ISIT'2013 1569719853

Distortion-Based Achievability Conditions for Joint Estimation of Sparse Signals and Measurement Parameters from Undersampled Acquisitions

Mehmet Akçakaya
Beth Israel Deaconess Medical Center
Harvard Medical School
Boston, USA
Email: akcakaya@fas.harvard.edu

Vahid Tarokh
School of Engineering and Applied Sciences
Harvard University
Cambridge, USA
Email: vahid@ses.harvard.edu

***Abstract*—In this paper, we consider an undersampling system model of the form $\mathbf{y} = \mathbf{A}(\mathcal{T}(\mathbf{x}, \boldsymbol{\theta})) + \mathbf{n}$, where $\mathbf{x}$ is a k-sparse signal, $\mathcal{T}(\cdot, \cdot)$ is a (possibly non-linear) function specified by a parameter vector $\boldsymbol{\theta}$ and acting on $\mathbf{x}$, $\mathbf{A}$ is a sensing matrix, and $\mathbf{n}$ is additive measurement noise. We consider an information theoretic decoder that aims to recover the sparse signal and the transformation parameter vector jointly, and study the achievability conditions for estimating the underlying signal within a specified ℓ_2 distortion for Gaussian sensing matrices. We compare the achievable distortion of the joint estimation process to that of the standard noisy compressed sensing model, where the sparse signal is directly measured with a sensing matrix with the same number of measurements. We also provide a numerical example to illustrate potential applications.**

I. Introduction

Let $\mathbf{x} \in \mathbb{C}^n$ be a sparse signal, such that the number of non-zero coefficients of $\mathbf{x}$, denoted $||\mathbf{x}||_0$, is $k << n$. Compressed (or compressive) sensing considers the problem of reconstructing a sparse signal from $m < n$ measurements acquired as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a sensing matrix and $\mathbf{n} \in \mathbf{C}^m$ is measurement noise. It has been shown that if $\mathbf{A}$ is chosen randomly from the Gaussian ensemble, then $C_1 k \log(n/k)$ measurements are sufficient for reconstructing $\hat{\mathbf{x}}$, such that $||\mathbf{x} - \hat{\mathbf{x}}||_2 \leq C_2 ||\mathbf{n}||_2$ for some constants C_1 and C_2 when minimizing the ℓ_1 norm of the estimate subject to distortion constraints on the measurements [1]–[3].

In certain practical applications, $\mathbf{A}$ may not be known exactly at the decoder. Sparse signal recovery under matrix uncertainty has been studied previously using an additive noise model [4]–[6]:

$$\mathbf{y} = (\mathbf{A} + \mathbf{E})\mathbf{x} + \mathbf{n},$$

where $\mathbf{E}$ is a perturbation matrix. It was shown that ℓ_1-regularized reconstruction procedures can reconstruct $\mathbf{x}$ within estimation error depending on $\mathbf{E}$ and $\mathbf{n}$ [4].

While these methods tackle the problem of an additive perturbation matrix, frequently knowledge of more inherent structure is available about the measurement system, for instance in magnetic resonance imaging (MRI). We consider a measurement system, which can be viewed as a combination of a function specified by a parameter vector acting on the sparse signal (potentially a non-linear function of the signal), followed by a sensing matrix:

$$\mathbf{y} = \mathbf{A}\big(\mathcal{T}(\mathbf{x}, \boldsymbol{\theta})\big) + \mathbf{n}, \tag{2}$$

where the underlying function $\mathcal{T}(\cdot, \cdot)$ is known, but the parameter vector $\boldsymbol{\theta}$ is unknown. Such models arise in the context of compressed sensing in MRI, for instance when multi-channel phased array receiver coils are utilized for image acquisition and reconstruction [7], [8] or when imaging in the presence of motion [9]. Details of such models will be discussed in Section II-A.

In this paper, we investigate information theoretic achievability conditions for jointly estimating the sparse signal $\mathbf{x}$ and parameter vector $\boldsymbol{\theta}$ from the measurement model in Eq. (2) based on a distortion-fidelity criterion, and explore how these differ from similar conditions for the standard model in Eq. (1). The main challenge in the proofs comes from the fact that $\mathcal{T}(\mathbf{x}, \boldsymbol{\theta})$ for a given $\boldsymbol{\theta}$ and a sparse $\mathbf{x}$ is not necessarily sparse, thus existing information theoretic decoders [10]–[13] are not directly applicable. The outline of the paper is given next. In Section II, we motivate the problem with examples from MRI, subsequently define the notation and measurement model in detail. We then state our main result. In Section III, we define our information theoretic decoder, and provide the proof for our main result. In Section IV, we provide a numerical simulation example for joint motion estimation and signal reconstruction based on a practical algorithm. Conclusions and directions for future research are discussed in Section V.

II. Measurement Model and Main Results

A. *Applications: $\mathcal{T}(\cdot, \cdot)$ in MRI*

The measurement model of Eq. (2) arises in certain MRI applications, where $\mathcal{T}(\cdot, \cdot)$ has the form:

$$\mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) = \begin{pmatrix} t_1(\mathbf{x}, \boldsymbol{\theta}_1) \\ \vdots \\ t_B(\mathbf{x}, \boldsymbol{\theta}_B) \end{pmatrix}, \tag{3}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_B^T]^T$ and B is an integer, determined by the application. Each $t_i(\cdot, \boldsymbol{\theta}_i) : \mathbb{C}^n \to \mathbb{C}^n$ is a mapping from the image space to the image space, affecting the intensity, contrast or the position of the image, and may be non-linear. We provide a simplified description of these functions in two scenarios:

Multi-channel phased array receiver coils: For multi-channel acquisitions, the image sensed in each receiver coil is different, for instance based on their proximity to the coil, according to the coil sensitivity maps [14]. In this case, $t_i(\cdot, \boldsymbol{\theta}_i)$ is a linear operator represented by a diagonal matrix, $\mathbf{S}_{\boldsymbol{\theta}_i}$, describing how the intensity and phase of the image is weighted voxel-wise based on the coil sensitivities; B is the number of receiver coils, and $\boldsymbol{\theta}_i$ are the coefficients of a polynomial weighting of the voxel locations, capturing the smooth variation in the coil sensitivities [14]. Furthermore, coil sensitivities are generally assumed to be normalized to satisfy $\sum_k \mathbf{S}_{\boldsymbol{\theta}_k} \mathbf{S}^*_{\boldsymbol{\theta}_k} = \mathbf{I}_n$ ($\mathbf{I}_n$ denoting the $n \times n$ identity matrix) to ensure uniqueness of the $(\mathbf{x}, \boldsymbol{\theta}_i)$ pairs.

Motion-corrupted acquisitions: If there is motion during an MRI scan, either due to the subject moving or due to the movement of the organs, there will be corruption in the measurements acquired during motion. In this case, $t_i(\cdot, \boldsymbol{\theta}_i)$ describes the motion that affects the underlying image for the i-th acquisition/segment, B is the number of acquisitions/segments and $\boldsymbol{\theta}_i$ can be described using a translational or a non-linear affine motion model [9]. To ensure uniqueness of $(\mathbf{x}, \boldsymbol{\theta}_i)$, typically the first segment is assumed to be a motion-free reference segment, i.e. $t_1(\cdot, \boldsymbol{\theta}_1)$ is the identity map.

B. Assumptions on $\mathcal{T}(\cdot, \cdot)$

We define the map which causes perturbation in the measurement system as $\mathcal{T} : \mathbb{C}^n \times \Theta \to \mathbb{C}^{nB}$, where $B \in \mathbb{Z}^+$ and $\Theta = [-U_{\boldsymbol{\theta}}, U_{\boldsymbol{\theta}}]^{M_{\boldsymbol{\theta}}}$, the $M_{\boldsymbol{\theta}}$-folded Cartesian of $[-U_{\boldsymbol{\theta}}, U_{\boldsymbol{\theta}}]$ by itself, and $U_{\boldsymbol{\theta}}, B, M_{\boldsymbol{\theta}}$ are constants knowing a-priori. The definition of Θ indicates that the search space for $\boldsymbol{\theta}$ is bounded and finite-dimensional irrespective of the signal dimension. We assume $\mathcal{T}(\mathbf{x}, \boldsymbol{\theta})$, has a number of properties. Since the measurement system itself is scale invariant as a function of $\mathbf{x}$, we have,

Property 1: $\qquad \mathcal{T}(\alpha \mathbf{x}, \boldsymbol{\theta}) = \alpha \mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) \qquad \forall \alpha \in \mathbb{C}$

We assume $\mathcal{T}(\cdot, \cdot)$ is locally Lipshitz continuous in $\mathbf{x}$:

Property 2: $\quad ||\mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) - \mathcal{T}(\mathbf{x}', \boldsymbol{\theta})||_2^2 \leq B(1+\mu)||\mathbf{x} - \mathbf{x}'||_2^2$

for $||\mathbf{x} - \mathbf{x}'||_2^2 \leq \kappa$, where $\kappa, \mu > 0$ are constants. Together with Property 1, this implies that $\mathcal{T}(\cdot, \cdot)$ is Lipschitz continuous, i.e. Property 2 holds for all $\mathbf{x}, \mathbf{x}' \in \mathbb{C}^n$. We note that this property readily holds for linear maps between finite-dimensional vector spaces. Furthermore we assume

Property 3: $\qquad ||\mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) - \mathcal{T}(\mathbf{x}, \boldsymbol{\theta}')||_2^2 \leq \eta ||\mathbf{x}||_2^2$

for $||\boldsymbol{\theta} - \boldsymbol{\theta}'||_2^2 \leq \xi$, where $\xi, \eta > 0$ are constants. Intuitively, this latter property enables a means to search the parameter space Θ for an approximate solution, since no other information about the extrema of $\mathcal{T}(\cdot, \cdot)$ is available. Finally, as alluded to in Section II-A, uniqueness of the pair $(\mathbf{x}, \boldsymbol{\theta})$ requires that

$$\mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) = \mathcal{T}(\mathbf{x}', \boldsymbol{\theta}') \Leftrightarrow \mathbf{x} = \mathbf{x}', \boldsymbol{\theta} = \boldsymbol{\theta}'.$$

We further assume a stronger condition holds:

Property 4: $\qquad ||\mathcal{T}(\mathbf{x}, \boldsymbol{\theta}) - \mathcal{T}(\mathbf{x}', \boldsymbol{\theta}')||_2^2 \geq \tau ||\mathbf{x} - \mathbf{x}'||_2^2$

for a constant $0 < \tau \leq B$. For instance for the motion models, this property holds with $\tau = 1$ (due to the motion-free reference).

C. Main Result

We consider an information theoretic formulation of the problem, similar to those for the standard model of Eq. (1) as in [10]–[13]. Let $\{\mathbf{x}^{(n)}\}_n$ be a sequence of vectors such that $\mathbf{x}^{(n)} \in \mathbb{C}^n$ and $\mathcal{I}^{(n)} = \text{supp}(\mathbf{x}^{(n)})$, with $|\mathcal{I}^{(n)}| = k^{(n)}$ and $||\mathbf{x}^{(n)}||_2^2 = P$ for a constant P. For $\mathbf{x}^{(n)}$ from this sequence, we consider $m \times nB$ random Gaussian measurement matrices, $\mathbf{A}^{(n)}$, where each entry is i.i.d. with $\mathcal{N}_C(0,1)$ complex circularly symmetric Gaussian distribution, and m is a function of n. We will omit the superscript showing the explicit dependence of $\mathcal{I}^{(n)}, k^{(n)}, \mathbf{A}^{(n)}$ on n for brevity, since this is implied by $\mathbf{x}^{(n)}$. Furthermore, let $\mathbf{n} \sim \mathcal{N}_C(\mathbf{0}, \sigma^2 \mathbf{I}_m)$ be additive noise vector with i.i.d. complex circularly symmetric Gaussian distribution, leading to a finite SNR (defined as $\mathbb{E}_{\mathbf{A}} ||\mathbf{A}\big(\mathcal{T}(\mathbf{x}^{(\mathbf{n})}, \boldsymbol{\theta})\big)||_2^2 / \mathbb{E}_{\mathbf{n}} ||\mathbf{n}||_2^2$).

For $\mathbf{y}$ generated according to Eq. (2), a decoder $\mathcal{D} : \mathbb{C}^m \to \mathbb{C}^n \times \Theta$ outputs a pair $(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \mathcal{D}(\mathbf{y})$ with $||\hat{\mathbf{x}}|| = k$. In defining the success at the decoder, we consider a distortion-based error metric between $\mathbf{x}$ and $\hat{\mathbf{x}}$

$$d_e(\mathbf{x}, \hat{\mathbf{x}}, D_{ge}) = 1 - \mathbb{I}\big(||\mathbf{x} - \hat{\mathbf{x}}||_2^2 < D_{ge}\big). \tag{4}$$

which is 1 in the case that the decoder fails to find $\hat{\mathbf{x}}$ that is sufficiently close to $\mathbf{x}$ for some pre-specified constant D_{ge}. We study the error probability, averaged over all Gaussian measurement matrices and noise,

$$p_{err}(\mathcal{D}, D_{ge}) = \mathbb{E}(d_e(\mathbf{x}, \hat{\mathbf{x}}, D_{ge}) \neq 1).$$

We say that the decoder achieves *asymptotically reliable* D_{ge}-distortion performance if $p_{err}(\mathcal{D}, D_{ge}) \to 0$ as $n \to \infty$. We now state our main result for these ensembles

Theorem 1: Let a sequence of sparse vectors, $\{\mathbf{x}^{(n)} \in \mathbb{C}^n\}_n$ with $||\mathbf{x}^{(n)}||_0 = k$ be given. Let $m > C_{ge} k \log(n/k)$ for a pre-specified constant $C_{ge} > 1$. Let $\mathcal{T}(\cdot, \cdot)$ be a map satisfying the properties in Section II-B. Then for the model in Eq. (2), asymptotically reliable $D_{\text{joint model}}$-distortion performance is achievable if

$$D_{\text{joint model}} > \frac{1}{\tau}\bigg(2B(1+\mu)\epsilon + 2\eta P + \sigma^2\bigg)$$

for a constant ϵ depending on C_{ge}. Furthermore, for the same number of measurements, a decoder that only estimates sparse $\mathbf{x}$ from the model in Eq. (1), can achieve asymptotically reliable $D_{\text{standard model}}$-distortion with

$$D_{\text{standard model}} > \epsilon + \sigma^2$$

Proof: Please see Section III-B. ∎

The theorem suggests that a distortion penalty is incurred via the joint estimation procedure as intuitively expected. However, the penalty is not severe in the sense that it does not alter the order of the sufficient number of measurements.

III. Decoder and Proof of the Main Result

A. Covering-Based Decoder

Our decoder relies on a covering of Θ and all k-dimensional subspaces of $\mathbb{C}^n$ by a finite number of points as in [15], and then using these points in an ℓ_2-norm based decoder that searches through these nets of points. We note that this approach builds on prior proof techniques with information theoretic decoders [10]–[13], which search through projections onto each k-dimensional subspace. However, these decoders are not applicable to the model in Eq. (2), since $\mathcal{T}(\mathbf{x},\boldsymbol{\theta})$ is not necessarily sparse for a given $\boldsymbol{\theta}$ and k-sparse $\mathbf{x}$, and furthermore $\mathcal{T}(\cdot,\cdot)$ may not be explicitly known. We also note that the brute force search proposed here may not be computationally feasible to implement.

Lemma 2: ([15]) Let $\mathbf{e}_k$ be the k^{th} canonical basis vector. Let $T=\{i_1,i_2,\dots,i_k\}$, and $V_T=\{\sum_k \lambda_k \mathbf{e}_{i_k} \big| \sum_k |\lambda_k|^2 = P\}$. Then a finite set of points $Q_T \subseteq V_T$ can be chosen with $|Q_T| \le (12P/\epsilon)^k$ such that for all $\mathbf{v}\in V_T$

$$\min_{\mathbf{q}\in Q_T} ||\mathbf{q}-\mathbf{v}||_2^2 \le \epsilon.$$

We let $Q=\bigcup_T Q_T$, where the union is over all index sets of size k. Similarly Θ can be covered with a set Λ and covering radius $\sqrt{\rho}$. We note the following about the cardinality of Q and Λ

$$|Q| \le \binom{n}{k}|Q_T| \le \exp\Bigg(k\Big(\log(en/k)+\log(12P/\epsilon)\Big)\Bigg) \quad (5)$$

and

$$|\Lambda| \le (24U_{\boldsymbol{\theta}}/\rho)^{M_{\boldsymbol{\theta}}}.$$

Note for Θ, the cardinality of the covering set is fixed as a function of n. Given Q, Λ, $\mathbf{y}$ measured according to Eq. (2), $\mathbf{A}$ and $\mathcal{T}(\cdot,\cdot)$, we define

$$W_\Delta = \Bigg\{(\mathbf{x}',\boldsymbol{\theta}')\in Q\times\Lambda \;\Bigg|\; \frac{1}{m}||\mathbf{y}-\mathbf{A}\big(\mathcal{T}(\mathbf{x}',\boldsymbol{\theta}')\big)||_2^2 < \Delta\Bigg\}. \quad (6)$$

The covering-based decoder $\mathcal{D}^{\Delta}_{\text{cover}}$ outputs any pair in W for a given $\mathbf{y}$. If there are more than one element in W_Δ, it picks one of these elements arbitrarily. If there are no elements W_Δ, all-zero vector is output and an error is declared.

B. Proof of Theorem 1

We first consider the distribution of $||\mathbf{y}-\mathbf{A}\big(\mathcal{T}(\mathbf{x}',\boldsymbol{\theta}')\big)||_2^2$ for a given $(\mathbf{x}',\boldsymbol{\theta}')$. Noting that each entry of $\mathbf{A}$ is i.i.d. $\mathcal{N}_C(0,1)$, $\mathbf{n}\sim\mathcal{N}_C(\mathbf{0},\sigma^2\mathbf{I}_m)$, and $\mathbf{y}=\mathbf{A}\big(\mathcal{T}(\mathbf{x},\boldsymbol{\theta})\big)+\mathbf{n}$, we have

$$\begin{aligned}||\mathbf{y}-\mathbf{A}\big(\mathcal{T}(\mathbf{x}',\boldsymbol{\theta}')\big)||_2^2 &= ||\mathbf{A}\big(\mathcal{T}(\mathbf{x},\boldsymbol{\theta})-\mathcal{T}(\mathbf{x}',\boldsymbol{\theta}')\big)+\mathbf{n}||_2^2\\ &\triangleq ||\mathbf{n}'||_2^2,\end{aligned}$$

where due to the distributions on $\mathbf{A}$ and $\mathbf{n}$, we have

$$\mathbf{n}'\sim\mathcal{N}_C\Bigg(\mathbf{0},\Big(||\mathcal{T}(\mathbf{x},\boldsymbol{\theta})-\mathcal{T}(\mathbf{x}',\boldsymbol{\theta}')||_2^2+\sigma^2\Big)\mathbf{I}_m\Bigg).$$

Next, we state a technical lemma about tails of chi-square random variables.

Lemma 3: Let $\mathbf{u}\sim\mathcal{N}_C(\mathbf{0},\nu^2\mathbf{I}_m)$. If $\omega<\nu^2$, then

$$\mathbb{P}\Big(\frac{1}{m}||\mathbf{u}||_2^2<\omega\Big)\le\exp\Bigg(-\frac{m}{4}\Big(1-\frac{\omega}{\nu^2}\Big)^2\Bigg), \quad (7)$$

and if $\omega>\nu^2$,

$$\mathbb{P}\Big(\frac{||\mathbf{u}||_2^2}{m}>\omega\Big)\le\exp\Bigg(-\frac{m}{4}\Big(\sqrt{2\frac{\omega}{\nu^2}-1}-1\Big)^2\Bigg). \quad (8)$$

Proof: First note $\Upsilon=||\mathbf{u}||_2^2/\nu^2$ is a chi-square random variable with m degrees of freedom. For the first part, we use the following tail bound for chi-squre random variables [10], [12]

$$\mathbb{P}\big(\Upsilon-m\le-2\sqrt{m\beta}\big)\le e^{-\beta}$$

Noting $\omega<\nu^2$ and letting

$$\beta=\Bigg(\frac{1}{2}\Big(1-\frac{\omega}{\nu^2}\Big)\sqrt{m}\Bigg)^2$$

leads to the desired result. For the second part, we use another tail bound [10], [12]

$$\mathbb{P}\big(\Upsilon-m\ge2\sqrt{m\beta}+2\beta\big)\le e^{-\beta}.$$

For $\omega>\nu^2$, we let

$$\beta=\Bigg(\frac{1}{2}\Big(\sqrt{1+2\Big(\frac{\omega}{\nu^2}-1\Big)}-1\Big)\sqrt{m}\Bigg)^2.$$

∎

We let $\Delta=2B(1+\mu)\epsilon+2\eta P+\sigma^2+\zeta>\tau D_{\text{joint model}}$ for an arbitrary constant $\zeta>0$ in the definition of the set W_Δ. We will show that with high probability, there is a signal within distortion $D_{\text{joint model}}$ of the original signal, and with high probability that the decoder will not choose a signal with a greater distortion. We define two types of error events for the decoder

- E_1: Let $\tilde{\mathbf{q}}\in Q$ be a point in the covering set with $||\mathbf{x}-\tilde{\mathbf{q}}||_2^2<\epsilon$, and let $\tilde{\boldsymbol{\theta}}\in\Lambda$ be a point with $||\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}||_2^2<\xi$. We note by its definition in Theorem 1, $D_{\text{joint model}}>\epsilon$. We define the following event

$$E_1=\{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})\notin W_\Delta \;|\; ||\mathbf{x}-\tilde{\mathbf{q}}||_2^2<\epsilon, ||\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}||_2^2<\xi\}$$

- $E_{(\mathbf{z}',\boldsymbol{\theta}')}$: Let $\mathbf{z}'\in Q$ be a point such that $||\mathbf{x}-\mathbf{z}'||_2^2>\frac{\Delta}{\tau}>D_{\text{joint model}}$, and $\boldsymbol{\theta}'\in\boldsymbol{\theta}$. We define

$$E_{(\mathbf{z}',\boldsymbol{\theta}')}=\{(\mathbf{z}',\boldsymbol{\theta}')\in W_\Delta \;|\; ||\mathbf{x}-\mathbf{z}'||_2^2>\frac{\Delta}{\tau},\boldsymbol{\theta}'\in\boldsymbol{\theta}\}$$

For $\tilde{\mathbf{q}} \in Q$ and $\tilde{\boldsymbol{\theta}} \in \Lambda$ as defined in E_1, we have

$$
\begin{aligned}
\sigma^2_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})} &\triangleq ||\mathcal{T}(\mathbf{x},\boldsymbol{\theta}) - \mathcal{T}(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})||_2^2 + \sigma^2 \\
&= ||\mathcal{T}(\mathbf{x},\boldsymbol{\theta}) - \mathcal{T}(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}}) + \mathcal{T}(\mathbf{x},\tilde{\boldsymbol{\theta}}) - \mathcal{T}(\mathbf{x},\tilde{\boldsymbol{\theta}})||_2^2 + \sigma^2 \\
&\leq 2||\mathcal{T}(\mathbf{x},\boldsymbol{\theta}) - \mathcal{T}(\mathbf{x},\tilde{\boldsymbol{\theta}})||_2^2 \\
&\qquad + 2||\mathcal{T}(\mathbf{x},\tilde{\boldsymbol{\theta}}) - \mathcal{T}(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})||_2^2 + \sigma^2 \\
&\leq 2\eta||\mathbf{x}||_2^2 + 2B(1+\mu)\epsilon + \sigma^2 \\
&= 2\eta P + 2B(1+\mu)\epsilon + \sigma^2 \triangleq \Delta_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})},
\end{aligned} \tag{9}
$$

where the last inequality follows from Properties 1, 2 and 3 of $\mathcal{T}(\cdot,\cdot)$. Noting $\sigma^2_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})} \leq \Delta_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})} < \Delta$, from Inequality (8) of Lemma 3, we have

$$
\begin{aligned}
\mathbb{P}(E_1) &= \mathbb{P}\left(\frac{1}{m}||\mathbf{y} - \mathbf{A}\big(\mathcal{T}(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})\big)||_2^2 > \Delta\right) \\
&\leq \exp\left(-\frac{m}{4}\left(\sqrt{2\frac{\Delta}{\sigma^2_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})}} - 1} - 1\right)^2\right) \\
&\leq \exp\left(-\frac{m}{4}\left(\sqrt{2\frac{\Delta}{\Delta_{(\tilde{\mathbf{q}},\tilde{\boldsymbol{\theta}})}} - 1} - 1\right)^2\right) \\
&= \exp\left(-\frac{m}{4}\left(\sqrt{\frac{2B(1+\mu)\epsilon + 2\eta P + \sigma^2 + 2\zeta}{2B(1+\mu)\epsilon + 2\eta P + \sigma^2}} - 1\right)^2\right)
\end{aligned} \tag{10}
$$

For $\mathbf{z}' \in Q$ such that $||\mathbf{x} - \mathbf{z}'||_2^2 > \frac{\Delta}{\tau}$, and any $\boldsymbol{\theta}' \in \boldsymbol{\theta}$ as defined in $E_{(\mathbf{z}',\boldsymbol{\theta}')}$, we have

$$
\begin{aligned}
\sigma^2_{(\mathbf{z}',\boldsymbol{\theta}')} &\triangleq ||\mathcal{T}(\mathbf{x},\boldsymbol{\theta}) - \mathcal{T}(\mathbf{z}',\boldsymbol{\theta}')||_2^2 + \sigma^2 \\
&\geq \tau||\mathbf{x} - \mathbf{z}'||_2^2 + \sigma^2 > \Delta + \sigma^2,
\end{aligned} \tag{11}
$$

using Property 4 of $\mathcal{T}(\cdot,\cdot)$. From Inequality (7) of Lemma 3, we have

$$
\begin{aligned}
\mathbb{P}(E_{(\mathbf{z}',\boldsymbol{\theta}')}) &= \mathbb{P}\left(\frac{1}{m}||\mathbf{y} - \mathbf{A}\big(\mathcal{T}(\mathbf{z}',\boldsymbol{\theta}')\big)||_2^2 < \Delta\right) \\
&\leq \exp\left(-\frac{m}{4}\left(1 - \frac{\Delta}{\sigma^2_{(\mathbf{z}',\boldsymbol{\theta}')}}\right)^2\right) \\
&\leq \exp\left(-\frac{m}{4}\left(\frac{\sigma^2}{\Delta + \sigma^2}\right)^2\right)
\end{aligned} \tag{12}
$$

We consider

$$
C_{ge} > 4\left(\frac{\Delta + \sigma^2}{\sigma^2}\right)^2\left(\frac{\log(n/k) + \log(12P/\epsilon) + 1}{\log(n/k)}\right),
$$

and evaluate

$$
\begin{aligned}
p_{err}(\mathcal{D}, D_{\text{joint model}}) &\leq \mathbb{P}\Bigg(E_1 \bigcup_{\substack{\boldsymbol{\theta}' \in \Lambda, \mathbf{z}' \in Q \\ ||\mathbf{x} - \mathbf{z}'||_2^2 > D_{\text{joint model}}}} E_{(\mathbf{z}',\boldsymbol{\theta}')}\Bigg) \\
&\leq \exp\left(-\frac{m}{4}\left(\sqrt{\frac{2B(1+\mu)\epsilon + 2\eta P + \sigma^2 + \zeta}{2B(1+\mu)\epsilon + 2\eta P + \sigma^2}} - 1\right)^2\right) \\
&\quad + |Q||\Lambda| \exp\left(-\frac{m}{4}\left(\frac{\sigma^2}{\Delta + \sigma^2}\right)^2\right)
\end{aligned} \tag{13}
$$

by the union bound. The first term goes to 0 with large n. With the specified choice of C_{ge}, and the cardinality of Q and Λ, the second term also goes to 0 as $n \to \infty$. Thus asymptotically reliable distortion of $D_{\text{joint model}}$ can be achieved.

The second part of the proof follows the first, let $\boldsymbol{\Phi}$ be a $m \times n$ matrix with i.i.d. $\mathcal{N}_C(0,1)$ entries, where m and C_{ge} are as given above. Then

$$
||\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}'||_2^2 = ||\boldsymbol{\Phi}\mathbf{x} - \mathbf{x}' + \mathbf{n}||_2^2 = ||\mathbf{n}''||_2^2
$$

with

$$
\mathbf{n}'' \sim \mathcal{N}_C\Big(\mathbf{0}, \big(||\mathbf{x} - \mathbf{x}'||_2^2 + \sigma^2\big)\mathbf{I}_m\Big).
$$

The same ϵ-covering, Q, can be applied since k, m, n are unchanged. Setting $\Delta = \epsilon + \sigma^2 + \zeta > D_{\text{standard model}}$ for $\zeta > 0$ in defining W_Δ, and considering the error events (only over the covering set Q since Λ does not exist in the model) with the same bounding techniques leads to the desired result.

IV. Numerical Example

In this section, we simulate a measurement model with rotational and translational motion during the acquisition. The 256×256 Shepp-Logan phantom (Figure 1a) was imaged using a Gaussian sensing matrix with $m = n/3 = 21844$ measurements. After $n/6$ measurements were taken, the Shepp-Logan phantom was rotated by $10.5°$, and translated by 5.3 pixels in the x-direction and 6.8 pixels in the y-direction (Figure 1b). The subsequent $n/6$ measurements were taken on this image. All measurements were corrupted with noise for a final SNR of 20 dB. We note that this would be considered a low SNR for MRI applications.

The image was reconstructed using four different approaches: 1) The first $n/6$ measurements were used to reconstruct the motion-free image, 2) The second $n/6$ measurements were used to reconstruct the image after motion, 3) All $n/3$ measurements were used with the standard model in Eq. (1) without accounting for the motion in $\mathbf{A}$, 4) A joint reconstruction was performed on all $n/3$ measurements with the model in Eq. (2). Reconstructions were performed using ℓ_1 minimization on the Haar wavelet coefficients. Although total variation minimization is frequently used for this phantom, Haar wavelets were utilized in this case to be compatible with the theory. Thus, for the joint model, $t_1(\mathbf{x}, 0) = \mathbf{W}\mathbf{x}$ for the first motion-free reference segment, where $\mathbf{W}$ is the inverse wavelet transform and $\mathbf{x}$ are the wavelet coefficients. For the second segment $t_2(\mathbf{x}, \boldsymbol{\theta}_2) = \mathcal{M}_{\boldsymbol{\theta}_2}(\mathbf{W}\mathbf{x})$, where $\mathcal{M}_{\boldsymbol{\theta}_2}$ is a rotation by $\boldsymbol{\theta}_{rot}$ degrees, x-translation by $\boldsymbol{\theta}_x$ pixels and y-translation by $\boldsymbol{\theta}_y$ pixels; and $\boldsymbol{\theta}_2 = [\boldsymbol{\theta}_{rot}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_y]$. To simulate the effect of the imperfect covering of $\boldsymbol{\theta}$, the search space was restricted to $1°$ angles for rotation and 0.5 pixel shifts for x and y translations. The motion parameters were estimated using the reconstructed images from approaches (1) and (2), and subsequently used in the reconstruction of (4).

The results are depicted in Fig 1(c-f). As seen in (c) and (d), $n/6$ measurements are not enough for a high-fidelity reconstruction, which leads to visible blurring. However, these were sufficient to estimate the motion parameters as

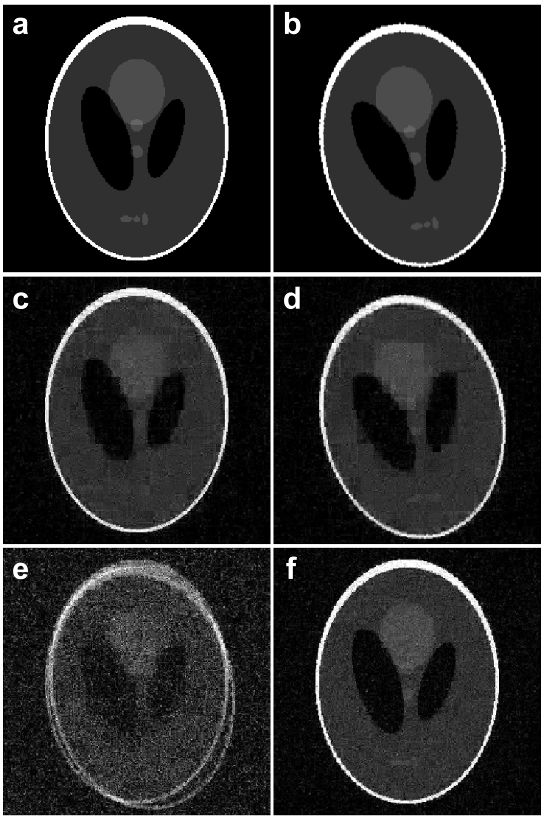

Fig. 1. Results of the numerical simulation using the Shepp-Logan phantom: (a) The motion-free image, (b) The image after undergoing 10.5° rotation, 5.3 pixel x-translation and 6.8 pixel y-translation, (c) Reconstruction of (a) from $n/6$ measurements with SNR = 20 dB, (d) Reconstruction of (b) from $n/6$ measurements, (e) Direct reconstruction from $n/3$ measurements (of (c) and (d)) according to Eq. (1), (f) Joint estimation of motion parameters and sparse signal based on the model in Eq (2). The joint reconstruction is sharper compared to (c) and (d) due to higher number of measurements, and does not suffer from motion artifacts as in (e) due to the motion estimation.

$\hat{\boldsymbol{\theta}} = [10^\circ, 5.5, 7]$. The reconstruction in (e) that does not take into account the motion corruption suffers from visible motion artifacts. The joint estimation procedure using $\hat{\boldsymbol{\theta}}$ is able to reconstruct the image with almost no visible motion artifacts and less blurring compared to (c) and (d), although noise is visible due to the imperfect estimation of the motion parameters, as well as the presence of measurement noise. In fact, more than a 3-fold distortion penalty is incurred by the joint estimation in (d) when compared to the case when $\mathbf{A}$ is used to directly measure the motion-free image with $n/3$ measurements at the same noise level and reconstructed according to Eq. (1) (reconstructed image not shown), with normalized (with respect to the the energy in the image) MSEs of 0.083 vs. 0.026.

V. Conclusions

In this paper, we studied a measurement system model for a k-sparse signal, consisting of a (possibly non-linear) transformation specified by a parameter vector, followed by a sensing matrix, $\mathbf{A} \in \mathbb{C}^{m \times n}$ with $m < n$, in the presence of additive measurement noise; a model that arises in certain applications in magnetic resonance imaging. We proposed an information theoretic decoder that attempts to recover the signal and the transformation parameter vector jointly, and characterized sufficiency results based on an ℓ_2 norm-based distortion criterion. We showed that a distortion penalty is incurred for the achievable distortion by the joint estimation process for the same number of measurements, when compared to the scenario where the sparse signal is measured without the parametric transformation, when using the same decoder. We also provided a numerical example to provide a proof-of-concept for potential applications, and to motivate the need for the development of computationally-feasible algorithms as in [16] for such joint estimation problems, which can be utilized efficiently for real-life applications. Extension of the proofs to partial Fourier matrices and the development of necessary conditions are also left for future work.

Acknowledgment

Mehmet Akçakaya would like to acknowledge grant support from NIH K99HL111410-01.

References

[1] E. J. Candès, J. Romberg and T. Tao, “Stable signal recovery for incomplete and inaccurate measurements,” *Commun. Pure Appl. Math.*, vol. 59, pp. 1207-1223, Aug. 2006.

[2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, April 2006.

[3] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589-592, May 2008.

[4] M. Rosenbaum and A. Tsybakov, “Sparse recovery under matrix uncertainty,” *The Annals of Statistics*, vol. 38, 99. 2620-2651, Oct. 2010.

[5] H. Zhu, G. Leus, and G. Giannakis, “Sparsity-cognizant total least-squares for perturbed compressive sampling,” *IEEE Trans. Sig. Proc.*, vol. 59, pp. 20022016, May 2011.

[6] J. T. Parker, V. Cevher, and P. Schniter, “Compressive sensing under matrix uncertainties: an approximate message passing approach,” *Proc. Conf. Forty Fifth ASILOMAR*, pp. 804-808, Nov. 2011.

[7] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R.A. Barth, J. M. Pauly, M. Lustig, “Improved pediatric MR imaging with compressed sensing,” *Radiology*, vol. 256, pp. 607-616, Aug. 2010.

[8] M. Akçakaya, T. A. Basha, R. H. Chan, W. J. Manning and R. Nezafat, “Accelerated isotropic sub-millimeter whole-heart coronary MRI: compressed sensing versus parallel imaging,” *Magn Reson Med*, in press.

[9] M. Usman, D. Atkinson, F. Odille, C. Kolbitsch, G. Vaillant, T. Schaeffter, P. G. Batchelor and C. Prieto, “Motion corrected compressed sensing for free-breathing dynamic cardiac MRI,” *Magn Reson Med*, in press.

[10] M. J. Wainwright, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5728-5741, Dec. 2009.

[11] A. K. Fletcher, S. Rangan and V. K. Goyal, “Necessary and sufficient conditions on sparsity pattern recovery,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5758-5772, Dec. 2009.

[12] M. Akçakaya and V. Tarokh, “Shannon theoretic limits on noisy compressive sampling,” *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492-504, Jan. 2010.

[13] S. Aeron, V. Saligrama, M. Zhao, “Information theoretic bounds for compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 5111-5130, Oct. 2010.

[14] K. P. Pruessmann, M. Weiger, M. B. Scheidegger and P. Boesiger, “SENSE: Sensitivity encoding for fast MRI” *Magn Reson Med*, vol. 42, pp. 952-962, Nov. 1999.

[15] R. Baraniuk, M. Davenport, R. DeVore and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approx.*, vol. 28, pp. 253-263, Dec. 2008.

[16] G. Puy and P. Vandergheynst, “Robust joint reconstruction of misaligned images using semi-parametric dictionaries,” *Proc. ICML*, June 2012.