RECONSTRUCTION OF SIGNALS FROM MAGNITUDES OF REDUNDANT REPRESENTATIONS: THE COMPLEX CASE

RADU BALAN

Abstract. This paper is concerned with the question of reconstructing a vector in a finite-dimensional complex Hilbert space when only the magnitudes of the coefficients of the vector under a redundant linear map are known. We present new invertibility results as well as an iterative algorithm that finds the least-square solution which is robust in the presence of noise. We analyze its numerical performance by comparing it to the Cramer-Rao lower bound.

AMS (MOS) Subject Classification Numbers: 15A29, 65H10, 90C26

Key words: Frame, phase retrieval, Cramer-Rao lower bound, phaseless reconstruction

Communicated by Emmanuel Candes

1. Introduction

This paper is concerned with the question of reconstructing a vector x in a finite-dimensional *complex* Hilbert space H when only the magnitudes of the coefficients of the vector under a redundant linear map are known.

Specifically our problem is to reconstruct $x \in H$ up to a global phase factor from the magnitudes $\{|\langle x, f_k\rangle| \ , \ 1 \leq k \leq m\}$ where $\{f_1, \cdots, f_m\}$ is a frame (complete system) for H. The real case was considered in [6]. Here we develop the theory for the complex case.

A previous paper [3] described the importance of this problem to signal processing, in particular to the analysis of speech. The problem appears in X-Ray crystallography under the name of "phase retrieval" problem (see [9]) where the frame vectors are the Fourier frame vectors. The case of the windowed Fourier transform was considered in the '80s [29]; see also [5] for a frame based approach to this case. A different approach is taken in [33]. The authors propose a novel algorithm adapted to compactly supported signals and FFT that uses individual signal spectral powers and two additional interferences between signals. A 3-term polarization identity has been used in [1] together with the angular synchronization algorithm.

Recently the authors of [13] developed a convex optimization algorithm (PhaseLift) and proved its ability to perform exact reconstruction in the absence of noise, as well as its stablity under noise conditions. In a separate paper, [14], the authors developed further a

similar algorithm in the case of windowed DFT. The original requirement of $m = O(n\,log(n))$ vectors has been relaxed to $m = O(n)$ in [15]. Similar convex optimization solutions have been proposed by the authors of [20] and [34]. Additionally, [23] studied the duality gap in this approach and obtained a necessary and sufficient condition for the existence of a dual certificate.

While writing this paper, we become aware of [8] where the authors obtained similar results to the injectivity criteria presented here, as well as to the Cramer-Rao Lower Bound derived in this paper. We will comment more later in the paper. We also acknowledge the paper [21] where certain Lipschitz bounds have been obtained in the real case. The real case of the stabilty bounds obtained here are presented in a separate paper [7] together with additional results for the real case.

The organization of the paper is as follows. In section 2 we define the problem explicitly. In section 3 we describe new analysis results. Specifically we analyze in more detail spaces of symmetric operators of constrained signature, and then we show how they are related to the phaseless recovery problem. Our results are canonical, meaning they are independent to a particular choice of basis. In section 4 we establish two robustness results: bi-Lipschitzianity of the nonlinear analysis map, and the Cramer-Rao Lower Bound (CRLB). In section 5 we present a new reconstruction algorithm based on the Least-Square method. We also obtain robustness bounds to noise. Its performance is analyzed in section 6 and is compared to the CRLB. Section 7 contains conclusions and is followed by references.

2. Background

Let H be an n-dimensional complex Hilbert space (such as $\mathbb{C}^n$ or $\mathbb{C}^{p_1\times p_2}$ the vector space of $p_1 \times p_2$ complex matrices) with scalar product $\langle , \rangle$ linear in the first term and antilinear in the second term and a *conjugation* $c : H \to H$ (see e.g. [22]; conjugation is an antilinear transformation that squares to the identity). Let $\mathbb{F} = \{f_1, \cdots, f_m\}$ be a spanning set of m vectors in H. As H has finite dimension such a set forms a *frame*. In the infinite dimensional case, the concept of frame involves a stronger property than completeness (see for instance [16]). We review additional terminology and properties which remain still true in the infinite dimensional setting. The set $\mathbb{F}$ is a frame if and only if there are two positive constants $0 < A \leq B < \infty$ (called frame bounds) such that

$$A\|x\|^2 \leq \sum_{k=1}^{m} |\langle x, f_k\rangle|^2 \leq B\|x\|^2 \ .$$

When we can choose $A = B$ the frame is said to be *tight*. For $A = B = 1$ the frame is called *Parseval*. A set of vectors $\mathbb{F}$ of the n-dimensional Hilbert space H is said to have *full spark* if every subset of n vectors is linearly independent (see [2] for a full discussion of such frames).

2.1. Problem Definition and Notations.

For a vector $x \in H$, the collection of coefficients $\{\langle x, f_k\rangle \ , \ 1 \leq k \leq m\}$ represents the analysis of the vector x given by the frame $\mathbb{F}$. In H we

consider the following equivalence relation:

$$x, y \in H \ , \ x \sim y \text{ iff } y = zx \text{ for some scalar z with } |z| = 1. \tag{2.1}$$

Note $z = e^{i\varphi}$ for some real number φ. Let $\hat{H} = H/\sim$ be the set of classes of equivalence induced by this relation. Thus $\hat{x} = \{e^{i\alpha}x, 0 \leq \alpha < 2\pi\}$. The analysis map induces the following nonlinear map

$$\alpha : \hat{H} \to (\mathbb{R}^+)^m \ , \ \alpha(\hat{x}) = (|\langle x, f_k\rangle|^2)_{1\leq k\leq m} \tag{2.2}$$

where $\mathbb{R}^+ = \{x \ , \ x \in \mathbb{R} \ , \ x \geq 0\}$ is the set of nonnegative real numbers. In [6] we studied when the nonlinear map α is injective, mostly in the real case, and we provided some necessary conditions of injectivity in the complex case. We review these results below. In this paper we obtain additional injectivity results. We then concentrate on the additive white Gaussian noise model

$$y = \alpha(x) + \nu \ , \ \nu \sim \mathcal{N}(0, \sigma^2). \tag{2.3}$$

We describe an algorithm (the Iterative Regularized Least-Square (IRLS) algorithm) to solve the estimation problem. We prove some convergence results and we study its performance in the noisy case. We shall derive the Cramer-Rao Lower Bound (CRLB) for this model and compare the algorithm performance to this bound.

We describe several objects that will be used in this paper.

The set $B(H)$ denotes the set of bounded linear operators on H. In $B(H)$ for any $1 \leq p \leq \infty$, the p-norm of $T \in B(H)$ denoted $\|T\|_p$ is given by the p-norm of its set of singular eigenvalues. In particular for $p = 1$, the 1-norm $\|T\|_1$ is called the nuclear norm of T; for $p = 2$, the 2-norm $\|T\|_2 = \sqrt{tr(T^*T)}$ is the Frobenius norm of T; for $p = \infty$, the ∞-norm $\|T\|_\infty$ is the same as the operator norm of T on H, simply denoted $\|T\|$.

In general, for two operators $T, S \in B(H)$ we denote $\langle T, S\rangle_{B(H)} = tr\{TS^*\}$ their Hilbert-Schmidt scalar product, where S^* is the adjoint of S and tr denotes the trace. Note the scalar product is basis independent, but it depends on the underlying Hilbert space structure. When no danger of confusion we drop the index $B(H)$ from the scalar product notation.

For each frame vector f_k we denote by F_k its associated rank-1 operator

$$F_k : H \to H \ , \ F_k(x) = \langle x, f_k\rangle f_k. \tag{2.4}$$

In general the *associated rank-1 operator* to a vector $x \in H$ is the operator $X : H \to H$, $X = xx^*$ which acts by $X(v) = \langle v, x\rangle x$. Here and throughout the paper x^* denotes the adjoint (or dual) of x, that is $x^* : H \to \mathbb{C}$, $x^*(v) = \langle v, x\rangle$. Note X has *at most* rank one. Specifically, X has rank one if and only if $x \neq 0$; otherwise X has rank zero.

For any two vectors $u, v \in H$ we define their *symmetric outer product* denoted $[\![u, v]\!]$ by

$$[\![u, v]\!] : H \to H \ , [\![u, v]\!] = \frac{1}{2}(uv^* + vu^*) \ , \ [\![u, v]\!](x) = \frac{1}{2}(\langle x, u\rangle v + \langle x, v\rangle u) \ . \tag{2.5}$$

Note the rank-1 operator associated to a vector x can be written as $[\![x, x]\!]$. In particular $F_k = [\![f_k, f_k]\!]$. Note also $[\![u, v]\!]$ is $\mathbb{R}$-bilinear but it is not $\mathbb{C}$-(bi)linear. Furthermore $[\![u, v]\!] = [\![v, u]\!]$.

Following [4] the nonlinear map α induces a linear map $\mathcal{A}$ on the set $B(H)$ of bounded operators on H:

$$\mathcal{A} : B(H) \to \mathbb{C}^m \ , \ (\mathcal{A}(T))_k = \langle Tf_k, f_k\rangle = tr\{TF_k\} \ , \ 1 \leq k \leq m. \tag{2.6}$$

Thus $(\mathcal{A}(T))_k = \langle T, F_k\rangle_{B(H)}$. Also note $\alpha(x) = \mathcal{A}(X)$ where $X = xx^*$ is the rank-1 operator associated to x. This remark was first observed by B. Bodmann in [4].

Let $Sym(H)$ denote the set of self-adjoint operators on H, $Sym(H) = \{T \in B(H) \ , \ T^* = T\}$. We denote by $\mathcal{S}^{p,q}$ or $\mathcal{S}^{p,q}(H)$, the set of self-adjoint operators on H that have at most p positive eigenvalues and at most q negative eigenvalues:

$$\begin{aligned}\mathcal{S}^{p,q} =& \{T \in Sym(H) \ , \ Sp(T) = \{\lambda_1, \cdots, \lambda_n\} \ , \\ & \lambda_1 \geq \cdots \geq \lambda_p \geq 0 = \lambda_{p+1} = \cdots = \lambda_{n-q} \geq \lambda_{n-q+1} \geq \cdots \geq \lambda_n\}\end{aligned} \tag{2.7}$$

where $Sp(T)$ denotes the spectrum of T (i.e. the set of its eigenvalues). Notice $\mathcal{S}^{p,q}$ is not a linear space, but instead it is a cone in $B(H)$. This cone property is key in deriving robustness and stability bounds later on.

We denote by $\lambda_{max}(T)$ the largest eigenvalue of T and by $\lambda_{min}(T)$ the smallest eigenvalue of T. In particular we are interested in $\mathcal{S}^{1,0}$ and $\mathcal{S}^{1,1}$:

$$\mathcal{S}^{1,0} = \{T \in Sym(H) \ , \ rank(T) \leq 1, \lambda_{min}(T) = 0\} \tag{2.8}$$

$$\begin{aligned}\mathcal{S}^{1,1} = & \{T \in Sym(H) \ , \ rank(T) \leq 2, Sp(T) = \{\lambda_{max}(T), 0^{(n-2)}, \lambda_{min}(T)\}, \\ & \lambda_{max}(T) \geq 0 \geq \lambda_{min}(T)\}\end{aligned} \tag{2.9}$$

Note the following obvious inclusions

$$\{0\} \subset \mathcal{S}^{1,0} \subset \mathcal{S}^{1,1} \subset Sym(H) \ , \ \{0\} \subset \mathcal{S}^{0,1} \subset \mathcal{S}^{1,1} \subset Sym(H) \tag{2.10}$$

We denote by $\mathring{\mathcal{S}}^{p,q}$ the subset of $\mathcal{S}^{p,q}$ of selfadjoint operators that have rank $p+q$, hence exactly p strictly positive eigenvalues and q strictly negative eigenvalues. Thus

$$\mathcal{S}^{p,q} = \mathring{\mathcal{S}}^{p,q} \cup \mathcal{S}^{p-1,q} \cup \mathcal{S}^{p,q-1} \tag{2.11}$$

represents a disjoint partition of $\mathcal{S}^{p,q}$. In particular

$$\mathcal{S}^{1,1} = \mathring{\mathcal{S}}^{1,1} \cup \mathcal{S}^{0,1} \cup \mathcal{S}^{1,0} \ , \ \mathcal{S}^{1,0} = \mathring{\mathcal{S}}^{1,0} \cup \{0\} \ . \tag{2.12}$$

Finally we let $GL(H)$ denote the group of invertible linear operators on H. A more detailed analysis of these sets is presented in subsection 3.1.

Next we describe the realification of the Hilbert space H. To do so canonically we need to fix a conjugation $c : H \to H$. To the complex Hilbert space H with conjugation c we associate its $2n$-dimensional real vector space $H_{\mathbb{R}}$ subset of $H \times H$ built from vectors $v_R = \frac{1}{2}(v + c(v))$ and $v_I = \frac{1}{2i}(v - c(v))$ as follows:

$$H_{\mathbb{R}} = \left\{ \left(\frac{1}{2}(v + c(v)), \frac{1}{2i}(v - c(v)) \right) \ , \ v \in H \right\}. \tag{2.13}$$

Thus $H_{\mathbb{R}}$ is the image of H through the $\mathbb{R}$-linear map,

$$\mathbf{j} : H \to H \times H \ , \ \mathbf{j}(v) = \left(\frac{1}{2}(v + c(v)), \frac{1}{2i}(v - c(v)) \right). \tag{2.14}$$

Note $\mathbf{j}$ is injective with range $H_{\mathbb{R}}$. Furthermore $\mathbf{j} : H \to H_{\mathbb{R}}$ is a norm preserving $\mathbb{R}$-isomorphism. Its inverse is given by

$$\mathbf{j}^{-1} : H_{\mathbb{R}} \to H \ , \ \mathbf{j}^{-1}(u, v) = u + iv \ . \tag{2.15}$$

Let J denote the linear map defined by

$$J : H \times H \to H \times H \ , \ J(v, w) = (-w, v) \ . \tag{2.16}$$

Note it is conjugate to the multiplication by i in H:

$$J : H_{\mathbb{R}} \to H_{\mathbb{R}} \ , \ J(\mathbf{j}(v)) = \mathbf{j}(iv) \ . \tag{2.17}$$

Hence $H_{\mathbb{R}}$ is J invariant. In $H_{\mathbb{R}}$ the induced scalar product is given by

$$\langle \mathbf{j}(v), \mathbf{j}(w) \rangle := \left\langle \frac{1}{2}(v + c(v)), \frac{1}{2}(w + c(w)) \right\rangle + \left\langle \frac{1}{2i}(v - ic(v)), \frac{1}{2i}(w - c(w)) \right\rangle = real(\langle v, w \rangle). \tag{2.18}$$

We denote by $\langle v, w \rangle_{\mathbb{R}} = real(\langle v, w \rangle)$ the $\mathbb{R}$-linear inner product on H. Note

$$\begin{aligned} \langle u, v \rangle &= real(\langle u, v \rangle) + i \, imag(\langle u, v \rangle) = \langle u, v \rangle_{\mathbb{R}} - i \langle iu, v \rangle_{\mathbb{R}} \\ &= \langle u, v \rangle_{\mathbb{R}} + i \langle u, iv \rangle_{\mathbb{R}} = \langle \mathbf{j}(u), \mathbf{j}(v) \rangle + i \langle \mathbf{j}(u), J\,\mathbf{j}(v) \rangle. \end{aligned} \tag{2.19}$$

If $\{e_1, \cdots, e_n\}$ is an orthonormal basis in H, then $\{\mathbf{j}(e_1), \cdots, \mathbf{j}(e_n), J\mathbf{j}(e_1), \cdots, J\mathbf{j}(e_n)\}$ is an orthonormal basis in $H_{\mathbb{R}}$. Which shows the real dimension of $H_{\mathbb{R}}$ is $2n$, $\dim_{\mathbb{R}} H_{\mathbb{R}} = 2n$.

On $H \times H$ there are two inner product structures:

$$\langle (x, y), (u, v) \rangle_{\mathbb{C}} = \langle x, u \rangle + \langle y, v \rangle \tag{2.20}$$

$$\langle (x, y), (u, v) \rangle_{\mathbb{R}} = \frac{1}{2} \left(\langle (x, y), (u, v) \rangle + \langle (u, v), (x, y) \rangle \right) = \langle x, u \rangle_{\mathbb{R}} + \langle y, v \rangle_{\mathbb{R}} \tag{2.21}$$

With respect to $\langle \cdot, \cdot \rangle_{\mathbb{C}}$, $H \times H$ is a $\mathbb{C}$-vector space of dimension $2n$. With respect to $\langle \cdot, \cdot \rangle_{\mathbb{R}}$, $H \times H$ is a $\mathbb{R}$-vector space of dimension $4n$. On $H_{\mathbb{R}}$ the two inner products coincide. Because of this fact we shall simply denote $\langle \xi, \eta \rangle_{H_{\mathbb{R}}}$ or $\langle \xi, \eta \rangle$ the scalar product on $H \times H$ whenever $\xi, \eta \in H_{\mathbb{R}}$. Furthermore $\mathbf{j}$ is norm preserving $\|x\|_H = \|\mathbf{j}(x)\|_{H_{\mathbb{R}}}$ and $\langle x, y \rangle_{\mathbb{R}} = \langle \mathbf{j}(x), \mathbf{j}(y) \rangle_{H_{\mathbb{R}}}$ for all $x, y \in H$. The orthogonal complement of $H_{\mathbb{R}}$ in $H \times H$ with respect to $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ is given by

$$H_{\mathbb{R}}^{\perp} = iH_{\mathbb{R}} = \left\{ \left(\frac{i}{2}(x + c(x)), \frac{1}{2}(x - c(x)) \right) \ , \ x \in H \right\}. \tag{2.22}$$

The orthogonal projection onto $H_{\mathbb{R}}$ with respect to the real structure is given by

$$P_{\mathbb{R}} : H \times H \to H_{\mathbb{R}} \subset H \times H \ , \ P_{\mathbb{R}}(u, v) = \left(\frac{1}{2}(u + c(u)), \frac{1}{2}(v + c(v)) \right). \tag{2.23}$$

The projection onto the orthogonal complement $H_{\mathbb{R}}^{\perp}$ is given by

(2.24)
$$P_{\mathbb{R}}^{\perp} = 1 - P_{\mathbb{R}} \ , \ P_{\mathbb{R}}^{\perp}(u,v) = \left(\frac{1}{2}(u-c(u)), \frac{1}{2}(v-c(v))\right) = i\left(\frac{1}{2i}(u-c(u)), \frac{1}{2i}(v-c(v))\right) .$$

Note:

$$P_{\mathbb{R}}^{\perp}(u,v) = -iP_{\mathbb{R}}(iu,iv) \ . \tag{2.25}$$

Fix $a, b \in H$ and define the linear operator on H,

$$T_{a,b} : H \to H \ , \ T_{a,b}(x) = \langle x, a\rangle b \ . \tag{2.26}$$

Associate the following $\mathbb{R}$-linear operator on $H \times H$

$$\tilde{T}_{a,b} : H \times H \to H_{\mathbb{R}} \subset H \times H \ , \ \tilde{T}_{a,b}(u) = \langle u, \mathbf{j}(a)\rangle_{\mathbb{R}} \mathbf{j}(b) + \langle u, \mathbf{j}(ia)\rangle_{\mathbb{R}} \mathbf{j}(ib) \ . \tag{2.27}$$

Note $\langle \mathbf{j}(b), \mathbf{j}(ib)\rangle_{\mathbb{R}} = 0$ and $H_{\mathbb{R}}$ is invariant under the action of $T_{a,b}$. Direct computations show the following diagram is commutative:

$$\begin{array}{rcccc} T_{a,b} & : & H & \stackrel{T_{a,b}}{\longrightarrow} & H \\ & & \mathbf{j}\downarrow & & \downarrow \mathbf{j} \\ \tilde{T}_{a,b} & : & H_{\mathbb{R}} & \stackrel{\tilde{T}_{a,b}}{\longrightarrow} & H_{\mathbb{R}} \end{array} \tag{2.28}$$

Similarly each symmetric operator T in $Sym(H)$ gets mapped into a symmetric operator in $Sym(H_{\mathbb{R}})$ of double rank. Denote by τ this mapping, $\tau : Sym(H) \to Sym(H_{\mathbb{R}})$ that makes the following diagram commutative:

$$\begin{array}{ccc} H & \stackrel{T}{\longrightarrow} & H \\ \mathbf{j}\downarrow & & \downarrow \mathbf{j} \\ H_{\mathbb{R}} & \stackrel{\tau(T)}{\longrightarrow} & H_{\mathbb{R}} \end{array} \tag{2.29}$$

If desired, $\tau(T)$ can be extended to $H \times H$ using the $\mathbb{R}$-linear scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ on $H \times H$. τ is $\mathbb{R}$-linear but not $\mathbb{C}$-linear. In particular:

$$T = [\![x, x]\!] \in \mathcal{S}^{1,0} \implies \tau(T) = [\![\mathbf{j}(x), \mathbf{j}(x)]\!] + [\![\mathbf{j}(ix), \mathbf{j}(ix)]\!] \in \mathcal{S}^{2,0}(H_{\mathbb{R}}) \tag{2.30}$$

$$T = [\![x, y]\!] \in \mathcal{S}^{1,1} \implies \tau(T) = [\![\mathbf{j}(x), \mathbf{j}(y)]\!] + [\![\mathbf{j}(ix), \mathbf{j}(iy)]\!] \in \mathcal{S}^{2,2}(H_{\mathbb{R}}) \tag{2.31}$$

Using the $\mathbb{R}$-linear operator J introduced in (2.16) the first relation above can be rewritten as

$$\tau(xx^*) = \xi\xi^* + J\xi\xi^*J^* \ , \text{ where } \xi = \mathbf{j}(x) \in H_{\mathbb{R}} \tag{2.32}$$

and the adjoints ξ^* and J^* are taken with respect to the scalar product of $H_{\mathbb{R}}$. In general an operator in $\mathcal{S}^{p,q}(H)$ gets mapped into an operator in $\mathcal{S}^{2p,2q}(H_{\mathbb{R}})$. Note the map τ preserves scalar products between selfadjoint operators up to a multiplicative constant:

$$\langle T, S\rangle_{B(H)} = 2\langle \tau(T), \tau(S)\rangle_{B(H_{\mathbb{R}})} \ . \tag{2.33}$$

Thus τ is a monomorphism (i.e. linear injective morphism) between the $\mathbb{R}$-vector spaces $Sym(H)$ and $Sym(H_{\mathbb{R}})$. We shall discuss in more details the linear map τ in subsection 3.2.

Note, in the case $H = \mathbb{C}^n$, $c(z) = (\bar{z}_1, \cdots, \bar{z}_n)^T$, and $\langle v, w\rangle = v^T c(w)$, the map $\mathbf{j}$ acts by $z \in \mathbb{C}^n \mapsto \mathbf{j}(z) = (real(z^T), imag(z^T))^T$. Here ^{T} denotes transposition. Thus $H_{\mathbb{R}} = \{(real(v^T), imag(v^T))^T \ , \ v \in \mathbb{C}^n\} = \mathbb{R}^{2n}$, and the $\mathbb{R}$-linear map J has the block form

$$J = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$$

where I is the identity matrix of size n. The scalar product in $H_{\mathbb{R}}$ is the usual real scalar product. For any vector $\xi \in H_{\mathbb{R}} = \mathbb{R}^{2n}$ the adjoint reduces to transposition $\xi^* = \xi^T$.

We return to the frame set $\mathbb{F} = \{f_1, \cdots, f_m\}$ and the Hilbert space H. We let

$$\Phi_k = \tau(f_k f_k^*) = \varphi_k \varphi_k^* + J\varphi_k \varphi_k^* J^* \tag{2.34}$$

denote the image of $F_k = [\![f_k, f_k]\!]$ under τ, where $\varphi_k = \mathbf{j}(f_k)$ is an element of $H_{\mathbb{R}}$.

The last notation we introduce here is the following map on $H_{\mathbb{R}}$:

$$R : H_{\mathbb{R}} \to Sym(H_{\mathbb{R}}) \ , \ R(\xi) = \sum_{k=1}^{m} \Phi_k \xi \xi^* \Phi_k. \tag{2.35}$$

As we will see later $R(\xi)$ is related to the Fisher information matrix for the measurement model (2.3), and it plays a key role in obtaining a necessary and sufficient condition of injectivity for the nonlinear map α. More explicit

$$R(\xi) = \sum_{k=1}^{m} v_k v_k^* \ , v_k = \Phi_k \xi = \langle \xi, \varphi_k\rangle \varphi_k + \langle \xi, J\varphi_k\rangle J\varphi_k. \tag{2.36}$$

We shall not overload the notation and use the same letter R to denote the map $R : H \to Sym(H_{\mathbb{R}})$, defined by $x \mapsto R(\mathbf{j}(x))$. Finally, we let $\delta_{i,j}$ denote the Kroneker symbol: $\delta_{i,j} = 1$ if $i = j$, and 0 otherwise.

2.2. **Existing Results.** We revise now existing results on injectivity of the nonlinear map α. A subset Z of a topological space (X, τ) is said to be *generic* if its open interior is dense. In the following statements, the term *generic* refers to the Zarisky topology: a set $Z \subset \mathbb{K}^{n\times m} = \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ is said to be *generic* if Z is dense in $\mathbb{K}^{n\times m}$ and its complement is a finite union of zero sets of polynomials in nm variables with coefficients in the field $\mathbb{K}$ (here $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$).

Theorem 2.1. *In the real case when $H = \mathbb{R}^n$ the following are equivalent:*

1. *The nonlinear map α is injective;*
2. *([3], Th.2.8) For any disjoint partition of the frame set $\mathbb{F} = \mathbb{F}_1 \cup \mathbb{F}_2$, either $\mathbb{F}_1$ spans H or $\mathbb{F}_2$ spans H.*

(3) *([6], Th.2.4(2)) For any two vectors $x, y \in H$ if $x \neq 0$ and $y \neq 0$ then*

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 > 0$$

(4) *([6], Th.2.4(3)) There is a real constant $a_0 > 0$ so that for all $x, y \in H$,*

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 \geq a_0 \|x\|^2 \|y\|^2 \tag{2.37}$$

(5) *([6], Th.2.4(4)) There is a real constant $a_0 > 0$ so that for all $x \in H$,*

$$R(x) := \sum_{k=1}^{m} |\langle x, f_k\rangle|^2 \langle \cdot, f_k\rangle f_k \geq a_0 I \tag{2.38}$$

where the inequality is in the sense of quadratic forms.

Additionally, the following statements hold true:

(1) *([3], Prop2.5) If α is injective then $m \geq 2n-1$;*
(2) *([3], Prop.2.5) If $m \leq 2n-2$ then α cannot be injective;*
(3) *([3], Cor.2.7(1)) If $m = 2n-1$ then α is injective if and only if $\mathbb{F}$ is full spark;*
(4) *([3], Cor.2.6) If $m \geq 2n-1$ and $\mathbb{F}$ is full spark then the map α is injective;*
(5) *([3], Th.2.2) If $m \geq 2n-1$ then for a generic frame $\mathbb{F}$ the map α is injective.*

In the complex case the following results are known:

Theorem 2.2. *In the complex case when $H = \mathbb{C}^n$ the following statements hold true:*

(1) *([3], Th.3.3) If $m \geq 4n-2$ then for a generic frame $\mathbb{F}$ the map α is injective;*
(2) *([12]) For any positive integer n there is a frame with $m = 4n-4$ elements so that the nonlinear map α is injective;*
(3) *([25], Corollary 4) If α is injective then*

$$m \geq 4n - 2 - 2\beta + \begin{cases} 2 & \text{if } n \text{ odd and } \beta = 3 \bmod 4 \\ 1 & \text{if } n \text{ odd and } \beta = 2 \bmod 4 \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

where $\beta = \beta(n)$ denotes the number of 1's in the binary expansion of $n-1$.

(4) *The following are equivalent:*
   (a) *The map α is injective;*
   (b) *([25] Prop. 2)*

$$ker(\mathcal{A}) \cap \left(\mathcal{S}^{1,0} - \mathcal{S}^{1,0}\right) = \{0\} \tag{2.40}$$

   (c) *([8], Theorem 4) $dim\, S(u) \geq 2n-1$ for every $u \in \mathbb{C}^n \setminus \{0\}$;*
   (d) *([8], Theorem 4) $S(u) = span_{\mathbb{R}}(iu)^{\perp}$ for every $u \in \mathbb{C}^n \setminus \{0\}$.*
   (e) *([19][1], Theorem 1.1) If $m \geq 4n-4$ then for a generic frame $\mathbb{F}$ the map α is injective;*

[1]This result was not known at the time the present paper was submitted for publication.

In the last two conditions, $S(u) = span_{\mathbb{R}}\{f_k f_k^* u\}_{1\le k\le m}$, where $f_k f_k^* u$ is seen as a $2n$ vector in $\mathbb{R}^{2n}$.

3. New Analysis Results

This section contains our injectivity results of the nonlinear map α as well as an in-depth analysis of the spaces $\mathcal{S}^{p,q}$.

Theorem 3.1. *Let H be a $\mathbb{C}$-vector space of dimension n, with scalar product $\langle,\rangle$ and conjugation $c : H \to H$. The following are equivalent:*

(1) *The nonlinear map $\alpha : \hat{H} \to \mathbb{R}^m$, $(\alpha(x))_k = |\langle x, f_k\rangle|^2$ is injective.*
(2) *There is a constant $a_0 > 0$ so that for every $u, v \in H$*

$$\sum_{k=1}^{m} |\langle F_k, [\![u, v]\!]\rangle|^2 \ge a_0 \|[\![u, v]\!]\|_1^2 \tag{3.41}$$

where $F_k = f_k f_k^*$. Explicitly, this means:

$$\sum_{k=1}^{m} (real(\langle u, f_k\rangle\langle f_k, v\rangle))^2 \ge a_0 \left[\|u\|^2\|v\|^2 - (imag(\langle u, v\rangle))^2\right]. \tag{3.42}$$

(3) *For any $\xi \in H_{\mathbb{R}}$, $\xi \neq 0$, $rank(R(\xi)) = 2n - 1$.*
(4) *There is $a_0 > 0$ so that for all $\xi \in H_{\mathbb{R}}$, $\xi \neq 0$,*

$$R(\xi) \ge a_0 \|\xi\|^2 P_{J\xi}^{\perp} \tag{3.43}$$

where the inequality holds in the sense of quadratic forms in $H_{\mathbb{R}}$, and

$$P_{J\xi}^{\perp} = 1 - \frac{1}{\|\xi\|^2} J\xi\xi^* J^* \tag{3.44}$$

is the orthogonal projection in $H_{\mathbb{R}}$ onto the orthogonal complement of $J\xi$ in $H_{\mathbb{R}}$.

The proof is given in section 3.3.

Remark 3.2. *The two constants a_0 in (2) and (4) can be chosen to be equal, hence the same notation. We will see in the next section, this common constant is related to robustness and stability of any reconstruction scheme.*

Remark 3.3. *The proof of (3) $\Leftrightarrow$ (4) shows that the optimal bound a_0 is given by*

$$a_0^{opt} = min_{\xi\in H_{\mathbb{R}}, \|\xi\|=1} a_{2n-1}(R(\xi)) \tag{3.45}$$

where $a_{2n-1}(R(\xi))$ denotes the next to the smallest eigenvalue of $R(\xi)$.

Remark 3.4. *The choice of the nuclear norm and the square in (3.41) is somewhat arbitrary. For any $p, q \ge 1$ (including infinity), there is a constant $a_{p,q} > 0$ so that*

$$\sum_{k=1}^{m} |\langle F_k, [\![u, v]\!]\rangle|^p \ge a_{p,q} \|[\![u, v]\!]\|_q^p. \tag{3.46}$$

An interesting corollary, which follows, is obtained in the case when α is restricted to a subspace of H. It turns out that sometimes the underlying signal is actually real. The canonical description of such a condition is to be conjugation invariant. Let H' denote this set:

$$H' = \{x \in H \ ; \ c(x) = x\}. \tag{3.47}$$

Note H' is a $\mathbb{R}$-linear space, but it is not a $\mathbb{C}$-linear space. Since x is restricted to H' it follows that the equivalence class in H' of x is given by $\hat{x} \cap H' = \{x, -x\}$. Consequently the appropriate quotient space is given by $\widehat{H'} = \{\{x, -x\} \ , \ x \in H_{\mathbb{R}}\}$. Let $\pi_1 : H \times H \to H$ and $\pi_2 : H \times H \to H$ be the canonical projections onto factors: $\pi_1((u, v)) = u$ and $\pi_2((u, v)) = v$. Then it is immediate to check that H' admits the following equivalent descriptions:

$$H' = \{x \in H \ ; \ \pi_2(\mathbf{j}(x)) = 0\} = \mathbf{j}^{-1}\left(\{H_{\mathbb{R}} \cap (H \times \{0\})\right). \tag{3.48}$$

Note in H', $\langle u, v\rangle$ is always real for all $u, v \in H'$ since $\langle u, v\rangle = \langle c(v), c(u)\rangle = \langle v, u\rangle$. Let $\mathbb{F} = \{f_1, \cdots, f_m\}$ be the frame set in H. Note we do not assume $\mathbb{F} \subset H'$. Let

$$g_k = \pi_1(\mathbf{j}(f_k)) \ , \ h_k = \pi_2(\mathbf{j}(f_k)) \tag{3.49}$$

with $1 \leq k \leq m$. Note $f_k = \mathbf{j}^{-1}((g_k, h_k)) = g_k + ih_k$ and $g_k, h_k \in H'$. Set:

$$\Phi'_k = g_k g_k^* + h_k h_k^* \in \mathcal{S}^{2,0}(H') \subset \mathcal{S}^{2,0}(H) \ , \ 1 \leq k \leq m. \tag{3.50}$$

Note:

$$\langle F_k, [\![x, x]\!]\rangle \ = \ |\langle f_k, x\rangle|^2 = \langle {\Phi'}_k x, x\rangle = \langle {\Phi'}_k, [\![x, x]\!]\rangle \tag{3.51}$$

$$\langle F_k, [\![u, v]\!]\rangle \ = \ real(\langle u, f_k\rangle\langle f_k, v\rangle) = \langle {\Phi'}_k u, v\rangle = \langle {\Phi'}_k, [\![u, v]\!]\rangle \tag{3.52}$$

for all $x, u, v \in H'$. Thus the linear map $\mathcal{A}$ restricted to $Sym(H')$ can be thought of as taking inner products with a family of rank-2 nonnegative operators. We have

Corollary 3.5. *Let H be an n-dimensional complex Hilbert space with scalar product $\langle , \rangle$ and conjugation $c : H \to H$. Let $H' = \{x \in H \ ; \ c(x) = x\}$ be the maximal c-invariant set. The following are equivalent:*

1. *The restriction to H' of the nonlinear map $\alpha|_{H'}$ is injective on $\widehat{H'}$.*
2. *There is a constant $a_0 > 0$ so that $\forall u, v \in H'$*

$$\sum_{k=1}^{m} |\langle {\Phi'}_k u, v\rangle|^2 \geq a_0 \|u\|^2 \|v\|^2. \tag{3.53}$$

3. *For all $u \neq 0$,*

$$\dim_{\mathbb{R}} span_{\mathbb{R}}\{\langle u, g_k\rangle g_k + \langle u, h_k\rangle h_k \ ; \ 1 \leq k \leq m\} = n. \tag{3.54}$$

4. *For any $u \neq 0$,*

$$rank\left(\sum_{k=1}^{m} {\Phi'}_k u u^* {\Phi'}_k\right) = n. \tag{3.55}$$

(5) *Let* $R'(u) = \sum_{k=1}^{m} \Phi'_k uu^* \Phi'_k$. *There is a constant* $a_0 > 0$ *so that* $\forall u, v \in H'$,

$$\langle R'(u)v, v\rangle \geq a_0 \|u\|^2 \|v\|^2. \tag{3.56}$$

The proof is given in section 3.3.

3.1. **Analyis of sets $\mathcal{S}^{p,q}$.** In addition to (2.11) the sets $\mathcal{S}^{p,q}$ introduced in (2.7) have the following properties summarized in the following lemma:

Lemma 3.6. (1) *For any* $p_1 \leq p_2$ *and* $q_1 \leq q_2$, $\mathcal{S}^{p_1,q_1} \subseteq \mathcal{S}^{p_2,q_2}$

(2) *For any nonnegative integers* p, q *the following disjoint decomposition holds true*

$$\mathcal{S}^{p,q} = \bigcup_{r=0}^{p} \bigcup_{s=0}^{q} \mathring{\mathcal{S}}^{r,s} \tag{3.57}$$

where by convention $\mathring{\mathcal{S}}^{0,0} = \mathcal{S}^{0,0} = \{0\}$, *and* $\mathring{\mathcal{S}}^{p,q} = \emptyset$ *for* $p + q > n$.

(3) *For any nonnegative integers* p, q,

$$-\mathcal{S}^{p,q} = \mathcal{S}^{q,p} \tag{3.58}$$

(4) *The mapping* $(T, X) \mapsto TXT^*$ *defines an action of* $B(H)$ *on* $\mathcal{S}^{p,q}$. *Specifically for any* $T \in B(H)$ *and integers* p, q,

$$T\mathcal{S}^{p,q}T^* \subseteq \mathcal{S}^{p,q} \tag{3.59}$$

The inclusion becomes equality if T *is invertible.*

(5) $GL(H)$ *acts transitively on* $\mathring{\mathcal{S}}^{p,q}$. *Specifically for any* $X, Y \in \mathring{\mathcal{S}}^{p,q}$ *there is an invertible* $T \in GL(H)$ *so that* $Y = TXT^*$.

(6) *For any integers* p, q, r, s,

$$\mathcal{S}^{p,q} + \mathcal{S}^{r,s} = \mathcal{S}^{p,q} - \mathcal{S}^{s,r} = \mathcal{S}^{p+r,q+s} \tag{3.60}$$

Proof of Lemma 3.6

First three assertions are trivial.

(4) Fix an orthonormal basis in H, fix a $T \in B(H)$ and let $T = UDV$ be its singular value decomposition, where U, V are unitary operators on H and D is a diagonal operator with non-negative entries. Let $X \in \mathcal{S}^{p,q}$ and set $R(t) = U(tD+(1-t)I)VXV^*(tD+(1-t)I)U^*$ for every $0 \leq t \leq 1$, where I denotes the identity operator on H. Note $R(0) = UVXV^*U^* \in \mathcal{S}^{p,q}$ and $R(1) = TXT^*$. For every $0 \leq t < 1$, the operator $U(tD + (1-t)I)V$ is invertible. Then by Sylvester's Law of Inertia (see Ex. 12.43 in Chapter 12 of [30]), for every $0 \leq t < 1$ the operator $R(t)$ has the same number of strictly positive eigenvalues and strictly negative eigenvalues as X does. Since the spectrum is continuous with respect to matrix entries, it follows the number of strictly positive eigenvalues cannot increase when passing to limit $t \to 1$. Same conclusion holds for the number of strictly negative eigenvalues. This shows $TXT^* = R(1) \in \mathcal{S}^{p,q}$. Finally, when T is invertible, TXT^* has the same number of strictly positive (negative) eigenvalues as X does. This shows $T\mathcal{S}^{p,q}T^* = \mathcal{S}^{p,q}$.

(5) The conclusion follows again from Sylvester's Law of Intertia. Indeed fix an orthonormal basis in H and let $T_1, T_2 \in GL(H)$ be invertible transformations so that both $T_1 X T_1^*$ and

$T_2YT_2^*$ have the same matrix representations that are diagonal with $+1$ repeated p times, -1 repeated q times and 0 repeated $n-p-q$ times. Thus $T_1XT_1^* = T_2YT_2^*$ from where the conclusion follows with $T = T_2^{-1}T_1$.

(6) One can see this statement as a special instance of the Witt decomposition theorem, a much more powerful tool in the theory of quadratic forms especially in the case of vector spaces over fields of charactersitics other than 2, see [27]. However for the benefit of those who prefer a more direct proof, here is a sketch of this result. Using spectral decomposition and rearranging the term, one can easily see that $\mathcal{S}^{p+r,q+s} \subset \mathcal{S}^{p,q} + \mathcal{S}^{r,s}$. For the converse inclusion, we need to show that if $T, S \in Sym(H)$ are so that T has at most p positive eigenvalues and q negative eigenvalues, and S has at most r positive eigenvalues and s negative eigenvalues, then $T+S$ has at most $p+r$ positive eigenvalues, and $q+s$ negative eigenvalues. Without loss of generality we can assume $T \in \mathring{\mathcal{S}}^{p,q}$ and $S \in \mathring{\mathcal{S}}^{r,s}$. Using spectral decompostions of T and S we obtain

$$T+S = \left(\sum_{k=1}^{p} a_k A_k - \sum_{k=p+1}^{p+q} a_k A_k\right) + \left(\sum_{k=1}^{r} b_k B_k - \sum_{k=r+1}^{r+s} b_k B_k\right) = U - V \ ,$$

$$U = \sum_{k=1}^{p} a_k A_k + \sum_{k=1}^{r} b_k B_k \geq 0 \ , \ V = \sum_{k=p+1}^{q} a_k A_k + \sum_{k=r+1}^{r+s} b_k B_k \geq 0 \ ,$$

where A_k, B_k are rank one orthogonal (spectral) projectors with $A_k A_j = 0$ and $B_l B_h = 0$ for all $k \neq j$ and $l \neq h$, and $a_k, b_k > 0$. Thus $U \in \mathcal{S}^{p+r,0}$ and $V \in \mathcal{S}^{q+s,0}$. The claim now follows by induction provided we show that for any $R \in \mathring{\mathcal{S}}^{a,b}$ and $E \in \mathcal{S}^{1,0}$, a rank one associated to vector $e \in H$ ($E = \langle \cdot, e\rangle e$), then $E + R \in \mathcal{S}^{a+1,b}$.

Indeed this last assertion is shown as follows. Let $R = \sum_{k=1}^{a} c_k \langle \cdot, g_k\rangle g_k - \sum_{k=a+1}^{a+b} c_k \langle \cdot, g_k\rangle g_k$ be its spectral decomposition. Two cases are treated distinctly.

Case 1: If $e \notin R(H)$ then $\Gamma' = \{g_1, \cdots, g_{a+b}, e\}$ is linearly independent. Let $\Gamma = \{\gamma_k \ , \ 1 \leq k \leq a+b+1\}$ be the (unique) biorthogonal system to Γ'. Let T be an invertible operator that maps an orthonormal set $\Delta = \{\delta_1, \cdots, \delta_{a+b+1}\}$ into Γ, $T\delta_k = \gamma_k$, $1 \leq k \leq a+b+1$, and maps the orthogonal space to Δ onto the orthogonal complement to Γ. Note $T^* g_k = \delta_k$, $1 \leq k \leq a+b$ and $T^* e = \delta_{a+b+1}$. Then a direct computation shows that

$$T^*(R+E)T = \sum_{k=1}^{a+b} c_k \langle \cdot, \delta_k\rangle \delta_k + \langle \cdot, \delta_{a+b+1}\rangle \delta_{a+b+1}.$$

Thus the spectrum of $T^*(R+E)T$ is composed of $\{c_1, \cdots, c_a, -c_{a+1}, \cdots, c_{a+b}, 1\}$ which shows that $T^*(R+E)T \in \mathcal{S}^{a+1,b}$, and by (4), $R+E \in \mathcal{S}^{a+1,b}$.

Case 2: $e \in R(H)$. The rank of $R+E$ is less than or equal to the rank of R. Hence $R+E \in \mathcal{S}^{a',b'}$ with $a'+b' = a+b$. Now by the continuity of spectrum with respect to small perturbations, it follows that, for a small perturbation $e \mapsto e' \notin R(H)$, $R+E' \in \mathcal{S}^{a'',b''}$ with $a'' \geq a', b'' \geq b'$. But the proof of case 1 shows $\mathcal{S}^{a'',b''} \subset \mathcal{S}^{a+1,b}$. Hence $b \geq b'' \geq b' \geq b$, and $a+1 \geq a'' \geq a' \geq a$. Thus $R+E \in \mathcal{S}^{a+1,b}$. $\square$

Next we analyze the space $\mathcal{S}^{1,1}$ in more detail. The special factorization we obtain here can be extended to other classes $\mathcal{S}^{p,q}$ but we do not plan to do so here. First set the following matrices:

$$K = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\ ,\ D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\ ,\ V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{3.61}$$

In the next result we use a generalized unitary group $U(1,1;K)$. Recall its definition.

Definition The groups $U(1,1)$ and $U(1,1;K)$ are defined by

$$U(1,1) = \{A \in \mathbb{C}^{2\times 2}\ ,\ A^*DA = D\} \tag{3.62}$$

$$U(1,1;K) = \{A \in \mathbb{C}^{2\times 2}\ ,\ A^*KA = K\} \tag{3.63}$$

These groups have been studied extensively in literature. See for instance [32], section 10.4. In particular the two groups above are unitarily equivalent to each other, and the matrix V provides such an equivalence:

$$A \in U(1,1;K) \Leftrightarrow B = VAV^* \in U(1,1). \tag{3.64}$$

The quadratic form $\omega(z_1,z_2) = |z_1|^2 - |z_2|^2$ is invariant under the action of $U(1,1)$, whereas $\phi(z_1,z_2) = \bar{z}_1 z_2 + z_1 \bar{z}_2$ is invariant under the action of $U(1,1;K)$.

Lemma 3.7.

(1) $\mathcal{S}^{1,1} = \mathcal{S}^{1,0} - \mathcal{S}^{1,0} = \mathcal{S}^{1,0} + \mathcal{S}^{0,1}$.

(2) *For any $T \in \mathcal{S}^{1,1}$ there are $u, v \in H$ so that*

$$T = \frac{1}{2}(uv^* + vu^*) = [\![u,v]\!]. \tag{3.65}$$

If $T = a_1 e_1 e_1^* - a_2 e_2 e_2^*$ with $a_1, a_2 \geq 0$ and $\langle e_k, e_j\rangle = \delta_{k,j}$ is its spectral factorization then

$$u_0 = \sqrt{a_1}e_1 + \sqrt{a_2}e_2,\ v_0 = \sqrt{a_1}e_1 - \sqrt{a_2}e_2 \tag{3.66}$$

provides a particular factorization in (3.65).

Lemma 3.8.

(1) *Let $T = [\![u,v]\!]$. Then traces and spectrum $Sp(T) = \{a_+, a_-\}$ are given by*

$$tr\{T\} = real(\langle u,v\rangle) = \langle u,v\rangle_{\mathbb{R}} \tag{3.67}$$

$$\begin{aligned} tr\{T^2\} &= \frac{1}{4}((\langle u,v\rangle)^2 + (\langle v,u\rangle)^2 + 2\|u\|^2\|v\|^2) \\ &= \frac{1}{2}\left(\|u\|^2\|v\|^2 + \langle u,v\rangle_{\mathbb{R}}^2 - \langle iu,v\rangle_{\mathbb{R}}^2\right) \end{aligned} \tag{3.68}$$

$$a_+ = \frac{1}{2}\left(\langle u,v\rangle_{\mathbb{R}} + \sqrt{\|u\|^2\|v\|^2 - \langle iu,v\rangle_{\mathbb{R}}^2}\right) \geq 0 \tag{3.69}$$

$$a_- = \frac{1}{2}\left(\langle u,v\rangle_{\mathbb{R}} - \sqrt{\|u\|^2\|v\|^2 - \langle iu,v\rangle_{\mathbb{R}}^2}\right) \leq 0 \tag{3.70}$$

The nuclear norm of T is given by

$$\|T\|_1 = a_+ + |a_-| = \sqrt{\|u\|^2\|v\|^2 - \langle iu, v\rangle_{\mathbb{R}}^2}. \tag{3.71}$$

Hence $T \in \mathcal{S}^{1,1}$.

(2) *Let $T = [\![u_0, v_0]\!] \in \mathring{\mathcal{S}}^{1,1}$. Then any pair (u, v) of vectors, with $u, v \in H$ so that $T = [\![u, v]\!]$ is given by*

$$u = a_{11}u_0 + a_{12}v_0 \ , \ v = a_{21}u_0 + a_{22}v_0 \tag{3.72}$$

for some matrix $A = (a_{k,l})_{1\leq k,l\leq 2}$ with $A \in U(1,1;K)$. Conversely, for any matrix $A \in U(1,1;K)$, $[\![u, v]\!] = [\![u_0, v_0]\!]$ where (u, v) are given by (3.72).

Lemma 3.9.

(1) *Let $T = xx^* - yy^*$ for some $x, y \in H$. Then $T \in \mathcal{S}^{1,1}$ with spectrum $Sp(T) = \{b_+, b_-\}$ and traces given by*

$$tr\{T\} = \|x\|^2 - \|y\|^2 \tag{3.73}$$

$$tr\{T^2\} = \|x\|^4 + \|y\|^4 - 2|\langle x, y\rangle|^2 \tag{3.74}$$

$$b_\pm = \frac{1}{2}\left(\|x\|^2 - \|y\|^2\right) \pm \frac{1}{2}\sqrt{(\|x\|^2 + \|y\|^2)^2 - 4|\langle x, y\rangle|^2} \tag{3.75}$$

$$\|T\|_1 = \sqrt{(\|x\|^2 + \|y\|^2)^2 - 4|\langle x, y\rangle|^2} \tag{3.76}$$

(2) *Let $T = xx^* - yy^* \in \mathcal{S}^{1,1}$. Any pair of vectors (x', y'), with $x', y' \in H$ so that $T = x'(x')^* - y'(y')^*$ is related to (x, y) via*

$$x' = b_{11}x + b_{12}y \ , \ y' = b_{21}x + b_{22}y \tag{3.77}$$

for some matrix $B = (b_{ij})_{1\leq i,j\leq 2}$ with $B \in U(1,1)$. Conversely, for any matrix $B \in U(1,1)$, $x'(x')^* - y'(y')^* = xx^* - yy^*$, where x', y' are given by (3.77).

Remark 3.10. *1. The need for studying $\mathcal{S}^{1,1}$ arose from the behavior of the IRLS algorithm described in* [6]*. However, it quickly became apparent that the space $\mathcal{S}^{1,1}$ and its factorization given by (3.65) are crucial for understanding the injectivity of the nonlinear map α, especially in light of Theorem 2.2 (4), equation (2.40).*

2. The choice in (3.66) has the following two additional properties:

$$\|u_0\| = \|v_0\| = \sqrt{a_1 + a_2} = \sqrt{a_+ - a_-} = \sqrt{\|T\|_1} \tag{3.78}$$

$$\langle u_0, v_0\rangle = a_1 - a_2 = a_+ + a_- = tr\{T\} \text{ (a real number!)} \tag{3.79}$$

where $\|T\|_1$ represents the nuclear norm of T, and $a_1 = a_+ \geq 0$ and $a_2 = -a_- \geq 0$ are its singular eigenvalues.

Proof of Lemma 3.7

(1) is a direct application of Lemma 3.6(5).

(2) Since $[\![,]\!]$ is $\mathbb{R}$-linear and $[\![e_1, e_2]\!] = [\![e_2, e_1]\!]$ we obtain

$$[\![u_0, v_0]\!] = a_1e_1e_1^* + \sqrt{a_1a_2}[\![e_1, e_2]\!] - \sqrt{a_1a_2}[\![e_1, e_2]\!] - a_2e_2e_2^* = a_1e_1e_1^* - a_2e_2e_2^* = T. \qquad \square$$

Proof of Lemma 3.8

(1) The equation (3.67) comes from the definition (2.5) and the fact that $tr(vu^*) = \langle v, u\rangle$. For T^2 compute first

$$T^2 = \frac{1}{4}\left(\langle v,u\rangle vu^* + \|u\|^2 vv^* + \|v\|^2 uu^* + \langle u,v\rangle uv^*\right).$$

Then (3.68) follows from this relation and $real((\langle u,v\rangle)^2) = (real(\langle u,v\rangle))^2 - (imag(\langle u,v\rangle))^2$. Finally, (3.69) and (3.70) come from solving:

$$\begin{array}{rcl} a_+ + a_- &=& tr(T) \\ a_+^2 + a_-^2 &=& tr(T^2) \end{array} \tag{3.80}$$

and observing

$$a_+a_- = \frac{1}{4}\left((\langle u,v\rangle_{\mathbb{R}})^2 + (\langle iu,v\rangle_{\mathbb{R}})^2 - \|u\|^2\|v\|^2\right) = \frac{1}{4}(|\langle u,v\rangle|^2 - \|u\|^2\|v\|^2) \le 0.$$

(2) A direct computation shows

$$\begin{aligned} [\![u,v]\!] = &\ \frac{1}{2}(\bar a_{11}a_{21} + \bar a_{21}a_{11})u_0u_0^* + \frac{1}{2}(\bar a_{11}a_{22} + \bar a_{21}a_{12})v_0u_0^* \\ &+ \frac{1}{2}(\bar a_{12}a_{21} + \bar a_{22}a_{11})u_0v_0^* + \frac{1}{2}(\bar a_{12}a_{22} + \bar a_{22}a_{12})v_0v_0^*. \end{aligned}$$

Since $\{u_0, v_0\}$ are linearly independent, $[\![u,v]\!] = [\![u_0,v_0]\!]$ implies

$$\bar a_{11}a_{21} + \bar a_{21}a_{11} = 0\ ,\ \bar a_{11}a_{22} + \bar a_{21}a_{12} = 1$$

$$\bar a_{12}a_{21} + \bar a_{22}a_{11} = 1\ ,\ \bar a_{12}a_{22} + \bar a_{22}a_{12} = 0.$$

which corresponds to $A^*KA = K$. Hence $A \in U(1,1;K)$. Conversely, if $A \in U(1,1;K)$ then the above relations are satisfied which imply $[\![u,v]\!] = [\![u_0,v_0]\!]$. □

Proof of Lemma 3.9

Claims (1) and (2) are similar to claims in lemma 3.8 and follow by direct computation. □

Topologically, $\mathcal{S}^{1,0}$ and $\mathcal{S}^{1,1}$ are not differentiable manifolds. Instead they are algebraic varieties since they are given by the zero loci of certain polynomials. We have the following result:

Lemma 3.11. *(1) The set $\mathring{\mathcal{S}}^{1,0}$ is an analytic manifold in $B(H)$ of real dimension $2n-1$. As a real manifold, its tangent space at $X = x_0x_0^*$ is given by*

$$\mathcal{T}_X\mathring{\mathcal{S}}^{1,0} = \{[\![x_0,y]\!]\ ,\ y \in H\}. \tag{3.81}$$

The $\mathbb{R}$-linear embedding $H \mapsto T_X\mathring{\mathcal{S}}^{1,0}$ given by $y \mapsto \varphi_{x_0}(y) = [\![x_0,y]\!]$ has null space given by $\ker \varphi_x = \{iax_0\ ;\ a \in \mathbb{R}\}$.

(2) The set $\mathring{\mathcal{S}}^{1,1}$ is an analytic manifold of real dimension $4n-4$. As a real manifold, its tangent space at $X = [\![x_0,y_0]\!]$ is given by

$$\mathcal{T}_X\mathring{\mathcal{S}}^{1,1} = \{[\![x_0,u]\!] + [\![y_0,v]\!]\ ,\ u,v \in H\}. \tag{3.82}$$

The $\mathbb{R}$-linear embedding $H\times H\mapsto \mathcal{T}_X\mathring{\mathcal{S}}^{1,1}$ given by $(u,v)\mapsto \varphi_{x_0,y_0}(u,v)=[\![x_0,u]\!]+[\![y_0,v]\!]$ has null space given by $\ker\varphi_{x_0,y_0}=\{a(ix_0,0)+b(0,iy_0)+c(y_0,-x_0)+d(iy_0,ix_0)\ ,\ a,b,c,d\in\mathbb{R}\}$.

Proof of Lemma 3.11

Let $c_1,\cdots,c_n: Sym(H)\to\mathbb{R}$ be the coefficients of the characteristic polynomial:

$$det(sI-T)=s^n+c_1(T)s^{n-1}+c_2(T)s^{n-2}+\cdots+c_n(T).$$

with $c_1(T)=-tr(T)$ and $c_n(T)=(-1)^n det(T)$. Note that the c_j's are polynomials.

(1) The manifold structure can be shown as follows. First note that

$$S^+=\{S\in Sym(H)\ ;\ c_1(T)=-tr(S)<0\}$$

is an open subset of $Sym(H)$ and therefore a manifold of same real dimension as $Sym(H)$ (which is n^2). Next note

$$\mathring{\mathcal{S}}^{1,0}=c_2^{-1}(0)\cap\cdots\cap c_n^{-1}(0)\cap S^+.$$

Hence $\mathring{\mathcal{S}}^{1,0}$ is an algebraic variety. Next we obtain that $\mathring{\mathcal{S}}^{1,0}$ is a homogeneous space and hence a real analytic manifold. Indeed by Lemma 3.6(5), $GL(H)$ acts transitively on $\mathring{\mathcal{S}}^{1,0}$. Therefore it is sufficient to verify the stabilizer group is closed. Fix $\{e_1,e_2,\cdots,e_n\}$ an orthonormal basis in H and consider the rank-1 operator $X=e_1e_1^*$. The stabilizer group for X is given by invertible transformations T so that $Te_1=ze_1$ with $z\in\mathbb{C}$, $|z|=1$. With respect to the fixed orthonormal basis, the stabilizer is represented by the group of matrices of the form:

$$\mathbb{H}_X^{1,0}=\left\{\begin{bmatrix} e^{i\theta} & v\\ 0 & M\end{bmatrix}\ ,\ \theta\in[0,2\pi)\ ,\ v\in\mathbb{C}^{1\times(n-1)}\ ,\ M\in\mathbb{C}^{(n-1)\times(n-1)}\ ,\ det(M)\neq 0\right\}.$$

One can easily verify that $\mathbb{H}_X^{1,0}$ is a closed subset of $GL(n,\mathbb{C})$, the Lie group of $n\times n$ invertible complex matrices. Thus $\mathring{\mathcal{S}}^{1,0}$ is diffeomorphic to the analytic manifold $GL(n,\mathbb{C})/\mathbb{H}_X^{1,0}$.

Next we determine the tangent space. Let $X=x_0x_0^*\in\mathring{\mathcal{S}}^{1,0}$. We consider the set of all differentiable curves

$$\Upsilon=\left\{\gamma: I\to\mathring{\mathcal{S}}^{1,0}\ ,\ \gamma(0)=X, 0\in I\subset\mathbb{R}\text{ open interval}\right\},$$

passing through X. Then the tangent space to $\mathring{\mathcal{S}}^{1,0}$ at X is given by

$$\mathcal{T}_X\mathring{\mathcal{S}}^{1,0}=\left\{\frac{d}{dt}\gamma(t)|_{t=0}\ ;\ \gamma\in\Upsilon\right\}.$$

For each such curve, $\gamma: I\to\mathring{\mathcal{S}}^{1,0}$ there is a unique differentiable curve $x: I\to H$ such that $\gamma(t)=x(t)(x(t))^*$ with $x(0)=x_0$. In fact, locally,

$$x(t)=\frac{1}{\sqrt{\langle\gamma(t)(x_0),x_0\rangle}}\gamma(t)(x_0)$$

which shows $x(t)$ is differentiable. Then a direct computation shows

$$\frac{d}{dt}\gamma(t)|_{t=0}=[\![\dot{x}(0),x(0)]\!]+[\![x(0),\dot{x}(0)]\!]=2[\![x_0,\dot{x}(0)]\!],\ \text{for any }\dot{x}(0)\in H.$$

Since $\dot{x}(0)$ can be chosen arbitrarily in H, it follows the tangent space to $\mathring{\mathcal{S}}^{1,0}$ at X is given by (3.81). The real dimension of the $\mathbb{R}$-vector space $\mathcal{T}_X\mathring{\mathcal{S}}^{1,0}$ is $2n-1$ once we notice the kernel of the $\mathbb{R}$-linear map φ_{x_0} is one dimensional and given by the real span of ix_0.

(2) The algebraic variety structure is given by the intersection

$$\mathring{\mathcal{S}}^{1,1} = c_3^{-1}(0) \cap \cdots \cap c_n^{-1}(0) \cap S^{--}$$

where

$$S^{--} = \{S \in Sym(H) \ ; \ c_2(T) < 0\}.$$

Next we obtain that $\mathring{\mathcal{S}}^{1,1}$ is a homogeneous space and hence a real analytic manifold. Indeed by Lemma 3.6(5), $GL(H)$ acts transitively on $\mathring{\mathcal{S}}^{1,1}$. Therefore it is sufficient to verify the stabilizer group is closed. Fix $\{e_1, e_2, \cdots, e_n\}$ an orthonormal basis in H and consider the rank-2 operator $X = e_1e_1^* - e_2e_2^* \in \mathring{\mathcal{S}}^{1,1}$. The stabilizer group for X is given by invertible transformations T whose matrix representations are of the form

$$\mathbb{H}_X^{1,1} = \left\{ \begin{bmatrix} R & v \\ 0 & M \end{bmatrix} \ , \ R \in U(1,1) \ , \ v \in \mathbb{C}^{2\times(n-2)} \ , \ M \in \mathbb{C}^{(n-2)\times(n-2)} \, , \, det(M) \neq 0 \right\}$$

where $U(1,1)$ was introduced in (3.62). One can easily verify that $\mathbb{H}_X^{1,1}$ is a closed subset of $GL(n,\mathbb{C})$, the Lie group of $n \times n$ invertible complex matrices. Thus $\mathring{\mathcal{S}}^{1,1}$ is diffeomorphic to the analytic manifold $GL(n,\mathbb{C})/\mathbb{H}_X^{1,1}$.

Next we determine the tangent space. Fix $X \in \mathring{\mathcal{S}}^{1,1}$, $X = [\![x_0, y_0]\!]$ with $\{x_0, y_0\}$ linearly independent, and let

$$\Upsilon' = \{\gamma : I \to \mathring{\mathcal{S}}^{1,1} \ , \ 0 \in I \subset \mathbb{R} \text{ open interval} \ , \ \gamma(0) = X\}$$

be the set of differentiable curves passing through X. Then the tangent space to $\mathring{\mathcal{S}}^{1,1}$ at X is given by

$$\mathcal{T}_X\mathring{\mathcal{S}}^{1,1} = \left\{ \frac{d}{dt}\gamma(t)|_{t=0} \ ; \ \gamma \in \Upsilon' \right\}.$$

By Lemma 3.6 we know $\gamma(t) = [\![x(t), y(t)]\!]$ for some $x : I \to H$ and $y : I \to H$. Note these functions are not unique. However we can choose them to be given by the spectral factorization of $\gamma(t)$ via (3.66). Furthermore a direct application of holomorphic functional calculus (see section 148, Decomposition Theorem in [28]) shows that $T \to P_1 \in \mathring{\mathcal{S}}^{1,0}$ and $T \mapsto P_2 \in \mathring{\mathcal{S}}^{1,0}$ are analytic, where $T = P_1 - P_2$ is its spectral decomposition with $P_1P_2 = 0$. Hence $t \in I \mapsto P_1(t)$ and $t \in I \mapsto P_2(t)$ are differentiable. The component functions $x(t)$ and $y(t)$ now uniquely defined by:

$$x(t) = \frac{1}{\sqrt{\langle P_1(t)(x(0)), x(0)\rangle}} P_1(t)(x(0)) \ , \ y(t) = \frac{1}{\sqrt{\langle P_2(t)(y(0)), y(0)\rangle}} P_2(t)(y(0)) \tag{3.83}$$

are differentiable. The derivative at $t = 0$ is given by

$$\frac{d}{dt}\gamma(t)|_{t=0} = 2[\![x(0), \dot{y}(0)]\!] + 2[\![y(0), \dot{x}(0)]\!]$$

And since $(x(0), y(0))$ and (x_0, y_0) are related by a $U(1,1;H)$ matrix, and the fact that $(\dot{x}(0), \dot{y}(0))$ is arbitrary in $H \times H$, we obtain the tangent space is given by (3.82), that is:

$$\mathcal{T}_X \mathring{\mathcal{S}}^{1,1} = \{\varphi_{x_0,y_0}(u,v) = [\![x_0, u]\!] + [\![y_0, v]\!] \ , \ u, v \in H\}.$$

A direct computation shows that $(ix_0, 0)$, $(0, iy_0)$, $(y_0, -x_0)$ and (iy_0, ix_0) are the only independent vectors in the null space of the $\mathbb{R}$-linear map $(u,v) \mapsto \varphi_{x_0,y_0}(u,v)$ which implies $\dim_{\mathbb{R}} \mathcal{T}_X \mathring{\mathcal{S}}^{1,1} = 4n-4$. Hence the real dimension of the manifold $\mathring{\mathcal{S}}^{1,1}$ is $4n-4$. □

Remark 3.12. *Compared to the complex projective manifold $\mathbb{CP}^{n-1} = \mathbb{P}(\mathbb{C}^n)$, $\mathring{\mathcal{S}}^{1,0}$ is diffeomorphic to $\mathbb{CP}^{n-1} \times \mathbb{R}^+$. The extra $\mathbb{R}^+$ component comes from the fact that rank-1 operators in $\mathring{\mathcal{S}}^{1,0}$ have arbitrary trace. This explains the real dimension of $\mathring{\mathcal{S}}^{1,0}$: $2\dim_{\mathbb{R}} \mathbb{CP}^n + 1 = 2(n-1) + 1 = 2n - 1$. On the other hand, using spectral factorization, $\mathring{\mathcal{S}}^{1,1}$ has real dimension given by $\dim_{\mathbb{R}} \mathring{\mathcal{S}}^{1,0} + \dim_{\mathbb{R}} \mathring{\mathcal{S}}^{1,0} - 2 = 2(2n-1) - 2 = 4n - 4$. The -2 term comes from the orthogonality of the two eigenvectors of an operator in $\mathring{\mathcal{S}}^{1,1}$. The $4n-4$ dimension estimate has been derived also heuristically in* [8] *right after proof of Lemma 9 (*[31]*).*

Remark 3.13. *As suggested by Bernhard Bodmann* [11]*, it can be shown that the subset of projections inside $\mathring{\mathcal{S}}^{1,0}$ is in fact a Kähler manifold (diffeomorphic to $\mathbb{CP}^{n-1}$). However $\mathring{\mathcal{S}}^{1,0}$ is not a Kähler manifold.*

3.2. **Analysis of the linear map τ .** We introduced earlier the $\mathbb{R}$-linear map τ that maps $Sym(H)$ operators into $Sym(H_{\mathbb{R}})$ operators, using the real linear structure on these spaces. In order for the diagram (2.29) to be commutative, the map $\tau(T)$ is given explicitly by

$$\tau(T)(\xi) = \mathbf{j}(T(\mathbf{j}^{-1}(\xi)) \ , \ \xi \in H_{\mathbb{R}} \tag{3.84}$$

The following lemma summarizes the basic properties of the map τ.

Lemma 3.14. (1) *Let P be an orthogonal projection of rank k in H. Then $\tau(P)$ is an orthogonal projection of rank $2k$ in $H_{\mathbb{R}}$. Furthermore if $\{e_1, \cdots, e_k\}$ is an orthonormal basis in the range of P, then $\{\mathbf{j}(e_1), \cdots, \mathbf{j}(e_k), J\mathbf{j}(e_1), \cdots, J\mathbf{j}(e_k)\}$ is an orthonormal basis in the range of $\tau(P)$.*

(2) *If $T \in Sym(H)$ has spectrum $(a_1, a_2, \cdots, a_n)$ then $\tau(T)$ in $Sym(H_{\mathbb{R}})$ has spectrum $(a_1, a_1, a_2, a_2, \cdots, a_n, a_n)$.*

(3) *For any two operators $T, S \in Sym(H)$, $\tau(T), \tau(S) \in Sym(H_{\mathbb{R}})$ and*

$$tr\{\tau(T)\tau(S)\} = \langle \tau(T), \tau(S) \rangle_{B(H_{\mathbb{R}})} = 2\langle T, S \rangle_{B(H)} = 2tr\{TS\} \tag{3.85}$$

(4) *Let $1 \leq p \leq \infty$. The p-norms of a symmetric operator $T \in Sym(H)$ and $\tau(T) \in Sym(H_{\mathbb{R}})$, are related by*

$$\|\tau(T)\|_p = 2^{1/p} \|T\|_p \ , \text{ if } p < \infty \tag{3.86}$$

$$\|T\| = \|\tau(T)\|_\infty = \|T\|_\infty = \|T\| \tag{3.87}$$

Proof of Lemma 3.14

(1) First we prove the statement for rank-1 projections. This comes from directly checking equation (2.30). Thus $P = ee^*$ gets mapped into $\tau(P) = \epsilon\epsilon^* + J\epsilon\epsilon^* J^*$, where $\epsilon = \mathbf{j}(e)$. Next if $\{e_1, \cdots, e_k\}$ is an orthonormal basis for the range of P then

$$P = \sum_{l=1}^{k} e_l e_l^*$$

By $\mathbb{R}$-linearity, $\tau(P)$ has the form

$$\tau(P) = \sum_{l=1}^{k} (\epsilon_l \epsilon_l^* + J\epsilon_l \epsilon_l^* J^*)$$

where $\epsilon_l = \mathbf{j}(e_l)$, $1 \leq l \leq k$. Note

$$\langle J\epsilon_l, J\epsilon_s \rangle = \langle \epsilon_l, \epsilon_s \rangle = \langle e_l, e_s \rangle_{\mathbb{R}} = \delta_{l,s} \ , \ \langle J\epsilon_l, \epsilon_s \rangle = \langle ie_l, e_s \rangle_{\mathbb{R}} = 0$$

Hence $\{\epsilon_1, \cdots, \epsilon_k, J\epsilon_1, \cdots, J\epsilon_k\}$ is an orthonormal set and since it is spanning the range of $\tau(P)$ it is an orthonormal basis in the range of $\tau(P)$. Hence $\tau(P)$ is an orthonormal projection on $H_{\mathbb{R}}$ of rank $2k$.

(2) Follows by using the spectral factorization of T,

$$T = \sum_{k=1}^{r} b_k P_k \mapsto \tau(T) = \sum_{k=1}^{r} b_k \tau(P_k) \tag{3.88}$$

where $P_1, \cdots, P_r$ are spectral projections and $b_1, \cdots, b_r$ are their associated distinct eigenvalues. For all $k \neq l$, $P_k P_l = 0$ which implies $\tau(P_k)\tau(P_l) = 0$. Thus the right hand side of the second equation in (3.88) represents the spectral factorization of $\tau(T)$. Hence each b_k is an eigenvalue of $\tau(T)$ but with multiplicity twice the multiplicity as an eigenvalue of T. The conclusion now follows.

(3) It is enought to show $tr\{\tau(T)\tau(S)\} = 2tr\{TS\}$. Fix an orthonormal basis in H, say $\{e_1, \cdots, e_n\}$. Then by (1) $\{\mathbf{j}(e_1), \cdots, \mathbf{j}(e_n), J\mathbf{j}(e_1), \cdots, J\mathbf{j}(e_n)\}$ is an orthonormal basis in $H_{\mathbb{R}}$. Note $J\mathbf{j}(e_k) = \mathbf{j}(ie_k)$. It follows

$$\begin{aligned} tr\{\tau(T)\tau(S)\} &= \sum_{k=1}^{n} \langle \tau(S)\mathbf{j}(e_k), \tau(T)\mathbf{j}(e_k) \rangle + \langle \tau(S)J\mathbf{j}(e_k), \tau(T)J\mathbf{j}(e_k) \rangle \\ &= \sum_{k=1}^{n} \langle Se_k, Te_k \rangle + \langle Sie_k, Tie_k \rangle = 2\sum_{k=1}^{n} \langle Se_k, Te_k \rangle \end{aligned}$$

which proves the claim.

(4) Follows from (2): for finite p,

$$\|\tau(T)\|_p = (|a_1|^p + |a_1|^p + \cdots + |a_n|^p + |a_n|^p)^{1/p} = 2^{1/p} (|a_1|^p + \cdots + |a_n|^p)^{1/p} = 2^{1/p} \|T\|_p$$

whereas for $p = \infty$,

$$\|\tau(T)\|_\infty = \max\{|a_1|, |a_1|, \cdots, |a_n|, |a_n|\} = \max\{|a_1|, \cdots, |a_n|\} = \|T\|_\infty.$$

□

3.3. **Proof of Theorem 3.1 and its Corollary 3.5.**

Proof of Theorem 3.1

(1)⇔(2). According to Theorem 2.2 (4), nonlinear map α is injective iff $ker(\mathcal{A}) \cap (\mathcal{S}^{1,0} - \mathcal{S}^{1,0}) = \{0\}$. But using Lemma 3.7(1) and (2) we get equivalently that α is injective iff for all $u, v \in H$ with $[\![u, v]\!] \neq 0$,

$$\mathcal{A}([\![u, v]\!]) \neq 0$$

Equivalently this means

$$\sum_{k=1}^{m} |\langle F_k, [\![u, v]\!]\rangle|^2 > 0$$

Consider now the unit ball in $\mathcal{S}^{1,1}$ with respect to the nuclear norm, say $S_1^{1,1}$. This set is compact in $Sym(H)$. Then let

$$a_0 = min_{T \in \mathcal{S}^{1,1}, \|T\|_1 = 1} \sum_{k=1}^{m} |\langle F_k, T\rangle|^2 \tag{3.89}$$

By homogeneity we obtain (3.41). Then

$$\langle F_k, [\![u, v]\!]\rangle = \frac{1}{2}(\langle u, f_k\rangle\langle f_k, v\rangle + \langle v, f_k\rangle\langle f_k, u\rangle) = real(\langle u, f_k\rangle\langle f_k, v\rangle)$$

and by Lemma 3.8(1),

$$\|[\![u, v]\!]\|_1^2 = \|u\|^2\|v\|^2 - \langle iu, v\rangle_{\mathbb{R}}^2 = \|u\|^2\|v\|^2 - (imag(\langle u, v\rangle))^2$$

Putting together all previous derivations we obtain the equivalence (1) ⇔ (2).

(2)⇔(4). Using Lemma 3.14 (3) we obtain (3.41) is equivalent to

$$\sum_{k=1}^{m} |\langle \tau(F_k), \tau([\![u, v]\!])\rangle|^2 \geq 4a_0 \left[\|u\|^2\|v\|^2 - (real(\langle iu, v\rangle))^2\right] \tag{3.90}$$

Now (2.30) and (2.31) imply

$$\tau(F_k) = [\![\varphi_k, \varphi_k]\!] + [\![J\varphi_k, J\varphi_k]\!] = [\![\varphi_k, \varphi_k]\!] + J[\![\varphi_k, \varphi_k]\!]J^*$$

$$\tau([\![u, v]\!]) = [\![\xi, \eta]\!] + [\![J\xi, J\eta]\!] = [\![\xi, \eta]\!] + J[\![\xi, \eta]\!]J^*$$

where $\varphi_k = \mathbf{j}(f_k)$ and $\xi = \mathbf{j}(u)$, $\eta = \mathbf{j}(v)$ and J^* is the adjoint of J. A direct computation using $J^* = -J$ shows that

$$\langle \tau(F_k), [\![J\xi, J\eta]\!]\rangle = \langle \tau(F_k), [\![\xi, \eta]\!]\rangle$$

Thus

$$\langle \tau(F_k), \tau([\![u, v]\!])\rangle = 2\langle \varphi_k\varphi_k^* + J\varphi_k\varphi_k^*J^*, [\![\xi, \eta]\!]\rangle = 2[\langle \xi, \varphi_k\rangle\langle \varphi_k, \eta\rangle + \langle \xi, J\varphi_k\rangle\langle J\varphi_k, \eta\rangle]$$

With the equation (2.36) we obtain

$$\langle \tau(F_k), \tau([\![u,v]\!])\rangle = 2\langle \Phi_k \xi, \eta\rangle \Rightarrow \sum_{k=1}^{m} |\langle \tau(F_k), \tau([\![u,v]\!])\rangle|^2 = 4\langle R(\xi)\eta, \eta\rangle$$

Now the right-hand side of (3.90) is processed as follows. Note $\|u\| = \|\xi\|$, $\|v\| = \|\eta\|$, and

$$real(\langle iu, v\rangle) = \langle iu, v\rangle_{\mathbb{R}} = \langle J\xi, \eta\rangle$$

Thus

$$\|u\|^2\|v\|^2 - (real(\langle iu, v\rangle))^2 = \|\xi\|^2\|\eta\|^2 - (\langle J\xi, \eta\rangle)^2 = \langle (\|\xi\|^2 1 - J\xi\xi^* J^*)\eta, \eta\rangle$$

Substituting in (3.90) we obtain (3.43).

(3)$\Leftrightarrow$(4). Assume $rank(R(\xi)) = 2n-1$ for all $\xi \neq 0$. A direct computation shows that $R(\xi)(J\xi) = 0$. Hence $J\xi$ is the only independent vector in $ker(R(\xi))$. It follows there is an $a = a(\xi) > 0$ so that

$$R(\xi) \geq a(\xi) P_{J\xi}^{\perp}$$

Note the $a(\xi)$ represents the smallest nonzero eigenvalue of $R(\xi)$ which must be the $2n-1^{th}$. Since the eigenvalues of a matrix depend continuously with the matrix entries, it follows that $a(\xi)$ is a continuous function on ξ. Let $a_0 = \min_{\|\xi\|=1} a(\xi)$. Since the minimum is achieved somewhere on the unit sphere, $a_0 > 0$. Using the homogeneity of degree 2 of $R(\xi)$, we get $a(\xi) = \|\xi\|^2 a(\frac{\xi}{\|\xi\|}) \geq a_0\|\xi\|^2$ which proves (3.43). Conversely, if (3.43) holds true, then $R(\xi)$ has rank at least $2n-1$. Again since $J\xi$ is in the kernel of $R(\xi)$, it follows that $R(\xi)$ must be of rank exactly $2n-1$. □

Proof of Corollary 3.5

(1)$\Leftrightarrow$(2) follows from Theorem 3.1 (2) and equation (3.52) and the fact that $imag(\langle u, v\rangle) = 0$ for all $u, v \in H'$.

(2)$\Leftrightarrow$(5) follows from the relation

$$\sum_{k=1}^{m} |\langle {\Phi'}_k u, v\rangle|^2 = \langle \left(\sum_{k=1}^{m} {\Phi'}_k u u^* {\Phi'}_k \right) v, v\rangle.$$

(4)$\Leftrightarrow$(5). Note first $\dim_{\mathbb{R}} H' = n$ since $H_{\mathbb{R}} = \mathbf{j}(H') \oplus \mathbf{j}(iH')$ is an orthogonal decomposition of the $2n$-dimensional real space $H_{\mathbb{R}}$ into two isomorphic subspaces. Hence $R'(u)$ is of rank-n if and only if it is bounded below by a multiple of the identity restricted to H'. Thus (3.56) follows by the homogeneity of $R'(u)$ with respect to $\|u\|$.

(2)$\Rightarrow$(3). For $u \neq 0$, (3.53) implies $\{{\Phi'}_k u \ , \ 1 \leq k \leq m\}$ spans H'. This is equivalent with (3.54).

(3)$\Rightarrow$(2). From (3.54) we obtain that $\{{\Phi'}_k u \ , \ 1 \leq k \leq m\}$ is a frame for H'. Then (3.53) follows from the lower frame bound condition. □

4. Performance Bounds on Reconstruction Algorithms

In this section we present two performance bounds applicable to any reconstruction algorithm. One bound is deterministic and is based on the constants a_0 introduced in Theorem 3.1. The second bound represents the Cramer-Rao Lower Bound for the stochastic model (2.3).

4.1. **Lipschitz bounds of the inverse map.** Consider the nonlinear map $\alpha : H \to \mathbb{R}^m$. We shall establish a deterministic performance bound for any inversion algorithm in terms of the Lipschitz bounds of the map:

$$\mathcal{A} : \mathcal{S}^{1,0} \to \mathbb{R}^m \ , \ \mathcal{A}(xx^*) = \alpha(x) \tag{4.91}$$

Specifically we want to bound from above and below the following expression:

$$U(x,y) = \frac{\|\mathcal{A}(xx^*) - \mathcal{A}(yy^*)\|^2}{\|xx^* - yy^*\|_1^2} \tag{4.92}$$

Since $xx^* - yy^* = [\![u,v]\!] \in \mathcal{S}^{1,1}$ for some $u, v \in H$ it follows:

$$\sup_{x,y\in H} U(x,y) = \sup_{u,v\in H} \frac{\sum_{k=1}^m |\langle F_k, [\![u,v]\!]\rangle|^2}{\|[\![u,v]\!]\|_1^2} \tag{4.93}$$

$$\inf_{x,y\in H} U(x,y) = \inf_{u,v\in H} \frac{\sum_{k=1}^m |\langle F_k, [\![u,v]\!]\rangle|^2}{\|[\![u,v]\!]\|_1^2} \tag{4.94}$$

These ratios can be further processed as follows

$$\frac{\sum_{k=1}^m |\langle F_k, [\![u,v]\!]\rangle|^2}{\|[\![u,v]\!]\|_1^2} = \frac{\langle R(\xi)\eta, \eta\rangle}{\|\xi\|^2 \langle P_{J\xi}^{\perp}\eta, \eta\rangle}$$

where $\xi = \mathbf{j}(u)$ and $\eta = \mathbf{j}(v)$. Since $R(\xi)\eta = R(\xi)P_{J\xi}^{\perp}\eta$ if follows:

$$\sup_{\xi,\eta\neq 0} \frac{\langle R(\xi)\eta, \eta\rangle}{\|\xi\|^2 \langle P_{J\xi}^{\perp}\eta, \eta\rangle} = \sup_{\xi\neq 0} \frac{\|R(\xi)\|}{\|\xi\|^2} = \max_{\xi\in H_{\mathbb{R}}, \|\xi\|=1} \|R(\xi)\|$$

and

$$\inf_{\xi,\eta\neq 0} \frac{\langle R(\xi)\eta, \eta\rangle}{\|\xi\|^2 \langle P_{J\xi}^{\perp}\eta, \eta\rangle} = a_0^{opt}.$$

Note the constant a_0^{opt} obtained above is the same as the one given in (3.45). Thus we proved:

Theorem 4.1. *Assume the nonlinear map α is injective. Then the map $\mathcal{A} : \mathcal{S}^{1,0} \to \mathbb{R}^m$ defined in (4.91) is bi-Lipschitz between $(\mathcal{S}^{1,0}, \|\cdot\|_1)$ and $(\mathbb{R}^m, \|\cdot\|)$ with the Euclidian norm, and it has the upper Lipschitz bound*

$$B_0 = \sqrt{\max_{\xi\in H_{\mathbb{R}}, \|\xi\|=1} \|R(\xi)\|} \tag{4.95}$$

and the lower Lipschitz bound

$$A_0 = \sqrt{a_0^{opt}} = \sqrt{\min_{\xi \in H_{\mathbb{R}}, \|\xi\|=1} a_{2n-1}(R(\xi))} \tag{4.96}$$

Specifically

$$A_0 \|xx^* - yy^*\|_1 \leq \|\mathcal{A}(xx^*) - \mathcal{A}(yy^*)\| \leq B_0 \|xx^* - yy^*\|_1 \tag{4.97}$$

for all $x, y \in H$.

4.2. **The Cramer-Rao Lower Bound (CRLB).** Consider now the noise model (2.3). We are interested in obtaining a lower bound for any unbiased estimator of x. The derivation of the CRLB in this paper coincides with the one presented in [8] (see Theorem 23 there). In turn this follows the recipe presented in [6]. We will just present the key steps of this derivation. Note that our derivation is canonical, that is basis independent.

Due to non-holomorphy of the nonlinear map α, the analysis is done in the realification space $H_{\mathbb{R}}$. We denote $\zeta = \mathbf{j}(x)$ for the signal $x \in H$. The frame set is $\mathbb{F} = \{f_1, \cdots, f_m\}$ and $F_k = f_k f_k^* \in \mathcal{S}^{1,0}(H)$ denotes the measurement operators. Recall our notation $\Phi_k = \tau(F_k) = \varphi_k \varphi_k^* + J\varphi_k \varphi_k^* J^* \in \mathcal{S}^{2,0}(H_{\mathbb{R}})$. First we compute the Fisher information matrix associated to ζ. The likelihood for this problem is

$$p(y|\zeta) = \frac{1}{(2\pi)^{m/2}\sigma^m} exp\left(-\frac{1}{2\sigma^2}\|y - \alpha(x)\|^2\right) = \frac{1}{(2\pi)^{m/2}\sigma^m} exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{m} |y_k - \langle \Phi_k \zeta, \zeta\rangle|^2\right) \tag{4.98}$$

where σ^2 is the noise variance. The Fisher information matrix is given by (see [26])

$$I(\zeta) = \mathbb{E}\left[(\nabla_\zeta \log(p(y|\zeta)))(\nabla_\zeta \log(p(y|\zeta)))^T\right]. \tag{4.99}$$

The canonical form of this operator is

$$I(\zeta) = \mathbb{E}\left[[\![\nabla_\zeta \log(p(y|\zeta)), \nabla_\zeta \log(p(y|\zeta))]\!]\right]. \tag{4.100}$$

A little bit of algebra shows

$$I(\zeta) = \frac{4}{\sigma^2}\sum_{k=1}^{m} \Phi_k \zeta \zeta^* \Phi_k = \frac{4}{\sigma^2} R(\zeta). \tag{4.101}$$

In general the covariance of any unbiased estimator is bounded below by the inverse of the Fisher information matrix (operator). However in this case the Fisher information operator is not invertible. This fact simply expresses the statement that x is not identifiable from the measurements $y = \alpha(x) \in \mathbb{R}^m$ alone. As we know the nonlinear map α is not injective on H but instead it is injective on $\hat{H}$. The nonuniqueness on H is reflected in having a singular Fisher information matrix on $H_{\mathbb{R}}$. To solve this issue we need to fix the global phase factor. One solution is to fix a basis and decide that a particular component (say the last component) is real. Such an approach was taken in [8]. Here we propose a canonical solution to this normalization. An oracle provides us with a vector $z_0 \in H$, so that $\langle x, z_0\rangle > 0$ is positive real. Assume z_0 is normalized $\|z_0\| = 1$. Note there are two pieces of information that can

be extracted from here: First the fact that x is not orthogonal to z_0; in particular $x \neq 0$. Second, the global phase to recover x from its rank-1 operator xx^* is uniquely determined by the fact that $imag(\langle x, z_0\rangle) = 0$ and $real(\langle x, z_0\rangle) > 0$.

Under this scenario we want to analyze the Fisher information operator obtained earlier. Let $\psi_0 = \mathbf{j}(z_0) \in H_{\mathbb{R}}$. We know

$$\langle \zeta, \psi_0\rangle = real(\langle x, z_0\rangle) > 0 \ , \ \langle \zeta, J\psi_0\rangle = imag(\langle x, z_0\rangle) = 0$$

with $\zeta = \mathbf{j}(x)$. Let Π denote the orthogonal projection onto the complement of $J\psi_0$,

$$\Pi : H_{\mathbb{R}} \to E \ , \ \Pi = 1 - J\psi_0\psi_0^* J^* \tag{4.102}$$

where $E = \{J\psi_0\}^{\perp}$. Let H_{z_0} denote the following closed set

$$H_{z_0} = \{\xi \in H_{\mathbb{R}} \ , \ \langle \xi, \psi_0\rangle \geq 0 \, , \, \langle \xi, J\psi_0\rangle = 0\} \subset E. \tag{4.103}$$

Note ζ belongs to the relative interior of H_{z_0}. The class of estimators for ζ should include only functions

$$\omega : \mathbb{R}^m \to H_{z_0} \tag{4.104}$$

In this case the appropriate Fisher information operator should be

$$\tilde{I}(\zeta) := \Pi I(\zeta)\Pi = \frac{4}{\sigma^2}\sum_{k=1}^{m} \Pi\Phi_k\zeta\zeta^*\Phi_k\Pi \tag{4.105}$$

The following lemma proves that under the scenario described here, $\tilde{I}(\zeta)$ is invertible on H_{z_0}.

Lemma 4.2. *Assume α is injective on $\hat{H}$ and $z_0 \in H$ is so that $\langle x, z_0\rangle > 0$. Let $\zeta = \mathbf{j}(x)$. Then*

$$\tilde{I}(\zeta) := \Pi I(\zeta)\Pi \geq \frac{4}{\sigma^2} a_0 |\langle x, z_0\rangle|^2 \, \Pi \tag{4.106}$$

where $a_0 = a_0^{opt}$ is the same lower bound introduced in Theorem 3.1 whose optimal value is given by (3.45). Furthermore this bound is tight.

Proof

Using (2.35) the left-hand side of (4.106) is $\tilde{I}(\zeta) = \frac{4}{\sigma^2}\Pi R(\zeta)\Pi$. We know $R(\zeta) \geq a_0\|\zeta\|^2 P_{J\zeta}^{\perp}$ from Theorem 3.1 (4). Therefore we only need to show

$$\|\zeta\|^2 \, \Pi \, P_{J\zeta}^{\perp} \, \Pi \ \geq \ |\langle \zeta, \psi_0\rangle|^2 \, \Pi$$

where $\psi_0 = \mathbf{j}(z_0)$, and the inequality is tight. Without loss of generality we can assume $\|\zeta\| = 1$ since all expressions are homogeneous in $\|\zeta\|$. Then we need to show that for any $\xi \in E$, $\|\xi\| = 1$,

$$\langle P_{J\zeta}^{\perp}\xi, \xi\rangle \geq |\langle \zeta, \psi_0\rangle|^2$$

This follows from

$$\inf_{\|\xi\|=1, \xi\in E} 1 - |\langle \xi, J\zeta\rangle|^2 = 1 - \max_{\|\xi\|=1, \xi\in E} |\langle \xi, J\zeta\rangle|^2 = 1 - \left|\langle \frac{\Pi J\zeta}{\|\Pi J\zeta\|}, J\zeta\rangle\right|^2 = |\langle \zeta, \psi_0\rangle|^2$$

The last equality follows by direct computation from:

$$\Pi J\zeta = J(\zeta - \langle \zeta, \psi_0\rangle \psi_0) \ , \ \|\zeta - \langle \zeta, \psi_0\rangle \psi_0\|^2 = 1 - |\langle \zeta, \psi_0\rangle|^2.$$

Note also the inequality in (4.106) is tight since the lower bound is achieved for $\zeta = \psi_0 = argmin_{\xi \in H_{\mathbb{R}}, \|\xi\|=1} a_{2n-1}(R(\xi))$, the optimizer in (3.45). □

Thus we established that $\tilde{I}(\xi)$ is invertible on H_{z_0}. See also [8], Lemma 22, for a similar statement.

Recall an estimator $\omega : \mathbb{R}^m \to H_{z_0}$ is said to be *unbiased* if

$$\mathbb{E}[\omega(y)|\zeta = \mathbf{j}(x)] = \zeta \tag{4.107}$$

We can now state the main result of this section:

Theorem 4.3. *Assume the nonlinear map α is injective and fix a vector $z_0 \in H$. For any vector $x \in H$ with $\langle x, z_0\rangle > 0$ the covariance operator of any unbiased estimator $\omega : \mathbb{R}^m \to H_{z_0}$ of x is bounded below by the Cramer-Rao Lower Bound (CRLB) given by*

$$Cov[\omega(y)|\zeta = \mathbf{j}(x)] \geq \frac{\sigma^2}{4} \left(\sum_{k=1}^{m} \Pi \Phi_k \zeta \zeta^* \Phi_k \Pi \right)^{\dagger} \tag{4.108}$$

where $\dagger$ denotes the pseudoinverse operator, $\zeta = \mathbf{j}(x)$ and $\Pi = 1 - J\psi_0\psi_0^* J^*$. In particular the Mean Square Error of ω is bounded below by

$$MSE(\omega) = \mathbb{E}[\|x - \omega(y)\|^2 | \zeta = \mathbf{j}(x)] \geq \frac{\sigma^2}{4} tr \left\{ \left(\sum_{k=1}^{m} \Pi \Phi_k \zeta \zeta^* \Phi_k \Pi \right)^{\dagger} \right\} \tag{4.109}$$

Proof

The key observation is that H_{z_0} is a relatively open subset of the real linear space E, the orthogonal complement of $J\psi_0$ in $H_{\mathbb{R}}$, $E = \{J\psi_0\}^{\perp} \cap H_{\mathbb{R}}$. Consider an orthonormal basis in E of the form $\{e_1, \cdots, e_{2n-1}\}$. Thus $\{e_1, \cdots, e_{2n-1}, J\psi_0\}$ is an orthonormal basis in $H_{\mathbb{R}}$. The (column vector) gradient with respect to E, ∇_{ζ}^{E} has the form $\nabla_{\zeta}^{E} = \Pi \nabla_{\zeta}$ where ∇ is the gradient with respect to the local coordinates in $H_{\mathbb{R}}$ and Π is the orthogonal projection onto E. This shows the Fisher information matrix associated to the Additive White Gaussian Noise (AWGN) measurement process (2.3) with $\zeta \in H_{z_0}$ is $\tilde{I}(\zeta)$ given by (4.105). Theorem 3.2 in [26] implies the covariance matrix of ω is bounded below by the inverse of $\tilde{I}(\zeta)$ restricted to E. This implies (4.108). Equation (4.109) follows from $MSE(\omega) = tr\{Cov[\omega(y)|\zeta = \mathbf{j}(x)]\}$ and (4.108). □

Lemma 4.2 allows us to predict an upper bound for the MSE of any efficient estimator (that is an unbiased estimator that achieves the CRLB):

Corollary 4.4. *Assume $\omega : \mathbb{R}^m \to H_{z_0}$ is an unbiased estimator that achieves the CRLB (4.108). Then its Mean-Square Error is bounded above by*

$$MSE(\omega) = \mathbb{E}[\|x - \omega(y)\|^2 | x] \leq \frac{(2n-1)\sigma^2}{4a_0^{opt} |\langle x, z_0\rangle|^2} \tag{4.110}$$

Proof

If ω is unbiased and achieves the CRLB (in other words, if ω is efficient) then

$$MSE(\omega) = \frac{\sigma^2}{4} tr\left\{\left(\sum_{k=1}^{m} \Pi\Phi_k\zeta\zeta^*\Phi_k\Pi\right)^{\dagger}\right\} = tr\left\{\left(\tilde{I}(\zeta)\right)^{\dagger}\right\}$$

Then (4.110) follows from (4.106) by noting that $tr\{\Pi\} = 2n-1$. □

5. The Iterative Regularized Least-Square (IRLS) Algorithm

Consider the additive noise model in (2.3). Our data is the vector $y \in \mathbb{R}^m$. Our goal is to find an $x \in H$ that minimizes $\|y - \alpha(x)\|$, where we use the Euclidian norm. Set

$$J_0(X) = \sum_{k=1}^{m} |y_k - \langle Xf_k, f_k\rangle|^2 \ , \ J_0 : Sym(H) \to \mathbb{R} \tag{5.111}$$

and notice $J_0(xx^*) = \|y - \alpha(x)\|^2$. The least-square error minimizer represents the Maximum Likelihood Estimator (MLE) when the noise is Gaussian. In this section we discuss an optimization algorithm for this criterion.

Consider now $J_0 = \|y - \mathcal{A}(X)\|^2$ where is X is restricted to $\mathring{\mathcal{S}}^{1,0}$, which is an analytic manifold. Consider a current point $X^{(t)} = x^{(t)}(x^{(t)})^*$ in an iterative process. Then a descent direction can be thought of as a vector in the tangent space to the manifold. According to Lemma 3.11 (1), the tangent space at $X^{(t)}$ is given by operators of the form $[\![x^{(t)}, \delta]\!]$. Since $X^{(t)} + [\![x^{(t)}, \delta]\!] = [\![x^{(t)}, x^{(t)} + \delta]\!] \in \mathcal{S}^{1,1}$, one would need to project $X^{(t)} + [\![x^{(t)}, \delta]\!]$ back into $\mathcal{S}^{1,0}$ and choose direction δ that minimizes (or at least decreases) $J_0(P(X^{(t)} + [\![x^{(t)}, \delta]\!]))$, where P is the (nonlinear) projection in $Sym(H)$ onto $\mathcal{S}^{1,0}$. However since J_0 is well defined on $\mathcal{S}^{1,1}$ we choose to optimize δ without projecting back onto $\mathcal{S}^{1,0}$. Thus we obtain the iterative process:

$$x^{(t+1)} = argmin_u J_0([\![x^{(t)}, u]\!])$$

However this process is not robust to noise, the main reason being ill-conditioning and multiple local minima of J_0 on $\mathcal{S}^{1,0}$. Instead we choose to regularize this process and thus to introduce a different optimization criterion.

Consider the following functional

$$J : H \times H \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+ \tag{5.112}$$

$$J(u, v, \lambda, \mu) = \sum_{k=1}^{m} |y_k - \frac{1}{2}(\langle u, f_k\rangle\langle f_k, v\rangle + \langle v, f_k\rangle\langle f_k, u\rangle)|^2 + \lambda\|u\|^2 + \mu\|u - v\|^2 + \lambda\|v\|^2.$$

Our ultimate goal is to minimize $J_0(uu^*) = \|y - \alpha(u)\|^2 = J(u, u, 0, \mu)$ over u, for some (and hence any) value of $\mu \in \mathbb{R}^+$. Our strategy is based on the following iterative process:

Algorithm 5.1. The Iterative Regularized Least-Square (IRLS) Algorithm

Step 0. Initialize x^0 as the global optimal solution for a specific pair (λ_0, μ_0).

Step 1. Iterate:

1.1 Solve for

$$x^{(t+1)} = argmin_u J(u, x^{(t)}, \lambda_t, \mu_t) \tag{5.113}$$

1.2 Update λ_{t+1}, μ_{t+1} according to a specific policy;
Step 2. Stop when some tolerance level is achieved.

As we describe below the update (5.113) can be modified to achieve a more robust behavior (see (5.131)).

5.1. **Initialization.** Consider the regularized least-square problem:

$$min_u J(u, u, \lambda, 0) = min_u \|y - \alpha(u)\|^2 + 2\lambda\|u\|^2$$

Note the following relation

$$\begin{aligned} J(u, u, \lambda, 0) &= \|y\|^2 + 2\lambda\|u\|^2 - 2\sum_{k=1}^{m} y_k |\langle u, f_k\rangle|^2 + \sum_{k=1}^{m} |\langle u, f_k\rangle|^4 \\ &= \|y\|^2 + 2\langle(\lambda I - Q)u, u\rangle + \sum_{k=1}^{m} |\langle u, f_k\rangle|^4 \end{aligned} \tag{5.114}$$

where

$$Q = \sum_{k=1}^{m} y_k f_k f_k^* = \sum_{k=1}^{m} y_k F_k \tag{5.115}$$

For $\lambda > \|Q\|$ the optimal solution is $u = 0$. Note that if $Q \leq 0$ (as a quadratic form) then the optimal solution of $min_u \|y - \alpha(u)\|^2$ is $u = 0$. Consequently in the following we assume the largest eigenvalue of Q is positive. As λ decreases the optimizer remains small. Hence we can neglect the forth order term in u in the expansion above and obtain:

$$J(u, u, \lambda, 0) \approx \|y\|^2 + 2\langle(\lambda I - Q)u, u\rangle$$

Thus the critical value of λ for which we may get a nonzero solution is $\lambda = maxeig(Q)$, which is the maximum eigenvalue of Q. Let us denote by a_1 this (positive) eigenvalue and e_1 its associated normalized eigenvector. This suggests to initialize $\lambda = \rho a_1$ for some $0 < \rho \leq 1$ and $x^{(0)} = \beta e_1$, for some nonzero scalar β. Substituting into (5.114) we obtain

$$J(\beta e_1, \beta e_1, \rho a_1, 0) = \|y\|^2 - 2(1-\rho)a_1\beta^2 + (\sum_{k=1}^{m} |\langle e_1, f_k\rangle|^4)\beta^4$$

For fixed ρ, the minimum over β is achieved at

$$\beta_0 = \sqrt{\frac{(1-\rho)a_1}{\sum_{k=1}^{m} |\langle e_1, f_k\rangle|^4}}\ ,\ x^{(0)} = \beta e_1 \tag{5.116}$$

The parameter μ controls the step size at each iteration. The larger the value the smaller the step. On the other hand, a small value of this parameter may produce an unstable behavior of the iterates. In our implementation we use the same initial value for both λ and μ:

$$\mu_0 = \lambda_0 = \rho a_1 \tag{5.117}$$

5.2. **Iterations.** Minimization (5.113) is performed in the space $H_{\mathbb{R}}$. Let $\zeta^{(t)} = \mathbf{j}(x^{(t)})$ and $\xi = \mathbf{j}(u)$. Then

(5.118)

$$J(u, x^{(t)}, \lambda_t, \mu_t) = \sum_{k=1}^{m} |\langle \left(\Phi_k \zeta^{(t)} (\zeta^{(t)})^* \Phi_k\right) \xi, \xi\rangle - y_k|^2 + \lambda_t \|\xi\|^2 + \mu_t \|\xi - \zeta^{(t)}\|^2 + \lambda_t \|\zeta^{(t)}\|^2$$

Note the criterion is quadratic in ξ. The unique minimum is given by solving the linear equation:

$$\left(\sum_{k=1}^{m} \Phi_k \zeta^{(t)} (\zeta^{(t)})^* \Phi_k + (\lambda_t + \mu_t) 1\right) \zeta^{(t+1)} = \left(\sum_{k=1}^{m} y_k \Phi_k + \mu_t 1\right) \zeta^{(t)} \tag{5.119}$$

for $\zeta^{(t+1)}$. In our implementations we decrease (λ_t, μ_t) but we limit μ_t to a minimum value. Thus our adaptation policy is

$$\lambda_{t+1} = \gamma \lambda_t \tag{5.120}$$

$$\mu_{t+1} = max(\gamma \mu_t, \mu^{min}) \tag{5.121}$$

where $0 < \gamma < 1$ is the rate parameter.

5.3. **Stopping Criterion.** One approach is to repeat the iterations until λ reaches a preset value λ^{min}. As proved later in this section, the error is linearly dependent on λ.

Alternatively, one can stop the iterations once the modeling error becomes comparable to the noise variance. Specifically, a stopping criterion could be

$$\sum_{k=1}^{m} |y_k - |\langle x^{(t)}, f_k\rangle|^2|^2 \leq \kappa m \sigma^2 \tag{5.122}$$

where $\kappa \geq 1$, for instance $\kappa = 3$.

5.4. **Convergence and Optimality.** Consider the following three functionals $J_1, J_2, J_3 : Sym(H) \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ defined by

$$J_1(X, \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, F_k\rangle|^2 + 2(\lambda + \mu)\|X\|_1 - 2\mu\, tr\{X\} \tag{5.123}$$

$$J_2(X, \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, F_k\rangle|^2 + 2\lambda\, a_{max}(X) - (2\lambda + 4\mu)\, a_{min}(X) \tag{5.124}$$

$$J_3(X, \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, F_k\rangle|^2 + 2\lambda\, \|X\|_1 - 4\mu\, a_{min}(X) \tag{5.125}$$

where $a_{max}(X)$ and $a_{min}(X)$ are the maximum and minimum eigenvalue of X, respectively. We can prove the following result:

Lemma 5.2. (1) *When restricted to $\mathcal{S}^{1,1}$ the three criteria coincide:*

$$J_1(X,\lambda,\mu) = J_2(X,\lambda,\mu) = J_3(X,\lambda,\mu)\ ,\ \forall X \in \mathcal{S}^{1,1}, \lambda,\mu \in \mathbb{R}. \tag{5.126}$$

(2) *On $Sym(H)$, the three criteria J_1, J_2, J_3 are convex.*

(3) *The minimum value of J_1, J_2, J_3 on $\mathcal{S}^{1,1}$ coincides with the minimum value of J on $H \times H$:*

$$\min_{X\in\mathcal{S}^{1,1}} J_1(X,\lambda,\mu) = \min_{X\in\mathcal{S}^{1,1}} J_2(X,\lambda,\mu) = \min_{X\in\mathcal{S}^{1,1}} J_3(X,\lambda,\mu) = \min_{u,v\in H} J(u,v,\lambda,\mu) \tag{5.127}$$

for any $\lambda,\mu \geq 0$. Any minimizer $\hat{X} \in \mathcal{S}^{1,1}$ for J_1, J_2, J_3 and $(\hat{u},\hat{v})$ for J satisfy

$$\hat{X} = [\![\hat{u},\hat{v}]\!]\ ,\ \|\hat{u}\| = \|\hat{v}\|\ ,\ imag(\langle\hat{u},\hat{v}\rangle) = 0 \tag{5.128}$$

(4) *Restricted to $\mathcal{S}^{1,0}$ all four criteria coincide:*

$$J(u,u,\lambda,\mu) = J_1(uu^*,\lambda,\mu) = J_2(uu^*,\lambda,\mu) = J_3(uu^*,\lambda,\mu) = \|y-\alpha(u)\|^2 + 2\lambda\|u\|^2 \tag{5.129}$$

and are independent of μ.

Proof

For (1), the quadratic error term is the same in all three criteria, whereas the regularization terms are equal to each other:

$$\begin{aligned} 2(\lambda+\mu)\|X\|_1 - 2\mu tr\{X\} &= 2(\lambda+\mu)(a_1+a_2) - 2\mu(a_1-a_2) = 2\lambda a_1 + (2\lambda+4\mu)a_2 \\ 2\lambda a_{max}(X) - (2\lambda+4\mu)a_{min}(X) &= 2\lambda a_1 + (2\lambda+4\mu)a_2 \\ 2\lambda\|X\|_1 - 4\mu a_{min}(X) &= 2\lambda(a_1+a_2) - 4\mu(-a_2) = 2\lambda a_1 + (2\lambda+4\mu)a_2 \end{aligned}$$

where $X = a_1 e_1 e_1^* - a_2 e_2 e_2^*$ with $a_1, a_2 \geq 0$ and $\{e_1, e_2\}$ orthonormal set.

For (2) notice that the following four functions defined on the real vector space $Sym(H)$ are convex: $X \mapsto |y_k - \langle X, F_k\rangle|^2$, $X \mapsto \|X\|_1$, $X \mapsto -tr\{X\}$, $X \mapsto a_{max}(X)$, whereas $X \mapsto a_{min}(X)$ is concave. The last two statements are consequences of the Weyl's Inequality, Theorem III.2.1 in [10] with $i = j = 1$ in (III.5), and $i - j = n$ in (III.6).

For (3) and (4) note first the following relation:

$$\begin{aligned} J(u,v,\lambda,\mu) - J_1([\![u,v]\!],\lambda,\mu) &= (\lambda+\mu)\left[(\|u\|-\|v\|)^2 + 2\|u\|\|v\| - 2\sqrt{\|u\|^2\|v\|^2 - (imag(\langle u,v\rangle))^2}\right] \\ &\geq (\lambda+\mu)(\|u\|-\|v\|)^2 \geq 0 \end{aligned} \tag{5.130}$$

that follows from (3.67) and (3.71). Using part (1) we obtain

$$\min_{X\in\mathcal{S}^{1,1}} J_1(X,\lambda,\mu) = \min_{X\in\mathcal{S}^{1,1}} J_2(X,\lambda,\mu) = \min_{X\in\mathcal{S}^{1,1}} J_3(X,\lambda,\mu) \leq \min_{u,v\in H} J(u,v,\lambda,\mu).$$

Let $\hat{X}$ denote the optimizer and let $\hat{X} = a_1\hat{e}_1\hat{e}_1^* - a_2\hat{e}_2\hat{e}_2^*$ be its spectral decomposition with $a_1, a_2 \geq 0$ and $\langle\hat{e}_i,\hat{e}_j\rangle = \delta_{i,j}$, $1 \leq i,j \leq 2$. Set $\hat{u} = \sqrt{a_1}\hat{e}_1 + \sqrt{a_2}\hat{e}_2$ and $\hat{v} =$

$\sqrt{a_1}\hat{e}_1 - \sqrt{a_2}\hat{e}_2$. Note $\hat{X} = [\![\hat{u}, \hat{v}]\!]$ and $\|\hat{u}\| = \|\hat{v}\|$, $imag(\langle \hat{u}, \hat{v}\rangle) = 0$. Then (5.130) implies $J(\hat{u}, \hat{v}, \lambda, \mu) = J_1(\hat{X}, \lambda, \mu)$ which proves (5.127). Furthermore, let $(\hat{u}, \hat{v})$ be a minimizer for $\min_{u,v\in H} J(u, v, \lambda, \mu)$ which achieves also equality in (5.127). By (5.130) it follows that $\|\hat{u}\| = \|\hat{v}\|$ and $imag(\langle \hat{u}, \hat{v}\rangle) = 0$ which proves statement (3).

(4) follows from the first equality in (5.130) and (5.126). □

Remark 5.3. *The criterion J_2 shows that the two regularization terms $\lambda(\|u\|^2 + \|v\|^2)$ and $\mu\|u - v\|^2$ have different effects on the optimizer: the larger the parameter μ the closer the lower eigenvalue is to zero; hence the closer the optimizer $\hat{X}$ is to a rank-1 operator; on the other hand, the larger the parameter λ the larger the cost of the $\mathcal{S}^{1,0}$ component in $\hat{X}$; hence $\|\hat{X}\|$ remains bounded.*

Remark 5.4. *In the optimization problem $\inf_{u,v\in H} J(u, v, \lambda, \mu)$ the constraint is convex but the criterion is not jointly convex in (u, v). It is however convex in each individual variable u and v. On the other hand in the optimization problem $\inf_{X\in\mathcal{S}^{1,1}} J_s(X, \lambda, \mu)$, $1 \leq s \leq 3$, the criterion is convex in X, but the underlying constraint $X \in \mathcal{S}^{1,1}$ does not define a convex set.*

The optimization procedure outlined in Algorithm 5.1 describes the following mechanism. Consider a path $(x^{(t)})_{t\geq 0}$ in H. Set

$$X^{(t+1)} = [\![x^{(t)}, x^{(t+1)}]\!] = \frac{1}{2}\left(x^{(t+1)}(x^{(t)})^* + x^{(t)}(x^{(t+1)})^*\right)$$

Thus a trajectory $(x^{(t)})_{t\geq 0}$ in H is mappped into a trajectory $(X^{(t)})_{t\geq 0}$ in $\mathcal{S}^{1,1}$. Then the algorithm chooses $X^{(t+1)}$ along a tangent direction to $\mathcal{S}^{1,1}$ at $[\![x^{(t)}, x^{(t-1)}]\!]$, namely a direction of the form $[\![x^{(t)}, u]\!]$ for some $u \in H$ that is of maximum descent for J_s, $1 \leq s \leq 3$. Since the algorithm performance is completely characterized by the sequence $(X^{(t)})_{t\geq 0}$, and since for a given $X \in \mathcal{S}^{1,1}$ the minimum of $J(u, v, \lambda, \mu)$ over $u, v \in H$ subject to $[\![u, v]\!] = X$ is achieved at a pair (u, v) so that $\|u\| = \|v\|$ and $imag\langle u, v\rangle = 0$, we chose to rescale the vector $x^{(t+1)}$ to achieve norm $\sqrt{\|\zeta^{(t)}\|\|\zeta^{(t+1)}\|}$. Thus:

$$x^{(t+1)} = \sqrt{\frac{\|\zeta^{(t)}\|}{\|\zeta^{(t+1)}\|}}\mathbf{j}^{-1}(\zeta^{(t+1)}) \tag{5.131}$$

5.5. **Robustness to noise and the effect of regularization.** In this subsection we make a stronger assumption of injectivity:

Assumption A: For every $x \in H$ there is a unique $X \in \mathcal{S}^{1,1}$ so that $\mathcal{A}(X) = \mathcal{A}(xx^*)$, namely $X = xx^*$.

A simple heuristic argument (that can be made more precise, but we do not intend to do so here) suggests that generically this assumption is satisfied for frames of redundancy 6 or more (that is for $m \geq 6n$).

This assumption turns out to be equivalent to a stability bound that we describe next

Lemma 5.5. *The following are equivalent:*

(1) *The frame* $\mathbb{F}$ *satisfies assumption A.*
(2) $\ker \mathcal{A} \cap \mathcal{S}^{2,1} = \{0\}$.
(3) *There is a constant* $A_3 > 0$ *so that*

$$A_3 \|X - Y\|_1^2 \leq \|\mathcal{A}(X) - \mathcal{A}(Y)\|^2 \tag{5.132}$$

for all $X \in \mathcal{S}^{1,1}$ *and* $Y \in \mathcal{S}^{1,0}$.

Proof

(1)⇒(2). Note first that $\mathcal{S}^{1,0} - \mathcal{S}^{1,1} = \mathcal{S}^{2,1}$ cf. Lemma 3.6 (5). Assume $\mathbb{F}$ satisfies assumption A and let $Z \in \ker \mathcal{A} \cap \mathcal{S}^{2,1}$. Then $Z = X - xx^*$ for some $X \in \mathcal{S}^{1,1}$ and $x \in H$. Then $\mathcal{A}(X) = \mathcal{A}(xx^*)$ and by Assumption A, $X = xx^*$. Hence $Z = 0$.

(2)⇒(1). Conversely if (2) holds true, then for every $x \in H$ and $X \in \mathcal{S}^{1,1}$ so that $\mathcal{A}(X) = \mathcal{A}(xx^*)$ it follows $Z = X - xx^* \in \ker \mathcal{A}$ and $Z \in \mathcal{S}^{2,1}$. Thus $Z = 0$ which means $\mathbb{F}$ satisfies assumption A.

(3)⇒(2) is immediate.

(2)⇒(3). Since $S = \{W \in \mathcal{S}^{2,1} \mid \|W\|_1 = 1\}$ is compact, it follows

$$A_3 = \inf_{W \in S} \|\mathcal{A}(W)\|^2 = \|\mathcal{A}(W_0)\|^2 > 0 \tag{5.133}$$

for some $W_0 \in \mathcal{S}^{2,1}$, $W_0 \neq 0$. Then since any $Z \in \mathcal{S}^{2,1}$ can be written as $Z = \|Z\|_1 W$, with $W = \frac{1}{\|Z\|_1} Z \in S$,

$$\|\mathcal{A}(Z)\|^2 = \|Z\|_1^2 \|\mathcal{A}(W)\|^2 \geq A_3 \|Z\|_1^2$$

Then (5.132) follows from noticing that $Y - X \in \mathcal{S}^{2,1}$. □

Now we can state the main stability result of the estimators described in this section.

Theorem 5.6. *Fix* $\mu \geq 0$, $x \in H$ *and let* $y = \alpha(x) + \nu$. *Assume* $\mathbb{F}$ *satisfies assumption A and let* A_3 *denote the Lipschitz bound in (5.132). Assume the optimization procedure finds a pair* $u, v \in H$ *so that* $J(u, v, \lambda, \mu) \leq J(x, x, \lambda, \mu)$. *Then*

$$\|[\![u, v]\!] - xx^*\|_1 \leq \frac{2\lambda}{A_3} + 2\sqrt{\frac{\lambda^2}{A_3^2} + \frac{\|\nu\|^2}{A_3}} \leq 4\frac{\lambda}{A_3} + 2\frac{\|\nu\|}{\sqrt{A_3}} \ . \tag{5.134}$$

Let $[\![u, v]\!] = a_1 e_1 e_1^* - a_2 e_2 e_2^*$, *with real* $a_1, a_2 \geq 0$ *and* $\{e_1, e_2\}$ *orthonormal set in* H, *be its spectral decomposition. Assume an oracle provides the global phase* φ_0 *so that* $e^{i\varphi_0} = \frac{\langle x, e_1 \rangle}{|\langle x, e_1 \rangle|}$. *Set*

$$\hat{x} = e^{i\varphi_0} \sqrt{a_1}\, e_1 \ . \tag{5.135}$$

Then

$$\|x - \hat{x}\|^2 \leq \|[\![u, v]\!] - xx^*\|_1 + a_2 \leq \frac{4\lambda}{A_3} + \frac{2\|\nu\|}{\sqrt{A_3}} + \frac{\|\nu\|^2}{4\mu} + \frac{\lambda \|x\|^2}{2\mu} \ . \tag{5.136}$$

If, additionally, $J(u, v, \lambda, \mu) \leq J(0, 0, \lambda, \mu)$ *then*

$$\|x - \hat{x}\|^2 \leq \frac{4\lambda}{A_3} + \frac{2\|\nu\|}{\sqrt{A_3}} + \frac{\|\nu\|^2}{4\mu} \ . \tag{5.137}$$

Proof

The proof follows from the Lipschitz bound (5.132). Let $Y = [\![u, v]\!]$ and $X = xx^*$. Note that

$$J_3(Y, \lambda, \mu) \leq J(u, v, \lambda, \mu) \leq J(x, x, \lambda, \mu) = J_3(xx^*, \lambda, \mu)$$

where the first inequality follows from (5.126) and (5.130). Explicitly this means

$$\|\mathcal{A}(Y - X) - \nu\|^2 + 2\lambda\|Y\|_1 - 4\mu a_{min}(Y) \leq \|\nu\|^2 + 2\lambda\|X\|_1 \ .$$

Then

$$A_3\|Y - X\|_1^2 \leq \|\mathcal{A}(Y - X)\|^2 \leq (\|\mathcal{A}(Y - X) - \nu\| + \|\nu\|)^2 \leq 2\|\mathcal{A}(Y - X) - \nu\|^2 + 2\|\nu\|^2 \ .$$

Substituting into the previous inequality we obtain

$$\|X - Y\|_1^2 - \frac{4\lambda}{A_3}\|X - Y\|_1 - \frac{4\|\nu\|^2 + 8\mu a_{min}(Y)}{A_3} \leq 0 \ .$$

Solving for $\|X - Y\|_1$ we obtain

$$\|X - Y\|_1 \leq \frac{2\lambda}{A_3} + 2\sqrt{\frac{\lambda^2}{A_3^2} + \frac{\|\nu\|^2}{A_3} + \frac{2\mu}{A_3}a_{min}(Y)} \leq \frac{2\lambda}{A_3} + 2\sqrt{\frac{\lambda^2}{A_3^2} + \frac{\|\nu\|^2}{A_3}}$$

since $\mu \geq 0$ and $a_{min}(Y) \leq 0$ for any $Y \in \mathcal{S}^{1,1}$. This proves the first inequality in (5.134). The second inequality in (5.134) follows from $\sqrt{a^2 + b^2} \leq a + b$ for any $a, b \geq 0$.

The second part of the Theorem is obtained as follows. First note $\hat{x}\hat{x}^* = a_1 e_1 e_1^*$. Hence

$$\|\hat{x}\hat{x}^* - xx^*\|_1 \leq \|[\![u, v]\!] - xx^*\|_1 + a_2 \ .$$

Next we show that $\|x - \hat{x}\|^2 \leq \|\hat{x}\hat{x}^* - xx^*\|_1$, from where the first inequality of (5.136) follows. Let $T = \hat{x}\hat{x}^* - xx^* \in \mathcal{S}^{1,1}$. Its nuclear norm is given by (3.76):

$$\|T\|_1 = \sqrt{(\|\hat{x}\|^2 + \|x\|^2)^2 - 4|\langle x, \hat{x}\rangle|^2} \ .$$

Recall $\hat{x}$ is given the global phase so that $\langle \hat{x}, x\rangle \geq 0$ is real and nonnegative. Thus $|\langle \hat{x}, x\rangle|^2 = (\langle \hat{x}, x\rangle)^2$ and

$$\|x - \hat{x}\|^4 = \|T\|_1^2 - 4\langle \hat{x}, x\rangle \cdot \|x - \hat{x}\|^2 \leq \|T\|_1^2$$

which shows $\|x - \hat{x}\|^2 \leq \|\hat{x}\hat{x}^* - xx^*\|_1$ and thus first inequality in (5.136). The second inequality follows from (5.134) and a bound for a_2 from

$$4\mu a_2 \leq J_3(Y, \lambda, \mu) \leq J_3(X, \lambda, \mu) = \|\nu\|^2 + 2\lambda\|x\|^2.$$

Finally, (5.137) follows from using the second inequality in (5.134) together with a bound for a_2 from

$$4\mu|a_2| \leq J_3(Y, \lambda, \mu) \leq J_3(0, \lambda, \mu) = \|\nu\|^2.$$

□

Remark 5.7. *Note the result does not require to finding the global minimum, but only an estimate that brings the criterion J below the level achieved by the true signal. On the other hand the result is not unexpected since MLE is asymptotically efficient (i.e. unbiased and achieves CRLB asymptotically with the number of measurements), see Theorem 7.1 in* [26].

6. Numerical Analysis

In this section we present numerical simulations for the Regularized Iterative Least-Square algorithm presented in the previous section.

We generated random frames or redundancy $\frac{m}{n} = 4$,6, and 8, as well as complex random signals x of size $n = 100$. All these vectors (frame and signal) are drawn from standard Gaussian distribution $\mathcal{N}(0, I)$. Then we scale the frame vectors to have norm 1. We set the first component of x to be a positive real, and so the global phase becomes uniquely determined. The algorithm stopped when λ_t reached a preset value. In these simulations we choose $\lambda^{min} = 0.01$, $\mu^{min} = 1$ and rate $\gamma = 0.8$ in (5.120,5.121).

The magnitude square of signal coefficients $\alpha(x)$ is perturbed additively by Gaussian noise with variance σ^2 to achieve a fixed signal-to-noise-ratio defined as

$$SNR = \frac{\sum_{k=1}^{m} |\langle x, f_k\rangle|^4}{m\sigma^2} \ , \ SNRdB = 10\log_{10}(SNR) \text{ [dB]}$$

We vary SNRdB over 15 values in 5dB increments from -30dB to +40dB. We average algorithm performance over 100 noise realizations.

Figure 1 includes the mean-square error averaged over 100 noise realizations for one fixed realization of x, and the C-R lower bound.

In Figure 2 we plot the bias and variance components of the mean-square error for the same results in Figure 1. Note the bias is relatively small. The bulk of mean-square error is due to estimation variance.

The mean number of iterations varied between 35 and 50, with a higher value for lower SNR (-30dB) and a smaller value for higher SNR (40dB).

For $m = 800$ and three values of SNR (-30dB, 0dB, and 40dB) Figures 3-5 plot traces of a particular noise realization (different for each SNR). In each figure, we plot four characteristics: top graph plots the estimation error of the clean signal $10\log_{10}\|x^{(t)} - x\|^2$; second graph presents the smallest eigenvalue of $X^{(t)} = [\![x^{(t)}, x^{(t-1)}]\!]$ (which is negative); third graph contains the current estimation error of the clean rank-one, that is $10\log_{10}\|X^{(t)} - xx^*\|_2^2$ (Frobenius norm); bottom graph plots $J_0 = 10\log_{10}\|y - \mathcal{A}(X^{(t)})\|^2$.

We also analyzed the algorithm sensitivity to initialization. Instead of using the eigenpair (5.116,5.117) we initialize by a random vector x^0 together with the eigenvalue (5.117). Results for the three values of redundancy are presented in Figure 6 for 0dB to 40dB range of SNR.

7. Conclusions

Novel necessary conditions for complex valued signal reconstruction from magnitudes of frame coefficients have been presented. Deterministic stability bounds (Lipschitz constants) and stochastic performance bounds (Cramer-Rao lower bound) have been presented. The entire analysis has been done canonically, that is independent of a particular choice of basis. Then an optimization algorithm based on the least-square error has been proposed and analyzed. The algorithm performance has been compared to the theoretical lower bound given by the Cramer-Rao inequality. Remarkably the algorithm performs very well on a large range of SNR. In particular, for high SNR, it seems to converge to the correct signal every time. This behavior suggests the algorithm presented here is able to track the global minimum of (5.112) very well. A future study shall analyze this tracking hypothesis.

Acknowledgments

The author was partially supported by NSF under DMS-1109498 and DMS-1413249 grants. The author thanks the Erwin Schrödinger Institute for the hospitality shown during the special workshop on "Phase Retrieval" in October 2012. Some of the results obtained here were presented at that workshop and later at the Workshop on "Phaseless Reconstruction", UMD, February 2013. The author also thanks Bernhard Bodmann, Jameson Cahill, Martin Ehler, Boaz Nadler, Oren Raz, and Yang Wang for fruitful discussions. He also thanks the annonymous referees for their helpful comments and careful reading of the first draft. Additionally he is grateful to Friedrich Philipp [31] for his comments and for pointing out several errors in the first draft, in particular the dimension of $\mathcal{S}^{1,1}$ is Lemma 3.11. Last but not least, the author thanks the anonymous referees for their comments and their patience with an earlier draft of this paper.

References

[1] B. Alexeev, A. S. Bandeira, M. Fickus, D. G. Mixon, Phase retrieval with polarization, SIAM J. Imaging Sci., 7(1) (2014), 35–66.
[2] B. Alexeev, J. Cahill, D. G. Mixon, Full spark frames, J. Fourier Anal. Appl. 18 (2012), 1167-1194.
[3] R. Balan, P. Casazza, D. Edidin, On signal reconstruction without phase, Appl.Comput.Harmon.Anal. 20 (2006), 345–356.
[4] R. Balan, B. Bodmann, P. Casazza, D. Edidin, Painless reconstruction from Magnitudes of Frame Coefficients, J.Fourier Anal.Applic., 15(4) (2009), 488–501.
[5] R. Balan, On Signal Reconstruction from Its Spectrogram, Proceedings of the CISS Conference, Princeton NJ, May 2010.
[6] R. Balan, Reconstruction of Signals from Magnitudes of Frame Representations, arXiv submission arXiv:1207.1134.
[7] R. Balan and Y. Wang, Invertibility and Robustness of Phaseless Reconstruction, available online arXiv:1308.4718v1, Appl. Comp. Harm. Anal., accepted July 2014.
[8] A. S. Bandeira, J. Cahill, D. Mixon, A. A. Nelson, Saving phase: Injectivity and Stability for phase retrieval, arXiv submission , arXiv: 1302.4618, Appl. Comp. Harm. Anal. 37(1) (2014), 106–125.
[9] R.H. Bates and D. Mnyama, The status of practical fourier phase retrieval, in W.H. Hawkes, ed., Advances in Electronics and Electron Physics, 67 (1986), 1–64.

[10] R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.
[11] B. G. Bodmann, private communications, October 2012 and March 2013.
[12] B. G. Bodmann, N. Hammen, Stable Phase Retrieval with Low-Redundancy Frames, arXiv submission:1302.5487v1, Adv. Comput. Math., accepted 10 April 2014.
[13] E. Candés, T. Strohmer, V. Voroninski, PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming, Communications in Pure and Applied Mathematics vol. 66, 1241–1274 (2013).
[14] E. Candés, Y. Eldar, T. Strohmer, V. Voloninski, Phase Retrieval via Matrix Completion Problem, SIAM J. Imaging Sci., 6(1) (2013), 199-225.
[15] E. J. Candès, X. Li, Solving quadratic equations via PhaseLift when there are about as many equations as unknowns, available online arXiv:1208.6247, Found. of Comput.Math. 14(5) (2012), 1017–1026.
[16] P. Casazza, The art of frame theory, Taiwanese J. Math., 4(2) (2000), 129–202.
[17] P. Casazza and J. Kovačević, Equal-norm tight frames with erasures, Adv. Comp. Math. 18 (2003), 387–430.
[18] P. G. Casazza and G. Kutyniok, Frames of subspaces, Wavelets, frames and operator theory, Contemp. Math., vol. 345, Amer. Math. Soc., Providence, RI, (2004), 87–113.
[19] A. Conca, D. Edidin, M. Hering, C. Vinzant, An algebraic characterization of injectivity in phase retrieval, Appl. Comput. Harmon. Anal. 38 (2015), 346–356.
[20] L. Demanet, P. Hand, Stable optimizationless recovery from phaseless linear measurements, available online arXiv:1208.1803v1, J. Fourier Anal. Appl., 20(1) (2014), 199–221.
[21] Y. Eldar, S. Mendelson, Phase Retrieval: Stability and Recovery Guarantees, available online arXiv:1211.0872v1, Appl.Comp.Harm.Anal. 36(3) (2014), 473–494.
[22] P. R. Halmos, Finite-Dimensional Vector Space, reprint of the 2nd ed. published by Van Nostrand, Princeton, NJ, Springer-Verlag, 1974.
[23] P. Hand, Conditions for Existence of Dual Certificates in Rank One Semidefinite Problems, available online arXiv:1303.1598v2, Commun. Math. Sci, 12(7) (2014), 1363–1378.
[24] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, Signal Reconstruction from Phase and Magnitude, IEEE Trans. ASSP 28(6) (1980), 672–680.
[25] T. Heinosaari, L. Mazzarella, M. M. Wolf, Quantum Tomography under Prior Information, available online arXiv:1109.5478v1, Commun. Math. Phys. 318, 355-374 (2013).
[26] S. M. Kay, Fundamentals of Statistical Signal Processing. I. Estimation Theory, Prentice Hall PTR, 18th Printing, 2010.
[27] T.-Y. Lam, Algebraic Theory of Quadratic Forms, W A Benjamin, 1973.
[28] F. Riesz and B. Sz. Nagy, Functional Analysis, Ungar Publications, 2nd Edition, New York, 1955.
[29] H. Nawab, T. F. Quatieri, and J. S. Lim, Signal Reconstruction from the Short-Time Fourier Transform Magnitude, in Proceedings of ICASSP 1984.
[30] B. Noble, Applied Linear Algebra, Prentice-Hall, Englewood Cliffs, NJ, 1969.
[31] F. Phillip, personal communication, Aug. 2014.
[32] B. Simon, Orthogonal Polynomials on the Unit Circle. Part 2: Spectral Theory, AMS Colloquium Publications, vol. 54, 2004.
[33] O. Raz, N. Dudovich, O. Raz, Vectorial Phase Retrieval of 1-D Signals, IEEE Trans. on Signal Processing, 61(7) (2013), 1632–1643.
[34] I. Waldspurger, A. dAspremont, S. Mallat, Phase recovery, MaxCut and complex semidefinite programming, available online arXiv:1206.0102, Mathematical Programming, 149(1-2) (2015), 47–81.

Department of Mathematics, University of Maryland, College Park MD 20742
E-mail address, R. Balan: rvbalan@math.umd.edu

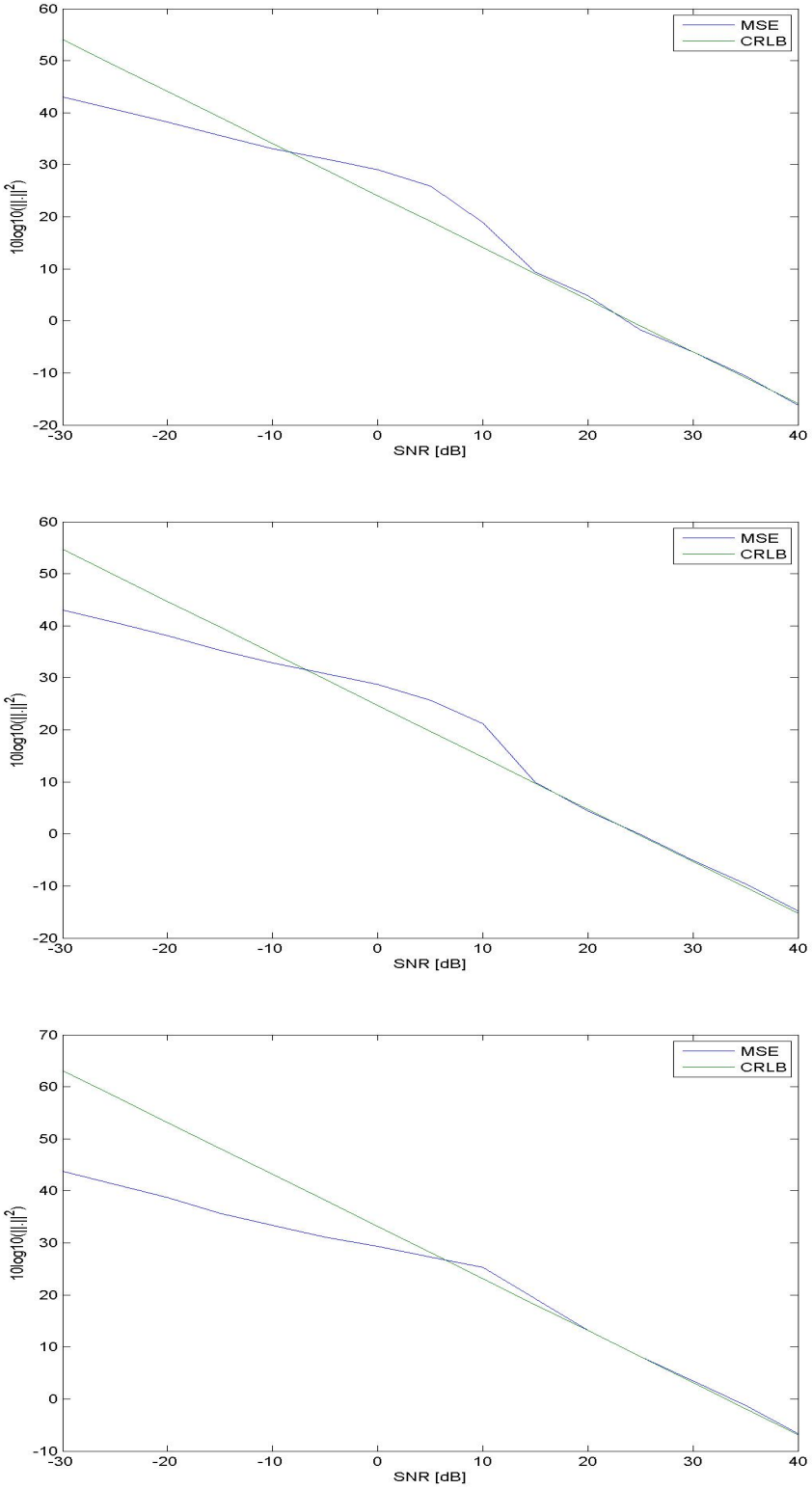

FIGURE 1. Mean-Square Error and CRLB bounds for $m = 800$ (top plot), $m = 600$ (middle plot), and $m = 400$ (bottom plot) when $n = 100$.

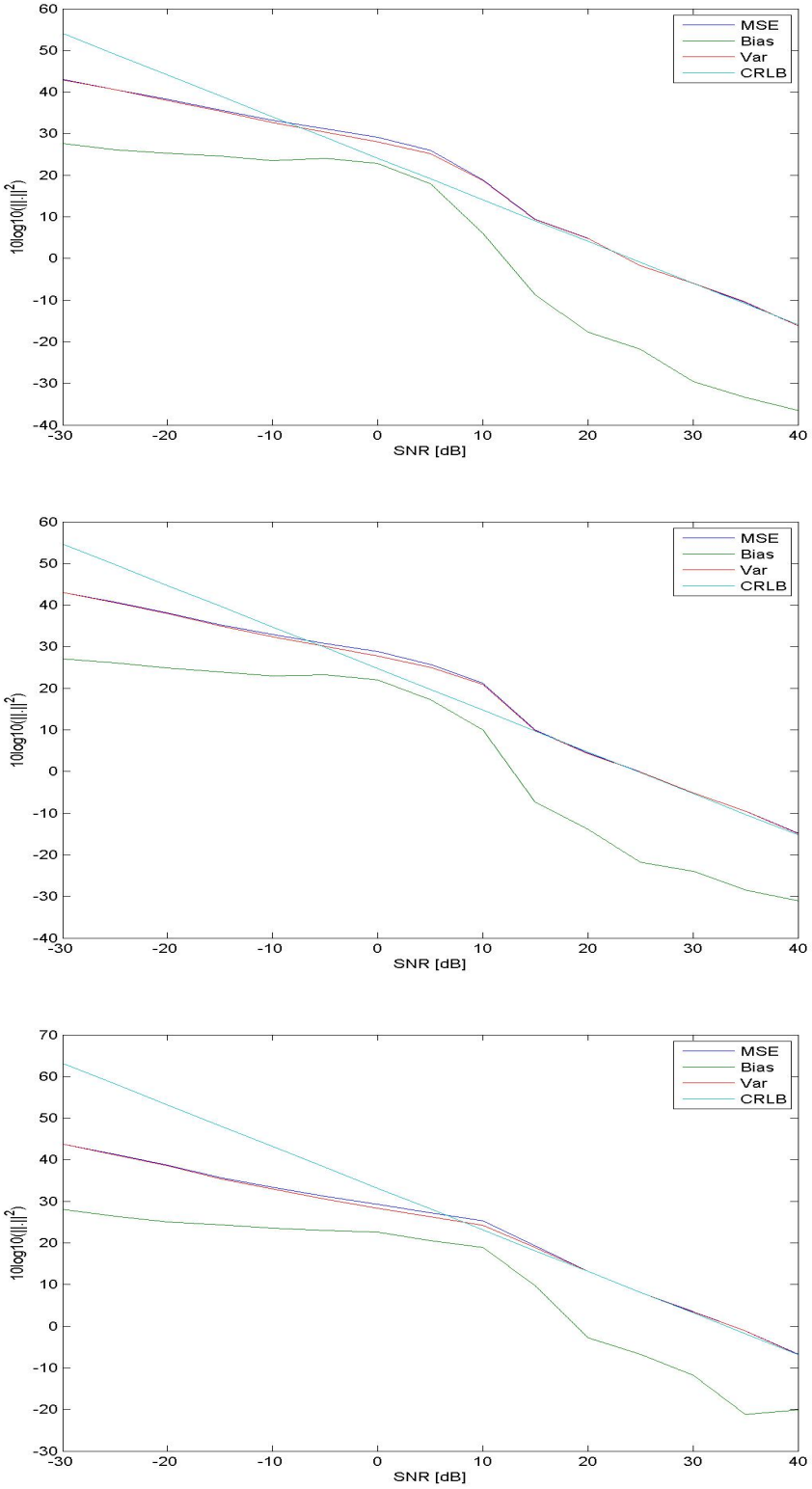

FIGURE 2. Bias and variance components of the mean-square error and CRLB bounds for $m = 800$ (top plot), $m = 600$ (middle plot), and $m = 400$ (bottom plot) when $n = 100$.

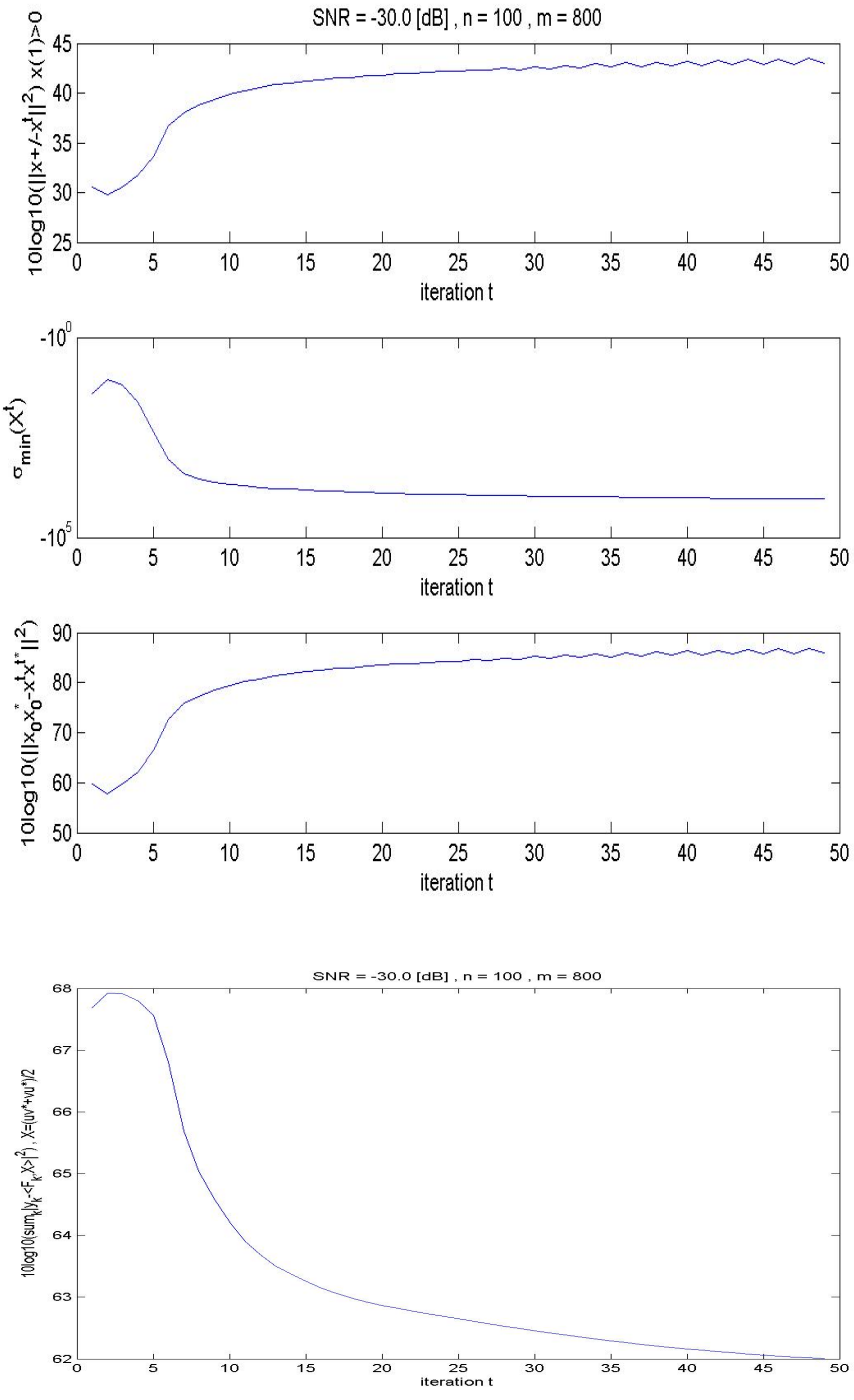

FIGURE 3. Traces for $m = 800$, $n = 100$ and $SNR = -30dB$:

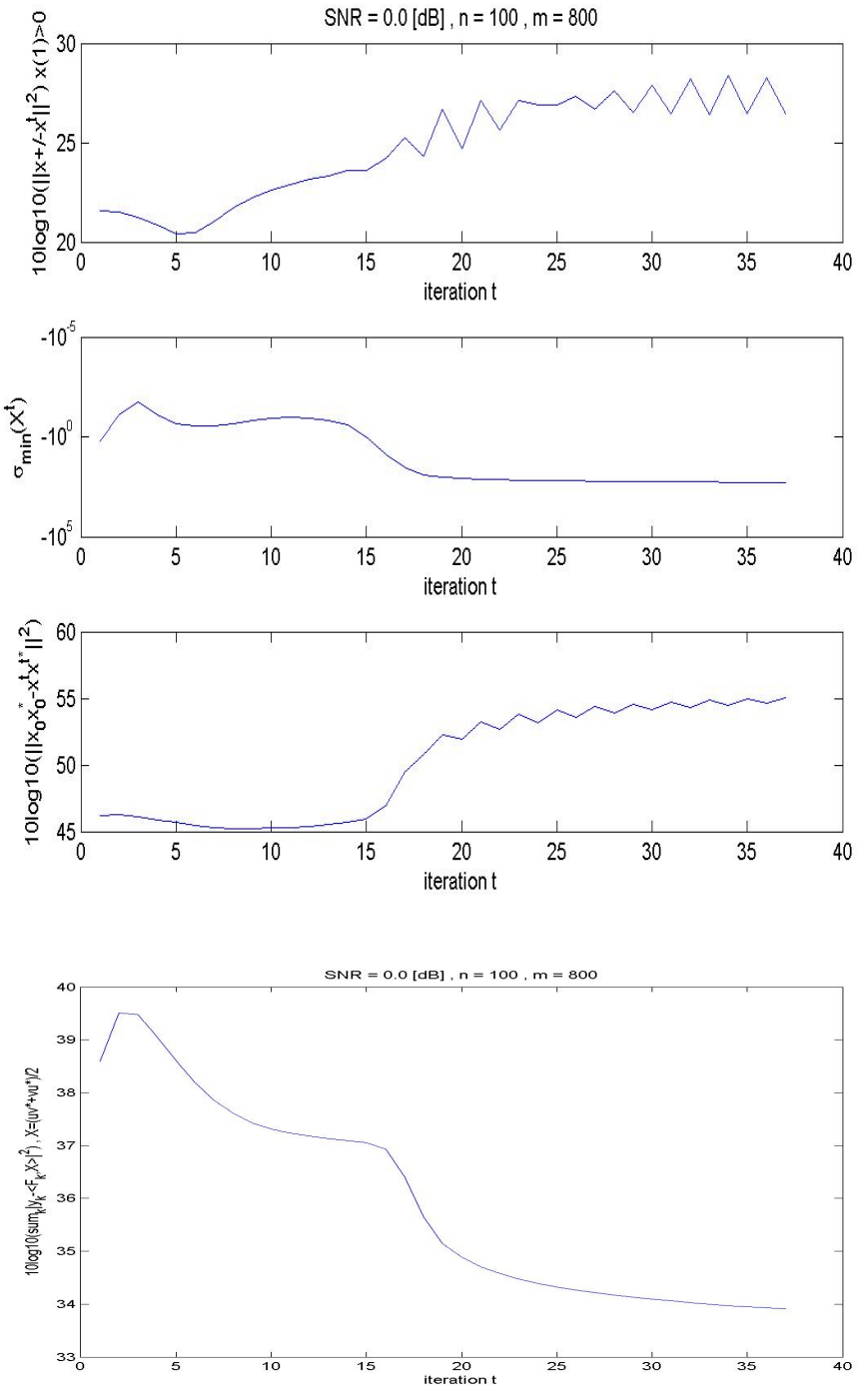

FIGURE 4. Traces for $m = 800$, $n = 100$ and $SNR = 0dB$:

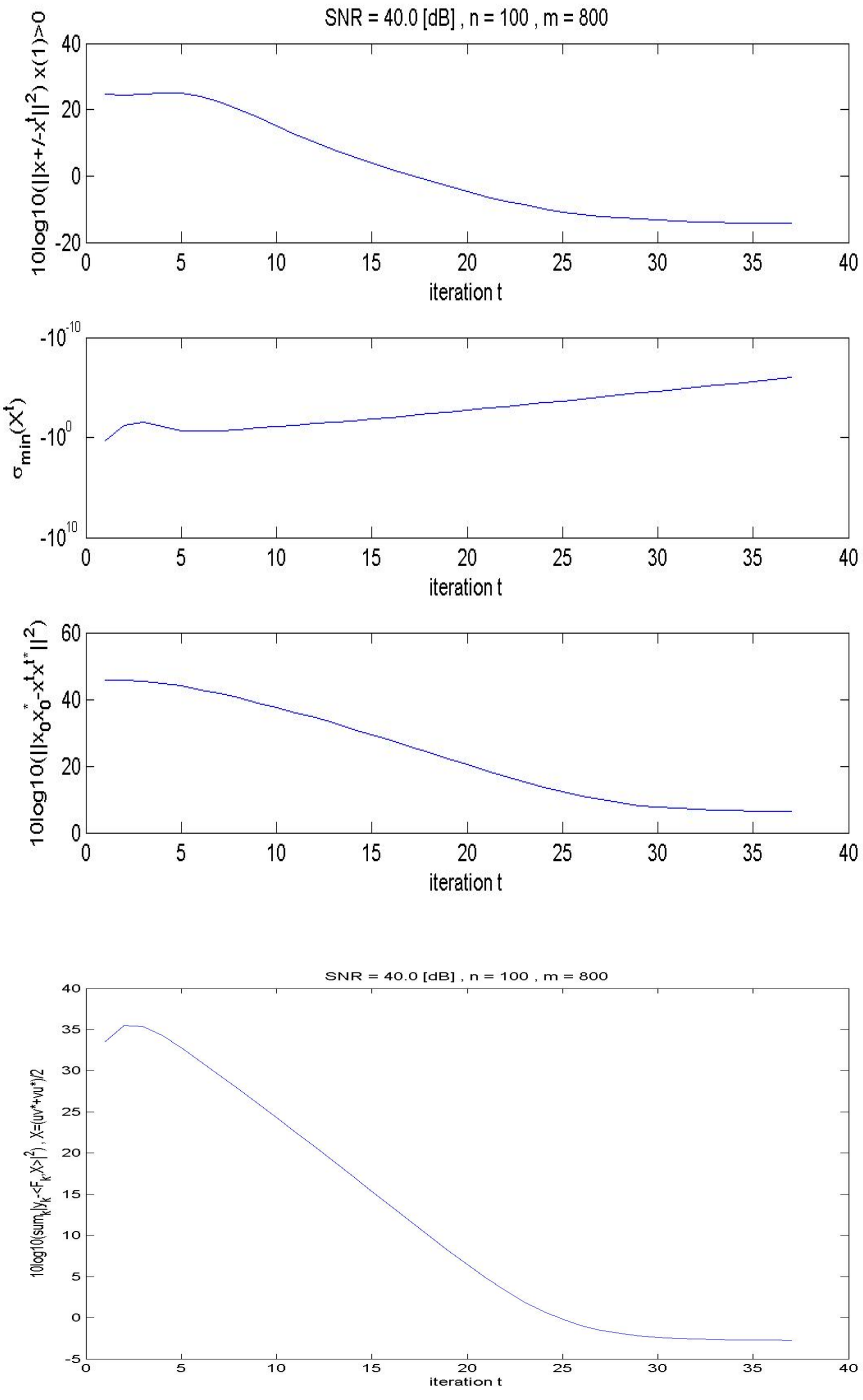

FIGURE 5. Traces for $m = 800$, $n = 100$ and $SNR = 40dB$:

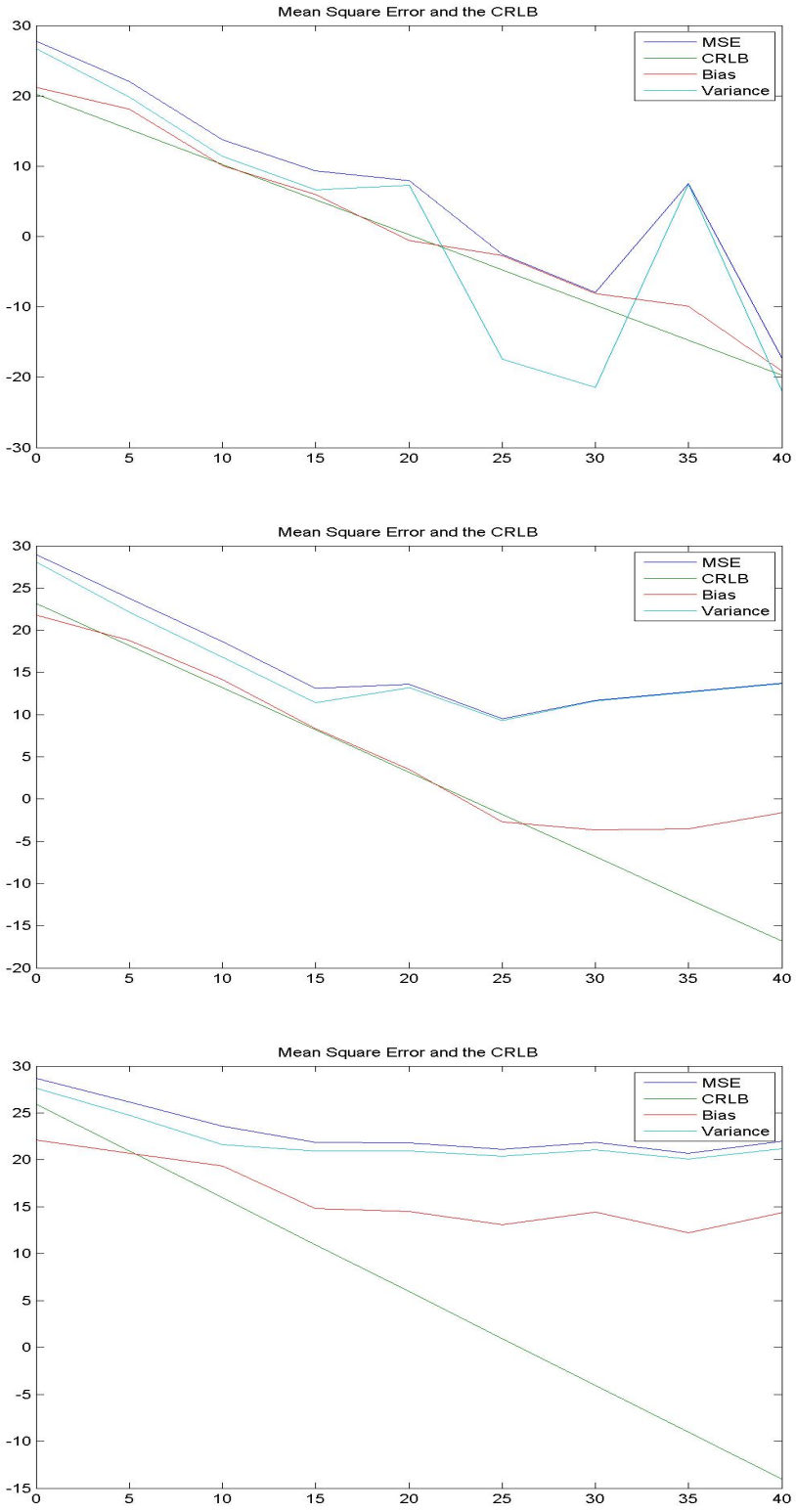

FIGURE 6. Mean-Square Error and CRLB bounds for $m = 800$ (top plot), $m = 600$ (middle plot), and $m = 400$ (bottom plot) when $n = 100$ and random initialization of x^0.