, C\}$, then reconstruct the test signal from its compressed measurements $\mathbf{y} = H\mathbf{x} + \mathbf{w}$ using Bayesian inference for each class ($\mathbf{x}_c = E[\mathbf{x}|\mathbf{y}, c]$), and finally perform classification by choosing the class under which the obtained signal posterior probability is maximized (i.e. $c^* = \max_c p(\mathbf{x}_c|\mathbf{y})$ or equivalently $c^* = \max_c p(c|\mathbf{y})$). Here, we use the postulated measurement matrix $A = H - E$ in the inference step while calculating $c^* = \max_c p(\mathbf{x}_c|\mathbf{y})$). We apply this method to letter dataset [67] and train a GMM distribution with 10 components for data samples represented by 16×1 feature vectors within each of the $C = 26$ classes. We use compression rate of $CR = m/n = 1/8, 1/4, 1/2$. The results in Fig. 6 show that the classification accuracy decreases with lower measurement mismatch SNR, as we observed in the signal reconstruction application. The performance degradation is more noticeable at higher values of the additive noise SNR. The classification success rate drop can be as high as 10% for a mismatch SNR of $mmSNR = 10\log_1 0(|E|_2/|H|_2) = 20\ dB$.

IV. CONCLUSION

In this paper, we considered signal recovery from its compressed version using Bayes inference when the measurement matrix $H = A + E$ is partially unknown (A is the known part and E is the unknown part). If E is priorly unknown but fixed, one may implement advanced learning methods to gradually learn E during the lifetime of the measurement device and eliminate its effects from the signal recovery stage. In order to quantify the performance loss due to the uncertainty in the measurement matrix (equivalently the gain one may obtain by learning the mismatch matrix E), we obtained an upper bound on the performance loss of the signal reconstruction in terms of MSE for signals with Gaussian prior based on the spectral properties of E. The results show that the upper bound grows linearly with $|E|^2$ for small perturbations. The results are somewhat consistent with those of former works in the context of compressive sensing, including the sparse signal reconstruction using BP algorithm. Also, the performance loss grows with the input signal energy under a given SNR value since $E\mathbf{x}$ acts as an additional noise term. If the measurement matrix experiences time-varying random perturbations, the results can be used as technical guarantees (almost surely) on the performance loss. A potential future extension of this work is extending the derived upper bound to a more general case of arbitrary prior distributions.

REFERENCES

[1] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. London, U.K.: Chapman & Hall, 2011.

[2] T. Kailath, *Linear Systems*, vol. 156. Englewood Cliffs, NJ, USA: Prentice-Hall, 1980.

[3] M. S. Grewal, *Kalman Filtering*. Berlin, Germany: Springer, 2011.

[4] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[5] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 2, pp. 18–48, 2nd Quart., 2007.

[6] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell, *Bayesian Multiple Target Tracking*. Norwood, MA, USA: Artech House, 2013.

[7] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[8] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1975, p. 777.

[9] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, vol. 1, no. 3. Basel, Switzerland: Birkhäuser, 2013.

[10] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, vol. 3, 2006, pp. 1433–1452.

[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[12] B. Adcock and A. C. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Found. Comput. Math.*, vol. 16, no. 5, pp. 1263–1323, Oct. 2016.

[13] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[14] G. Oliveri, M. Salucci, N. Anselmi, and A. Massa, "Compressive sensing as applied to inverse problems for imaging: Theory, applications, current trends, and open challenges," *IEEE Antennas Propag. Mag.*, vol. 59, no. 5, pp. 34–46, Oct. 2017.

[15] L. Wang, A. Razi, M. Rodrigues, R. Calderbank, and L. Carin, "Nonlinear information-theoretic compressive measurement design," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1161–1169.

[16] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, 2016.

[17] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, p. 2791, 2009.

[18] Y. Zhang, Y. Xiang, L. Y. Zhang, Y. Rong, and S. Guo, "Secure wireless communications based on compressive sensing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1093–1111, 2nd Quart., 2018.

[19] S. K. Sharma, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Application of compressive sensing in cognitive radio communications: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1838–1860, 3rd Quart., 2016.

[20] A. Massa, M. Bertolli, G. Gottardi, A. Hannan, D. Marcantonio, G. Oliveri, A. Polo, F. Robol, P. Rocca, and F. Viani, "Compressive sensing as applied to antenna arrays: Synthesis, diagnosis, and processing," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[21] R. G. Baraniuk and P. Steeghs, "Compressive sensing: A new approach to seismic data acquisition," *Lead. Edge*, vol. 36, no. 8, pp. 642–645, 2017.

[22] A. C. Fannjiang, T. Strohmer, and P. Yan, "Compressed remote sensing of sparse objects," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 595–618, 2010.

[23] L. Wang, K. Lu, and P. Liu, "Compressed sensing of a remote sensing image based on the priors of the reference image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 736–740, Apr. 2015.

[24] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of L_p minimization," *Math. Program.*, vol. 129, no. 2, pp. 285–299, 2011.

[25] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," *Preprint*, vol. 93, no. 1, p. 2, 2011.

[26] L. Kämmerer, S. Kunis, and D. Potts, "Interpolation lattices for hyperbolic cross trigonometric polynomials," *J. Complex.*, vol. 28, no. 1, pp. 76–92, 2012.

[27] R. E. Carrillo, L. F. Polanía, and K. E. Barner, "Iterative hard thresholding for compressed sensing with partially known support," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4028–4031.

[28] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2312–2351, 2007.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[30] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*. vol. 58, no. 2, pp. 1094–1121, Feb. 2006.

[31] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[32] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[33] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[34] N. E. Nahi, *Estimation Theory and Applications*. New York, NY, USA: Wiley, 1969.

[35] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[36] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[37] W. Chen, D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, "Simultaneous Bayesian sparse approximation with structured sparse models," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6145–6159, Dec. 2016.

[38] W. Chen, "Simultaneously sparse and low-Rank matrix reconstruction via nonconvex and nonseparable regularization," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5313–5323, Oct. 2018.

[39] B. Xin, Y. Wang, W. Gao, and D. Wipf, "Building invariances into sparse subspace clustering," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 449–462, Jan. 2017.

[40] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[41] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, no. 2, pp. 349–396, 2003.

[42] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[43] G. Yu and G. Sapiro, "Statistical compressive sensing of Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 3728–3731.

[44] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of Gaussian mixture models," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 585–600, Feb. 2013.

[45] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, Apr. 2013.

[46] H. Reboredo, F. Renna, R. Calderbank, and M. R. D. Rodrigues, "Bounds on the number of measurements for reliable compressive classification," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5778–5793, Nov. 2016.
[47] J. T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 804–808.
[48] J. Zhu, Q. Zhang, X. Meng, and Z. Xu, "Vector approximate message passing algorithm for structured perturbed sensing matrix," 2018, *arXiv:1808.08579*. [Online]. Available: https://arxiv.org/abs/1808.08579
[49] R. Ayanzadeh, S. Mousavi, M. Halem, and T. Finin, "Quantum annealing based binary compressive sensing with matrix uncertainty," 2019, *arXiv:1901.00088*. [Online]. Available: https://arxiv.org/abs/1901.00088
[50] M. Rosenbaum and A. B. Tsybakov, "Sparse recovery under matrix uncertainty," *Ann. Statist.*, vol. 38, no. 5, pp. 2620–2651, 2010.
[51] M. Rosenbaum and A. B. Tsybakov, "Improved matrix uncertainty selector," in *From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner*. Institute of Mathematical Statistics Collections, 2013, pp. 276–290. [Online]. Available: https://projecteuclid.org/euclid.imsc/1362751194 and https://www.imstat.org/contact-ims/
[52] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.
[53] M. A. Herman and D. Needell, "Mixed operators in compressed sensing," in *Proc. 44th Annu. Conf. Inf. Sci. Syst. (CISS)*, 2010, pp. 1–6.
[54] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.
[55] A. Aldroubi, X. Chen, and A. M. Powell, "Perturbations of measurement matrices and dictionaries in compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 33, no. 2, pp. 282–291, 2012.
[56] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
[57] Y. Tang, L. Chen, and Y. Gu, "On the performance bound of sparse estimation with sensing matrix perturbation," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4372–4386, Sep. 2013.
[58] J. Ding, L. Chen, and Y. Gu, "Perturbation analysis of orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 398–410, Jan. 2013.
[59] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.
[60] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.
[61] L. El Ghaoui, "Inversion error, condition number, and approximate inverses of uncertain matrices," *Linear Algebra Appl.*, vol. 343, pp. 171–193, Mar. 2002.
[62] F. Rezakhanlou, "Lectures on random matrices," Dept. Math., UC Berkeley, Berkeley, CA, USA, Aug. 2012.
[63] V. Cevher, "Learning with compressible priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 261–269.
[64] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, no. 2, pp. 475–501, 1964.
[65] A. J. Izenman, "Introduction to random-matrix theory," Tech. Rep., 2008. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C3&q=Izenman+Introduction+to+random-matrix+theory+2008&btnG= and https://astro.temple.edu/~alan/MMST/IntroRMT.PDF
[66] M. Chen, W. R. Carson, M. R. D. Rodrigues, L. Carin, and A. R. Calderbank, "Communications inspired linear discriminant analysis," in *Proc. ICML*, 2012, pp. 1–8.
[67] D. J. Slate. (1991). *Letter Recognition Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/letter+recognition

ABOLFAZL RAZI received the B.S. degree from Sharif University, in 1998, the M.S. degree from Tehran Polytechnic, in 1998, and the Ph.D. degree from the University of Maine, in 2001, all in electrical engineering. He held two postdoctoral positions in machine learning and predictive modeling with Duke University from 2013 to 2014 and Case Western Reserve University from 2014 to 2015. He is currently an Assistant Professor of electrical engineering with Northern Arizona University (NAU). His current research interests include AI-enabled wireless networking, nonlinear compressive sensing, topology control, UAV networks, and predictive modeling. He was a recipient of several competitive awards, including the Best Research of MCI in 2008, the Best Graduate Research Assistant of the Year Award from the College of Engineering, University of Maine, in 2011, and the Best Paper Award from the IEEE/CANEUS Fly by Wireless Workshop in 2011.