Appl. Comput. Harmon. Anal. 38 (2015) 469–488

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

# Invertibility and robustness of phaseless reconstruction

Radu Balan [a,*], Yang Wang [b]

[a] *Department of Mathematics, Center for Scientific Computation and Mathematical Modeling, University of Maryland, College Park, MD 20742, United States*
[b] *School of Mathematics, Michigan State University, East Lansing, MI 48824, United States*

ARTICLE INFO

*Article history:*
Received 2 September 2013
Received in revised form 7 July 2014
Accepted 12 July 2014
Available online 17 July 2014
Communicated by Jared Tanner

*Keywords:*
Frames
Redundant representations
Phase retrieval
Phaseless reconstruction

ABSTRACT

This paper is concerned with the question of reconstructing a vector in a finite-dimensional real Hilbert space when only the magnitudes of the coefficients of the vector under a redundant linear map are known. We analyze various Lipschitz bounds of the nonlinear analysis map and we establish theoretical performance bounds of any reconstruction algorithm. The discussion of robustness is with respect to random noise and with respect to deterministic perturbations. We show that robust and uniformly stable reconstruction is not achievable with the minimum redundancy for phaseless reconstruction. Robust reconstruction schemes require additional redundancy than the critical threshold.



## 1. Introduction

This paper is concerned with the question of reconstructing a vector $x$ in a finite-dimensional *real* Hilbert space $H$ of dimension $n$ 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 an overall change of sign from the magnitudes $\{|\langle x, f_k\rangle|,\ 1 \le k \le m\}$ where $\mathcal{F} = \{f_1, \ldots, f_m\}$ is a frame (complete system) for $H$.

A previous paper [6] described the importance of the phaseless reconstruction problem. One particular case is when the coefficients are obtained from an Undecimated Wavelet Transform. This case is relevant for instance in some audio and image signal processing applications, as well as in neural computations as performed by the auditory cortex [13].

While [6] presents some necessary and sufficient conditions for reconstruction, the general problem of finding fast/efficient algorithms is still open. In [3] we describe one solution in the case of STFT coefficients.

For vectors in real Hilbert spaces, the reconstruction problem is easily shown to be equivalent to a combinatorial problem. In [7] this problem is further proved to be equivalent to a (nonconvex) optimization problem.

* Corresponding author.
*E-mail addresses:* rvbalan@math.umd.edu (R. Balan), ywang@math.msu.edu (Y. Wang).

http://dx.doi.org/10.1016/j.acha.2014.07.003
1063-5203/

A different approach (which we called the *algebraic approach*) was proposed in [2]. While it applies to both real and complex cases, noiseless and noisy cases, the approach requires solving a linear system of size exponentially in the space dimension. This algebraic approach generalizes the approach in [8] where reconstruction is performed with complexity $O(n^2)$ (plus computation of the principal eigenvector for a matrix of size $n$). However this method requires $m = O(n^2)$ frame vectors.

Recently the authors of [10] developed a convex optimization algorithm (a SemiDefinite Program called *PhaseLift*) and proved its ability to perform exact reconstruction in the absence of noise, as well as its stability under noise conditions. In a separate paper [11], the authors further developed a similar algorithm in the case of windowed DFT transforms. Inspired by the PhaseLift and MaxCut algorithms, but operating in the coefficients space, the authors of [16] proposed a SemiDefinite Program called *PhaseCut*. They show the algorithm yields the exact solution in the absence of noise under similar conditions as PhaseLift.

The paper [4] presents an iterative regularized least-square algorithm for inverting the nonlinear map and compares its performance to a Cramer–Rao lower bound for this problem in the real case. The paper also presents some new injectivity results which are incorporated into this paper.

A different approach is proposed in [1]. There the authors use a 4-term polarization identity together with a family of spectral expander graphs to design a frame of bounded redundancy ($\frac{m}{n} \leq 236$) that yields an exact reconstruction algorithm in the absence of noise.

The authors of [14] study several robustness bounds to the phase recovery problem in the real case. However their approach is different from ours in several respects. First they consider a probabilistic setup of this problem, where data $x$ and frame vectors $f_j$'s are random vectors with probabilities from a class of subgaussian distributions. Additionally, their focus is on classes of $k$-sparse signals. In our paper we analyze stability bounds of reconstruction for a fixed frame using deterministic analytic tools. After that we present asymptotic behavior of these bounds for random frames.

Finally, the authors of [9] analyze the phaseless reconstruction problem for both the real and complex case. In the real case the authors obtain the exact upper Lipschitz constant for the nonlinear map $\alpha_{\mathcal{F}}$, namely $\sqrt{B}$ where $B$ is the upper frame bound. For the lower Lipschitz constant, they give an estimate between two computable singular eigenvalues. Our results have overlaps with their results. However, in our paper we improve the lower Lipschitz constant by giving its exact value. There are some significant differences between this paper and [9]. In addition to studying of the Lipschitz property of the map $\alpha_{\mathcal{F}}$ we focus also on two related but different settings. First we study the robustness of the reconstruction given a fixed error allowance in measurements. Second we also consider the Lipschitz property of the map ${\alpha_{\mathcal{F}}}^2$. The authors of [9] point out that the map ${\alpha_{\mathcal{F}}}^2$ is not bi-Lipschitz. However in our paper we show ${\alpha_{\mathcal{F}}}^2$ becomes bi-Lipschitz for a different metric on the domain. With this metric (the one induced by the nuclear norm on the set of symmetric operators) the nonlinear map ${\alpha_{\mathcal{F}}}^2$ is bi-Lipschitz with constants indicated in Theorem 4.5. Furthermore the same conclusion holds true in the complex case, although this will be studied elsewhere.

The organization of the paper is as follows. Section 2 formally defines the problem and reviews existing inversion results in the real case. Section 3 establishes information theoretic performance bounds, namely the Cramer–Rao lower bound. Section 4 contains robustness measures of any reconstruction algorithm. Section 5 presents a stochastic analysis of these bounds. Section 6 presents a numerical example and is followed by references.

## 2. Background

Let us denote by $H = \mathbb{R}^n$ the n-dimensional real Hilbert space $\mathbb{R}^n$ with scalar product $\langle , \rangle$. Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be a spanning set of $m$ vectors in $H$. In finite dimension (as it is the case here) such a set forms a *frame*. In the infinite dimensional case, the concept of frame involves a stronger property than completeness (see for instance [12]). We review additional terminology and properties which remain still

true in the infinite dimensional setting. The set $\mathcal{F}$ is a frame if and only if there are two positive constants $0 < A \leq B < \infty$ (called frame bounds) so that

$$A\|x\|^2 \leq \sum_{k=1}^{m} \big|\langle x, f_k\rangle\big|^2 \leq B\|x\|^2. \tag{2.1}$$

When we can choose $A = B$ the frame is said *tight*. For $A = B = 1$ the frame is called *Parseval*. The *frame matrix* corresponding to $\mathcal{F}$ is defined as $F = [f_1, f_2, \ldots, f_m]$ with the vectors $f_j \in \mathcal{F}$ as its columns. We shall frequently identify $\mathcal{F}$ with its corresponding frame matrix $F$. The largest $A$ and smallest $B$ in (2.1) are called the *lower frame bound* and *upper frame bound* of $\mathcal{F}$, and they are given by

$$A = \lambda_{\max}\big(FF^*\big) = \sigma_1^2(F), \qquad B = \lambda_{\min}\big(FF^*\big) = \sigma_n^2(F) \tag{2.2}$$

where $\lambda_{\max}, \lambda_{\min}$ denote the largest and smallest eigenvalues respectively, while $\sigma_1, \sigma_n$ denote the first and $n$-th singular values respectively. A set of vectors $\mathcal{F}$ of the $n$-dimensional Hilbert space $H$ is said to be *full spark* if any subset of $n$ vectors is linearly independent.

For a vector $x \in H$, the collection of coefficients $\{\langle x, f_j\rangle : \ 1 \leq j \leq m\}$ represents the analysis map of vector $x$ given by the frame $\mathcal{F}$, and from which $x$ can be completely reconstructed. In the phaseless reconstruction problem, we ask the following question: Can $x$ be reconstructed from $\{|\langle x, f_j\rangle| : \ 1 \leq j \leq m\}$? Consider the following equivalence relation $\sim$ on $H$: $x \sim y$ if and only if $y = cx$ for some unimodular constant $c$, $|c| = 1$. Since we focus on the real vector space $H = \mathbb{R}^n$, we have $x \sim y$ if and only if $x = \pm y$. Clearly the phaseless reconstruction problem cannot distinguish $x$ and $y$ if $x \sim y$, so we will be looking at reconstruction on $\hat{H} := H/\sim = \mathbb{R}^n/\sim$ whose elements are given by equivalent classes $\hat{x} = \{x, -x\}$ for $x \in \mathbb{R}^n$. The analogous analysis map for phaseless reconstruction is the following nonlinear map

$$\alpha_{\mathcal{F}} : \hat{H} \to \mathbb{R}_+^m, \qquad \alpha_{\mathcal{F}}(\hat{x}) = \big[\big|\langle x, f_1\rangle\big|, \big|\langle x, f_2\rangle\big|, \ldots, \big|\langle x, f_m\rangle\big|\big]^T. \tag{2.3}$$

Note that $\alpha_{\mathcal{F}}$ can also be viewed as a map from $\mathbb{R}^n$ to $\mathbb{R}_+^m$. Throughout the paper we will not make an explicit distinction unless such a distinction is necessary.

Thus the phaseless reconstruction problems aims to reconstruct $\hat{x} \in \hat{H}$ from the map $\alpha_{\mathcal{F}}(x)$. We say a frame $\mathcal{F}$ is *phase retrievable* if one can reconstruct $\hat{x} \in \hat{H}$ for all $\hat{x}$, or in other words, $\alpha_{\mathcal{F}}$ is injective on $\hat{H}$. The main objective of this paper is to analyze robustness and stability of the inversion map, and to give performance bounds of any reconstruction algorithm.

Before proceeding further we first review existing results on injectivity of the nonlinear map $\alpha_{\mathcal{F}}$. In general a subset $Z$ of a topological space is said *generic* if its open interior is dense. However in the following statements, the term *generic* refers to Zarisky topology: a set $Z \subset \mathbb{K}^{n\times m} = \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ is said *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 $\mathbb{K} = \mathbb{R}$).

**Theorem 2.1.** *Let $\mathcal{F}$ be a frame in $H = \mathbb{R}^n$ with $m$ elements. Then the following hold true:*

1. *The frame $\mathcal{F}$ is phase retrievable in $\hat{H}$ if and only if for any disjoint partition of the frame set $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$, either $\mathcal{F}_1$ spans $\mathbb{R}^n$ or $\mathcal{F}_2$ spans $\mathbb{R}^n$.*
2. *If $\mathcal{F}$ is phase retrievable in $\hat{H}$ then $m \geq 2n-1$. Furthermore, for a generic $\mathcal{F}$ with $m \geq 2n-1$ the map $\alpha_{\mathcal{F}}$ is phase retrievable in $\hat{H}$.*
3. *Let $m = 2n-1$. Then $\mathcal{F}$ is phase retrievable in $\hat{H}$ if and only if $\mathcal{F}$ is full spark.*

4. *Let*

$$a_0 := \min_{\|x\|=\|y\|=1} \sum_{j=1}^{m} \big|\langle x, f_j\rangle\big|^2 \big|\langle y, f_j\rangle\big|^2 \geq 0, \tag{2.4}$$

*so that*

$$\sum_{k=1}^{m} \big|\langle x, f_k\rangle\big|^2 \big|\langle y, f_k\rangle\big|^2 \geq a_0 \|x\|^2 \|y\|^2. \tag{2.5}$$

*Then $\mathcal{F}$ is phase retrievable on $\hat{H}$ if and only if $a_0 > 0$.*

5. *For any $x \in \mathbb{R}^n$ define the matrix $R(x)$ by*

$$R(x) := \sum_{j=1}^{m} \big|\langle x, f_j\rangle\big|^2 f_j f_j^*. \tag{2.6}$$

*Let $\lambda_{min}(R(x))$ denote the smallest eigenvalue of $R(x)$, and let $a_0 = \min_{\|x\|=1} \lambda_{\min}(R(x))$. Equivalently let $a_0$ be the largest constant so that $R(x) \geq a_0\|x\|^2 I$ for all $x \in H$, where $I$ is the identity matrix.*
*Then $\mathcal{F}$ is phase retrievable on $\hat{H}$ if and only if $a_0 > 0$.*
*Additionally the constant $a_0$ introduced here is the same as the constant $a_0$ given by (2.4).*

The results (1)–(3) are in [6], and (4)–(5) are in [4].

## 3. Information theoretic performance bounds

In this section we derive expressions for the *Fisher Information Matrix* and obtain performance bounds for reconstruction algorithms in the noisy case.

Consider the following noisy measurement process:

$$y_k = \big|\langle x, f_k\rangle\big|^2 + \nu_k, \quad \nu_k \sim \mathcal{N}\big(0, \sigma^2\big), \quad 1 \leq k \leq m \tag{3.1}$$

where the noise model is AWGN (additive white Gaussian noise): each random variable $\nu_k$ is independent and normally distributed with zero mean and $\sigma^2$ variance.

Consider the noiseless case first (that is $\nu_k = 0$). Obviously one cannot obtain the exact vector $x \in H$ due to the global sign ambiguity. Instead the best outcome is to identify (that is, to estimate) the class $\hat{x} = \{x, -x\}$ from $\alpha_{\mathcal{F}}(x)$. As such, we fix a disjoint partition of the punctured Hilbert space $H$, $\mathbb{R}^n \setminus \{0\} = \Omega_1 \cup \Omega_2$, such that $\Omega_2 = -\Omega_1$. We make the choice that the vector $x$ belongs to $\Omega_1$. Hence any estimator of $x$ is a map $\omega : \mathbb{R}^m \longrightarrow \Omega_1 \cup \{0\}$. Denote by $\mathring{\Omega}_1$ its interior as a subset of $\mathbb{R}^n$. Such a decomposition is, for example

$$\Omega_1 = \bigcup_{k=1}^{n} \big\{x \in \mathbb{R}^n : \ x_k \geq 0, \ x_j = 0 \text{ for } j < k\big\}.$$

Note its interior is given by $\mathring{\Omega}_1 = \{x \in \mathbb{R}^n, \ x_1 > 0\}$.

Under these assumptions we compute the Fisher Information matrix (see [15]). This is given by

$$\big(\mathbb{I}(x)\big)_{k,j} = \mathbb{E}\big[\big(\nabla \log L(x)\big)\big(\nabla \log L(x)\big)^T\big] \tag{3.2}$$

where the likelihood function $L(x)$ is given by

$$L(x) = p(y|x) = \frac{1}{(2\pi)^{m/2}\sigma^m} \exp\left( -\frac{1}{2\sigma^2} \sum_{k=1}^{m} \left| y_k - \left| \langle x, f_k \rangle \right|^2 \right|^2 \right). \tag{3.3}$$

After some algebra (see [4]) we obtain

$$\mathbb{I}(x) = \frac{4}{\sigma^2} R(x), \qquad R(x) = \sum_{j=1}^{m} \left| \langle x, f_j \rangle \right|^2 f_j f_j^T. \tag{3.4}$$

Note the matrix $R(x)$ is exactly the same as the matrix introduced in (2.6). Thus we obtain the following results:

**Theorem 3.1.** *The frame $\mathcal{F}$ is phase retrievable if and only if the Fisher information matrix $\mathbb{I}(x)$ is invertible for any $x \neq 0$.*

*When $\mathcal{F}$ is phase retrievable let $a_0$ be the positive constant introduced in (2.4). Then*

$$\mathbb{I}(x) \geq \frac{4a_0}{\sigma^2} \|x\|^2 I \tag{3.5}$$

*where $I$ is the $n \times n$ identity operator.*

This allows to state the following performance bound result (see [15] for details on the Cramer–Rao lower bound).

**Theorem 3.2.** *Assume $x \in \mathring{\Omega}_1$. Let $\omega : \mathbb{R}^m \to \Omega_1$ be any unbiased estimator for $x$. Then its covariance matrix is bounded below by the Cramer–Rao lower bound:*

$$\mathrm{Cov}\big[\omega(y)\big] \geq \big(\mathbb{I}(x)\big)^{-1} = \frac{\sigma^2}{4}\big(R(x)\big)^{-1}. \tag{3.6}$$

*Furthermore, any efficient estimator (that is, any unbiased estimator $\omega$ that achieves the Cramer–Rao Lower Bound (3.6)) has the covariance matrix bounded from above by*

$$\mathrm{Cov}\big[\omega(y)\big] \leq \frac{\sigma^2}{4a_0\|x\|^2} I \tag{3.7}$$

*and Mean-Square error bounded above by*

$$\mathrm{MSE}(\omega) = \mathbb{E}\big[\big\|\omega(y) - x\big\|^2\big] \leq \frac{n\sigma^2}{4a_0\|x\|^2}. \tag{3.8}$$

**Remark 3.3.** We point out the importance of the constant $a_0$ introduced in (2.4). On the one hand it represents a necessary and sufficient condition for phase retrievability as stated in Theorem 2.1. On the other hand the above results prove that $a_0$ provides also a bound for the Fisher Information matrix and hence a bound for any efficient estimator of $\hat{x}$. The larger this constant $a_0$, the smaller the variance of the efficient estimator. As we prove in the next section, the same constant $a_0$ represents the lower Lipschitz bound for the map $\alpha_{\mathcal{F}^2}$ (4.13) considered between $(\hat{H}, d_1)$ and the Euclidean space $(\mathbb{R}^m, \|\cdot\|)$ – see Theorem 4.5.

Additionally, similar expressions involving the bound $a_0$ occur in the complex case as well. Both the stochastic bound above and the bi-Lipschitz result in Theorem 4.5 can be extended to the complex case – see [5].

## 4. Robustness measures for reconstruction

In this section we analyze the robustness of deterministic phaseless reconstruction. Additionally we connect the constant $a_0$ introduced earlier in Theorem 2.1 and used in Theorem 3.1 to quantities directly computable from the frame $\mathcal{F}$.

Our approach is to analyze the stability in the worst case scenario, for which we consider the following measures. Denote $d(x,y) := \min(\|x-y\|, \|x+y\|)$. For any $x \in \mathbb{R}^n$ and $\varepsilon > 0$ define

$$Q_\varepsilon(x) = \max_{\{y:\|\alpha_{\mathcal{F}}(x)-\alpha_{\mathcal{F}}(y)\|\le\varepsilon\}} \frac{d(x,y)}{\varepsilon}. \tag{4.1}$$

The size of $Q_\varepsilon(x)$ measures the worst case stability of the reconstruction for the vector $x$, under the assumption that the total noise level is controlled by $\varepsilon$. We also study the global stability by analyzing the measures

$$q_\varepsilon := \max_{\|x\|=1} Q_\varepsilon(x), \qquad q_0 := \limsup_{\varepsilon\to 0} q_\varepsilon, \qquad q_\infty := \sup_{\varepsilon>0} q_\varepsilon. \tag{4.2}$$

Here $\|.\|$ denotes usual Euclidian norm. Note that $Q_\varepsilon(x)$ has the scaling property $Q_\varepsilon(x) = Q_{|c|\varepsilon}(cx)$ for any real $c \neq 0$. Thus it is natural to focus on unit vectors $x$.

We introduce now some quantities that play key roles in the estimation of these robustness measures. For the frame $\mathcal{F}$ let $F = [f_1, f_2, \cdots, f_m]$ be its frame matrix. Denote by $\mathcal{F}[S] = \{f_k,\ k \in S\}$ the subset of $\mathcal{F}$ indexed by a subset $S \subseteq \{1,2,\dots,m\}$, and by $F_S$ the frame matrix corresponding to $\mathcal{F}[S]$ (which is the matrix with vectors in $\mathcal{F}[S]$ as its columns). Set

$$A[S] := \sigma_n^2(F_S) = \lambda_{\min}\big(F_S F_S^*\big), \tag{4.3}$$

where as usual $\sigma_n$ and $\lambda_{\min}$ denote the $n$-th singular value and the minimal eigenvalue, respectively. Note that $A[S]$ is in fact the lower frame bound of $\mathcal{F}[S]$.

Let $\mathcal{S}$ denote the collection of subsets $S$ of $\{1,2,\dots,m\}$ so that $\dim(\mathrm{span}(\mathcal{F}[S^c])) < n$, where $S^c = \{1,2,\dots,m\} \setminus S$ is the complement of $S$. In other words, $\mathrm{rank}(F_{S^c}) < n$. Denote by $\Delta$ and $\omega$ the following expressions:

$$\Delta = \min_S \sqrt{A[S] + A\big[S^c\big]} \tag{4.4}$$

$$\omega = \min_{S\in\mathcal{S}} \sigma_n(F_S). \tag{4.5}$$

All of them depend of course on $\mathcal{F}$. However since we fix $\mathcal{F}$ throughout the paper, we shall not explicitly reference $\mathcal{F}$ in the notation for simplicity as there will not be any confusion. Clearly

$$\Delta \le \omega. \tag{4.6}$$

**Proposition 4.1.** *Let $\varepsilon > 0$. Then the stability measurement function $Q_\varepsilon(x)$ is given by*

$$Q_\varepsilon(x) = \frac{1}{\varepsilon} \max_{(w_1,w_2)\in\Upsilon} \min\big\{\|w_1\|, \|w_2\|\big\} \tag{4.7}$$

*where the constraint set $\Upsilon$ is given by*

$$\Upsilon = \left\{ (w_1, w_2) \mid \frac{1}{2}(w_1 + w_2) = x,\ \sum_{j=1}^{m} \min\big(|\langle f_j, w_1\rangle|^2, |\langle f_j, w_2\rangle|^2\big) = \|F_S^* w_1\|^2 + \|F_{S^c}^* w_2\|^2 \le \varepsilon^2 \right\}, \tag{4.8}$$

*where* $S := S(w_1, w_2) = \{j :\ |\langle f_j, w_1\rangle| \le |\langle f_j, w_2\rangle|\}$.

**Proof.** For any $x, y \in \mathbb{R}^n$ let $w_1 = x + y$ and $w_2 = x - y$. Then $x = \frac{1}{2}(w_1 + w_2)$ and $y = \frac{1}{2}(w_1 - w_2)$. It is easy to check that for $S = \{j :\ |\langle f_j, w_1\rangle| \le |\langle f_j, w_2\rangle|\}$ we have

$$\big|\langle f_j, x\rangle\big| - \big|\langle f_j, y\rangle\big| = \begin{cases} \pm\langle f_j, w_1\rangle & j \in S, \\ \pm\langle f_j, w_2\rangle & j \in S^c. \end{cases}$$

In other words,

$$\big|\langle f_j, x\rangle\big| - \big|\langle f_j, y\rangle\big| = \min\big(|\langle f_j, w_1\rangle|, |\langle f_j, w_2\rangle|\big). \tag{4.9}$$

Let $F$ be the frame matrix of $\mathcal{F}$. We thus have

$$\big\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\big\|^2 = \sum_{j\in S} \big|\langle f_j, w_1\rangle\big|^2 + \sum_{j\in S^c} \big|\langle f_j, w_2\rangle\big|^2 = \|F_S^* w_1\|^2 + \|F_{S^c}^* w_2\|^2.$$

Note that $d(x, y) = \min(\|w_1\|, \|w_2\|)$. The proposition now follows. □

The above proposition allows us to establish the following stability result for the worst case scenario.

**Theorem 4.2.** *Assume that the frame $\mathcal{F}$ is phase retrievable. Let $A > 0$ be the lower frame bound for the frame $\mathcal{F}$ and let* $\tau := \min\{\sigma_n(F_S) :\ S \subseteq \{1, \dots, m\},\ \operatorname{rank}(F_S) = n\}$.

(A) *For any $\varepsilon > 0$ we have*

$$\min\left\{\frac{1}{\varepsilon}, \frac{1}{\omega}\right\} \le q_\varepsilon \le \frac{1}{\Delta}. \tag{4.10}$$

(B) *If $\varepsilon < \tau$ then $q_\varepsilon = \frac{1}{\omega}$. Consequently $q_0 = \frac{1}{\omega}$.*

(C) *For any nonzero $x \in \mathbb{R}^n$ and any $0 < \varepsilon < \Delta_x$ we have*

$$Q_\varepsilon(x) = \frac{1}{\sqrt{A}}, \tag{4.11}$$

*where*

$$\Delta_x := \frac{2\tau}{\max(\|f_j\|) + \tau} \min\big\{\big|\langle f_j, x\rangle\big| :\ \langle f_j, x\rangle \ne 0\big\}.$$

(D) *The upper bound $q_\infty$ equals the reciprocal of $\Delta$:*

$$q_\infty = \frac{1}{\Delta}. \tag{4.12}$$

**Proof.** To prove (A) we first establish the upper bound in (4.10). Let $x \in \mathbb{R}^n$. By Proposition 4.1 we have

$$Q_\varepsilon(x) = \frac{1}{\varepsilon} \max_{w_1,w_2} \min\{\|w_1\|, \|w_2\|\}$$

under the constraints $\frac{1}{2}(w_1 + w_2) = x$ and

$$\left\|F_S^* w_1\right\|^2 + \left\|F_{S^c}^* w_2\right\|^2 \le \varepsilon^2$$

for some $S$. Now assume without loss of generality that $\|w_1\| \le \|w_2\|$. Then

$$\begin{aligned}\frac{\varepsilon^2}{\|w_1\|^2} &\ge \frac{\|F_S^* w_1\|^2 + \|F_{S^c}^* w_2\|^2}{\|w_1\|^2} \\ &\ge \sigma_n^2(F_S) + \sigma_n^2(F_{S^c}) \frac{\|w_2\|^2}{\|w_1\|^2} \\ &\ge \Delta.\end{aligned}$$

It follows that

$$\frac{1}{\varepsilon} \min\{\|w_1\|, \|w_2\|\} \le \frac{1}{\Delta}.$$

Thus $Q_\varepsilon(x) \le \frac{1}{\Delta}$.

To establish the lower bound in (4.10) we construct for any $\varepsilon > 0$ an $x \in \mathbb{R}^n$ and vectors $w_1, w_2$ satisfying the imposed constraints. Let $S$ be a subset of $\{1, 2, \dots, m\}$ such that $\operatorname{rank}(F_{S^c}) < n$ and $\sigma_n(F_S) = \omega$. Choose $v_1, v_2 \in \mathbb{R}^n$ with the property $\|v_1\| = \|v_2\| = 1$ and

$$\left\|F_S^* v_1\right\| = \omega, \qquad F_{S^c}^* v_2 = 0.$$

Set

$$t = \min\left\{\frac{\varepsilon}{\omega}, 1\right\}, \quad \text{and} \quad w_1 = t v_1.$$

Hence $\|w_1\| = t \le 1$. Now we select an $s \in \mathbb{R}$ so that $\|w_1 + s v_2\| = 2$. This is always possible since $s \mapsto \|w_1 + s v_2\|$ is continuous and $\|w_1 + 0 v_2\| = t \le 1 \le 2 \le \|w_1 + 3 v_2\|$. Set $w_2 = s v_2$ so $\|w_1 + w_2\| = 2$. We have

$$|s| = \|s v_2\| \ge \|w_1 + s v_2\| - \|w_1\| = 2 - t \ge 1.$$

Thus $\|w_2\| \ge \|w_1\|$. Now let

$$x = \frac{1}{2}(w_1 + w_2) \quad \text{and} \quad y = \frac{1}{2}(w_1 - w_2).$$

We have then

$$\begin{aligned}\left\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\right\|^2 &= \sum_{j=1}^m \min\left(\left|\langle f_j, w_1\rangle\right|^2, \left|\langle f_j, w_2\rangle\right|^2\right) \\ &\le \sum_{j\in S} \left|\langle f_j, w_1\rangle\right|^2 + \sum_{j\in S^c} \left|\langle f_j, w_2\rangle\right|^2 \\ &= t^2\omega^2 \le \varepsilon^2.\end{aligned}$$

Furthermore

$$d(x,y) = \min(\|w_1\|, \|w_2\|) = \|w_1\| = t.$$

Hence for this $x$ we have

$$Q_\varepsilon(x) \geq \frac{d(x,y)}{\varepsilon} = \min\left\{\frac{1}{\varepsilon}, \frac{1}{\omega}\right\}.$$

It follows that $q_\varepsilon \geq \min\{\frac{1}{\varepsilon}, \frac{1}{\omega}\}$. Now by taking $\varepsilon > 0$ sufficiently small we have $q_\varepsilon \geq \frac{1}{\omega}$.

We now prove (B). Assume that $\varepsilon \leq \min\{\sigma_n(F_S) : \ \operatorname{rank}(F_S) = n\}$. Then clearly we have $\varepsilon \leq \omega$. Thus by (4.10) we have $q_\varepsilon \geq \frac{1}{\omega}$. Again for each $x \in \mathbb{R}^n$ with $\|x\| = 1$ we consider $w_1, w_2$ for the estimation of $q_\varepsilon(x)$. The constraint $\|w_1 + w_2\| = 2$ implies either $\|w_1\| \geq 1$ or $\|w_2\| \geq 1$. Without loss of generality we assume that $\|w_1\| \geq 1$. For the constraint $\|F_S^* w_1\|^2 + \|F_{S^c}^* w_2\|^2 \leq \varepsilon^2$ for some $S$, assume that $\operatorname{rank}(F_S) = n$ then we have

$$\left\|F_S^* w_1\right\| \geq \sigma_n(F_S)\|w_1\| \geq \min\left\{\sigma_n(F_S):\ \operatorname{rank}(F_S) = n\right\} > \varepsilon.$$

This is a contradiction. So $\operatorname{rank}(F_S) < n$ and hence

$$\varepsilon^2 \geq \left\|F_S^* w_1\right\|^2 + \left\|F_{S^c}^* w_2\right\|^2 \geq \left\|F_{S^c}^* w_2\right\|^2 \geq \omega^2 \|w_2\|^2.$$

Thus $\|w_2\| \leq \frac{\varepsilon}{\omega}$. Proposition 4.1 now yields $q_\varepsilon = \frac{1}{\omega}$, proving part (B).

Now we prove (C). We go back to the formulation in Proposition 4.1.

$$Q_\varepsilon(x) = \frac{1}{\varepsilon} \max_{w_1, w_2} \min\left\{\|w_1\|, \|w_2\|\right\}$$

under the constraints $\frac{1}{2}(w_1 + w_2) = x$ and

$$\left\|F_S^* w_1\right\|^2 + \left\|F_{S^c}^* w_2\right\|^2 \leq \varepsilon^2$$

where $S := S(w_1, w_2) = \{j:\ |\langle f_j, w_1\rangle| \leq |\langle f_j, w_2\rangle|\}$. Since $\alpha_{\mathcal{F}}$ is injective, either $\operatorname{rank}(F_S) = n$ or $\operatorname{rank}(F_{S^c}) = n$ by Theorem 2.1 (1). Without loss of generality we assume $\operatorname{rank}(F_S) = n$. Thus $\varepsilon \geq \|F_S^* w_1\| \geq \tau\|w_1\|$. So $\|w_1\| \leq \varepsilon/\tau$. We show that for any $k \in S^c$ we must have $\langle f_k, x\rangle = 0$. Assume otherwise and write $w_2 = 2x - w_1$, $L_x := \min\{|\langle f_j, x\rangle| :\ \langle f_j, x\rangle \neq 0\}$. Then

$$\left|\langle f_k, w_2\rangle\right| \geq 2\left|\langle f_k, x\rangle\right| - \left|\langle f_k, w_1\rangle\right| \geq 2L_x - \max\left(\|f_j\|\right)\|w_1\| \geq 2L_x - \max\left(\|f_j\|\right)\frac{\varepsilon}{\tau} > \varepsilon.$$

This is a contradiction. Thus for $k \in S^c$ we have $\langle f_k, x\rangle = 0$ and

$$\left|\langle f_j, w_2\rangle\right| = \left|\langle f_j, 2x - w_1\rangle\right| = \left|\langle f_j, w_1\rangle\right|.$$

It follows that

$$\left\|F_S^* w_1\right\|^2 + \left\|F_{S^c}^* w_2\right\|^2 = \left\|F^* w_1\right\|^2 \leq \varepsilon^2.$$

Thus $\|w_1\| \leq \varepsilon/\sqrt{A}$ and hence $Q_\varepsilon(x) \leq \frac{1}{\sqrt{A}}$. Now we show the bound can be achieved. Let $w_1$ satisfy $\|F^* w_1\| = \sqrt{A}\|w_1\| = \varepsilon$. Such a $w_1$ always exists. Then clearly $w_1$ and $w_2 = 2x - w_1$ satisfy the required constraints, and it is easy to check that $\min(\|w_1\|, \|w_2\|) = \|w_1\| = \varepsilon/\sqrt{A}$.

Finally we prove (D). By the result at part (A), $q_\infty \leq \frac{1}{\Delta}$. It is therefore sufficient to show that $Q_\varepsilon(x) \geq \frac{1}{\Delta}$ for some $x$ and $\varepsilon$. Let $S_0$ be the subset that achieves the minimum in (4.4). Let $u, v \in H$ be unit eigenvectors corresponding to the lowest eigenvalues of $F_{S_0}F_{S_0}^*$ and $F_{S_0^c}F_{S_0^c}^*$ respectively. Thus

$$\left\|F_{S_0}^* u\right\|^2 = A[S_0], \qquad \left\|F_{S_0^c}^* v\right\|^2 = A\left[S_0^c\right]$$

Let $x = (u+v)/2$ and $\varepsilon = \Delta$, and set $w_1 = u$, $w_2 = v$. Then by Proposition 4.1

$$Q_\varepsilon(x) \geq \frac{\min(\|w_1\|, \|w_2\|)}{\varepsilon} = \frac{1}{\Delta}$$

since

$$\sum_{j=1}^m \min\left(\left|\langle f_j, w_1\rangle\right|^2, \left|\langle f_j, w_2\rangle\right|^2\right) \leq \left\|F_{S_0}^* w_1\right\|^2 + \left\|F_{S_0^c}^* w_2\right\|^2 = \varepsilon^2$$

This concludes the proof. □

**Remark.** It may seem strange that $Q_\varepsilon(x) = \frac{1}{\sqrt{A}}$ for all $x \neq 0$ and sufficiently small $\varepsilon$ while $q_0 = \frac{1}{\omega}$, where $\omega$ is typically much smaller than $\sqrt{A}$. The reason is that for $Q_\varepsilon(x) = \frac{1}{\sqrt{A}}$ to hold, $\varepsilon$ depends on $x$. Thus we cannot exchange the order of $\limsup_{\varepsilon\to 0}$ and $\max_{\|x\|=1}$.

Related to the study of stability of phaseless reconstruction is the study of the Lipschitz property of the map $\alpha_{\mathcal{F}}$ on $\hat{H} := \mathbb{R}^n/\sim$. We analyze the bi-Lipschitz bounds of both $\alpha_{\mathcal{F}}$ and ${\alpha_{\mathcal{F}}}^2$, which is simply the map $\alpha_{\mathcal{F}}$ with all entries squared, i.e.

$${\alpha_{\mathcal{F}}}^2(x) := \left[\left|\langle f_j, x\rangle\right|^2, \ldots, \left|\langle f_m, x\rangle\right|^2\right]^T. \tag{4.13}$$

We shall consider two distance functions on $\hat{H} = \mathbb{R}^n/\sim$: the standard distance $d(x,y) := \min(\|x-y\|, \|x+y\|)$ and the distance $d_1(x,y) := \|xx^* - yy^*\|_1$ where $\|X\|_1$ denotes the *nuclear norm* of $X$, which is the sum of all singular values of $X$. Specifically we are interested in examining the local and global behavior of the following ratios

$$U(x,y) := \frac{\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\|}{d(x,y)}, \qquad V(x,y) := \frac{\|{\alpha_{\mathcal{F}}}^2(x) - {\alpha_{\mathcal{F}}}^2(y)\|}{d_1(x,y)}. \tag{4.14}$$

While all norms in finite dimensional spaces are equivalent, we choose to consider $d_1$, the nuclear norm induced distance on $\hat{H}$, because the Lipschitz lower and upper bounds are very much related to the matrix $R(x)$ introduced in Theorem 2.1.

We first investigate the bounds for $U(x,y)$. For this the upper bound is relatively straightforward. Let $w_1 = x - y$ and $w_2 = x + y$. We have already shown in the proof of Theorem 4.2 using (4.9) that

$$\begin{aligned}
\left\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\right\|^2 &= \sum_{j=1}^m \min\left(\left|\langle f_j, w_1\rangle\right|^2, \left|\langle f_j, w_2\rangle\right|^2\right) \\
&\leq \min\left\{\sum_{j=1}^m \left|\langle f_j, w_1\rangle\right|^2, \sum_{j=1}^m \left|\langle f_j, w_2\rangle\right|^2\right\} \\
&\leq B\min\left\{\|w_1\|^2, \|w_2\|^2\right\} = Bd^2(x,y),
\end{aligned}$$

where $B$ is the upper frame bound of the frame $\mathcal{F}$. Thus $U(x,y)$ has an upper bound $U(x,y) \leq \sqrt{B}$. Furthermore, the bound is sharp. To see this, pick a unit vector $x \in \mathbb{R}^n$ such that $\sum_{j=1}^m |\langle f_j, w_1\rangle|^2 = B$ and set $y = 2x$. Then $U(x,y) = \sqrt{B}$.

To study the lower bound $U(x,y)$ we now consider the following quantities:

$$\begin{aligned}
\rho_\varepsilon(x) &:= \inf_{\{y: d(x,y)\leq \varepsilon\}} U(x,y),\\
\rho(x) &:= \liminf_{\{y: d(x,y)\to 0\}} U(x,y) = \liminf_{\varepsilon\to 0} \rho_\varepsilon(x),\\
\rho_0 &:= \inf_x \rho(x),\\
\rho_\infty &:= \inf_{d(x,y)>0} U(x,y).
\end{aligned}$$

We apply the equality

$$U^2(x,y) = \frac{\sum_{j=1}^m \min(|\langle f_j, w_1\rangle|^2, |\langle f_j, w_2\rangle|^2)}{\min(\|w_1\|^2, \|w_2\|^2)}$$

where again $w_1 = x - y$ and $w_2 = x + y$. Now fix $x$ and let $d(x,y) < \varepsilon$. Without loss of generality we may assume $\|y - x\| < \varepsilon$. Thus $\|w_1\| < \varepsilon$ and $\|w_2 - 2x\| = \|w_1\| < \varepsilon$. Let $S = \{j,\ \langle f_j, x\rangle \neq 0\}$ and set

$$\varepsilon_0(x) := \frac{\min_{k\in S} |\langle f_k, x\rangle|}{\max_{k\in S} \|f_k\|}. \tag{4.15}$$

Note for any $w_1$ with $\|w_1\| < \varepsilon_0$ and $k \in S$ we have

$$\big|\langle f_k, w_2\rangle\big| = \big|2\langle f_k, x\rangle - \langle f_k, w_1\rangle\big| \geq 2\big|\langle f_k, x\rangle\big| - \big|\langle f_k, w_1\rangle\big| \geq 2\varepsilon_0(x)\|f_k\| - \|w_1\|\|f_k\| \geq \big|\langle f_k, w_1\rangle\big|,$$

whereas for $k \in S^c$ we have

$$\big|\langle f_k, w_2\rangle\big| = \big|\langle f_k, w_1\rangle\big|.$$

Hence $\min(|\langle f_j, w_1\rangle|^2, |\langle f_j, w_2\rangle|^2) = |\langle f_j, w_1\rangle|^2$ for all $j$ whenever $\varepsilon < \varepsilon_0(x)$. It follows that

$$U^2(x,y) = \frac{\sum_{j=1}^m |\langle f_j, w_1\rangle|^2}{\|w_1\|^2} = \sum_{j=1}^m \left|\left\langle \frac{w_1}{\|w_1\|}, f_j\right\rangle\right|^2.$$

Thus $U^2(x,y) \geq A$ where $A$ is the lower frame bound for the frame $\mathcal{F}$. Furthermore this lower bound is achieved whenever $w_1 = x - y$ is an eigenvector corresponding to the smallest eigenvalue of $FF^*$. This implies that

$$\rho_\varepsilon(x) = \sqrt{A}$$

whenever $\varepsilon < \varepsilon_0(x)$. Consequently $\rho(x) = \sqrt{A}$. We have the following theorem:

**Theorem 4.3.** *Assume that the frame $\mathcal{F}$ is phase retrievable. Let $A$, $B$ be the lower and upper frame bounds for the frame $\mathcal{F}$, respectively and for each $x \in \mathbb{R}^n$, let $\varepsilon_0(x)$ be given in (4.15). Then*

(1) *$U(x,y) \leq \sqrt{B}$ for any $x, y \in \mathbb{R}^n$ with $d(x,y) > 0$.*
(2) *Assume that $\varepsilon < \varepsilon_0(x)$. Then $\rho_\varepsilon(x) = \sqrt{A}$. Consequently $\rho(x) = \rho_0 = \sqrt{A}$.*

(3) $\Delta = \rho_\infty \leq \omega \leq \rho_0 = \rho(x) = \sqrt{A}$.
(4) *The map $\alpha_{\mathcal{F}}$ is bi-Lipschitz with (optimal) upper Lipschitz bound $\sqrt{B}$ and lower Lipschitz bound $\rho_\infty$:*

$$\rho_\infty d(x,y) \leq \big\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\big\| \leq \sqrt{B} d(x,y), \quad \forall x, y \in \hat{H}$$

**Proof.** We have already proved (1) and (2) of the theorem in the discussion. It remains only to prove (3) since (4) is just a restatement of (1) and (3). Note that

$$\rho_\infty^2 = \inf_{d(x,y)>0} U^2(x,y) = \inf_{w_1, w_2 \neq 0} \frac{\sum_{j=1}^m \min(|\langle f_j, w_1\rangle|^2, |\langle f_j, w_2\rangle|^2)}{\min(\|w_1\|^2, \|w_2\|^2)}.$$

For any $w_1, w_2$, assume without loss of generality that $0 < \|w_1\| \leq \|w_2\|$. Let $S = \{j : |\langle f_j, w_1\rangle| \leq |\langle f_j, w_2\rangle|\}$. Set $v_1 = w_1/\|w_1\|$, $v_2 = w_2/\|w_2\|$ and $t = \|w_2\|/\|w_1\| \geq 1$. Then

$$\begin{aligned}
\frac{\sum_{j=1}^m \min(|\langle f_j, w_1\rangle|^2, |\langle f_j, w_2\rangle|^2)}{\min(\|w_1\|^2, \|w_2\|^2)} &= \sum_{j\in S} \big|\langle f_j, v_1\rangle\big|^2 + t^2 \sum_{j\in S^c} \big|\langle f_j, v_2\rangle\big|^2 \\
&\geq \sum_{j\in S} \big|\langle f_j, v_1\rangle\big|^2 + \sum_{j\in S^c} \big|\langle f_j, v_2\rangle\big|^2 \\
&\geq \Delta^2.
\end{aligned}$$

Hence $\rho_\infty \geq \Delta$.

Let $S$ and $u, v \in H$ be normalized (eigen) vectors that achieve the bound $\Delta$, that is:

$$\|u\| = \|v\| = 1, \qquad \sum_{k\in S} \big|\langle u, f_k\rangle\big|^2 + \sum_{k\in S^c} \big|\langle v, f_k\rangle\big|^2 = \Delta^2.$$

Set $x = u + v$ and $y = u - v$. Then, following [9]

$$\begin{aligned}
\big\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\big\|^2 &= \sum_{k\in S} \big|\,|\langle u+v, f_k\rangle| - |\langle u-v, f_k\rangle|\,\big|^2 + \sum_{k\in S^c} \big|\,|\langle u+v, f_k\rangle| - |\langle u-v, f_k\rangle|\,\big|^2 \\
&\leq 4\bigg(\sum_{k\in S} \big|\langle u, f_k\rangle\big|^2 + \sum_{k\in S^c} \big|\langle v, f_k\rangle\big|^2\bigg) = 4\Delta^2.
\end{aligned}$$

On the other hand

$$d(x,y) = \min\big(\|x-y\|, \|x+y\|\big) = 2.$$

Thus we obtain

$$\frac{\|\alpha_{\mathcal{F}}(x) - \alpha_{\mathcal{F}}(y)\|}{d(x,y)} \leq \Delta.$$

The theorem is now proved. □

**Remark.** The two quantities, $\rho_\infty$ and $q_\infty$ satisfy $\rho_\infty = \frac{1}{q_\infty}$. However there are subtle differences between $Q_\varepsilon(x)$ and $\rho_\varepsilon(x)$ so that the simple relationship $\rho_\varepsilon(x) = 1/Q_\varepsilon(x)$ does not usually hold. One such difference is due to the significance of $\varepsilon$ for the two bounds. See the numerical example presented in the last section.

**Remark.** The upper Lipschitz bound $\sqrt{B}$ has been obtained independently in [9]. The lower Lipschitz bound we obtained here strenghtens the estimates given in [9]. Specifically their estimate for $\rho_\infty$ reads $\boldsymbol{\sigma} \leq \rho_\infty \leq \sqrt{2}\boldsymbol{\sigma}$ where

$$\boldsymbol{\sigma} = \min_S \max\big(\sigma_n(F_S), \sigma_n(F_{S^c})\big) \tag{4.16}$$

Clearly $\boldsymbol{\sigma} \leq \Delta \leq \sqrt{2}\boldsymbol{\sigma}$.

We conclude this section by turning our attention to the analysis of $V(x,y)$. A motivation for studying it is that in practical problems the noise is often added directly to $\alpha_{\mathcal{F}}{}^2$ as in (3.1) rather than to $\alpha_{\mathcal{F}}$. Such noise model is used in many studies of phaseless reconstruction, e.g. in the Phaselift algorithm [10], or in the IRLS algorithm in [4].

Let $\mathrm{Sym}_n(\mathbb{R})$ denote the set of $n \times n$ symmetric matrices over $\mathbb{R}$. It is a Hilbert space with the standard inner product given by $\langle X, Y\rangle := \mathrm{tr}(XY^T) = \mathrm{tr}(XY)$. The nonlinear map $\alpha_{\mathcal{F}}{}^2$ actually induces a linear map on $\mathrm{Sym}_n(\mathbb{R})$. Write $X = xx^T$ for any $x \in \mathbb{R}^n$. Then the entries of $\alpha_{\mathcal{F}}{}^2(x)$ are

$$\big(\alpha_{\mathcal{F}}{}^2(x)\big)_j = \big|\langle f_j, x\rangle\big|^2 = x^T f_j f_j^T x = \mathrm{tr}(F_j X) = \langle F_j, X\rangle, \tag{4.17}$$

where $F_j := f_j f_j^T$. Now we denote by $\mathcal{A}$ the linear operator $\mathcal{A}:\ \mathrm{Sym}_n(\mathbb{R}) \longrightarrow \mathbb{R}^m$ with entries

$$\big(\mathcal{A}(X)\big)_j = \langle F_j, X\rangle = \mathrm{tr}(F_j X).$$

Let $S_n^{p,q}$ be the set of $n \times n$ real symmetric matrices that have at most $p$ positive and $q$ negative eigenvalues. Thus $S_n^{1,0}$ denotes the set of $n \times n$ real symmetric non-negative definite matrices of rank at most one. Note that spectral decomposition easily shows that $X \in S_n^{1,0}$ if and only if $X = xx^T$ for some $x \in \mathbb{R}^n$.

The following lemma will be useful in this analysis

**Lemma 4.4.** *The following are equivalent.*

(A) $X \in S_n^{1,1}$.
(B) $X = xx^T - yy^T$ *for some* $x, y \in \mathbb{R}^n$.
(C) $X = \frac{1}{2}(w_1 w_2^T + w_2 w_1^T)$ *for some* $w_1, w_2 \in \mathbb{R}^n$.

*Furthermore, for* $X = \frac{1}{2}(w_1 w_2^T + w_2 w_1^T)$ *its nuclear norm is* $\|X\|_1 = \|w_1\|\|w_2\|$.

**Proof.** (A) $\Rightarrow$ (B) is a direct result of spectral decomposition, which yields $X = \beta_1 u_1 u_1^T - \beta_2 u_2 u_2^T$ for some $u_1, u_2 \in \mathbb{R}^n$ and $\beta_1, \beta_2 \geq 0$. Thus $X = xx^T - yy^T$ where $x := \sqrt{\beta_1}u_1$ and $y := \sqrt{\beta_2}u_2$.

(B) $\Rightarrow$ (C) is proved directly by setting $w_1 = x - y$ and $w_2 = x + y$.

We now prove (C) $\Rightarrow$ (A) by computing the eigenvalues of $X = \frac{1}{2}(w_1 w_2^T + w_2 w_1^T)$. Obviously $\mathrm{rank}(X) \leq 2$. Let $\lambda_1, \lambda_2$ be the two (possibly) nonzero eigenvalues of $X$. Then

$$\begin{aligned}
\lambda_1 + \lambda_2 &= \mathrm{tr}\{X\} = \langle w_1, w_2\rangle, \\
\lambda_1^2 + \lambda_2^2 &= \mathrm{tr}\{X^2\} = \big(\|w_1\|^2\|w_2\|^2 + \big|\langle w_1, w_2\rangle\big|^2\big)/2.
\end{aligned}$$

Solving for eigenvalues we obtain

$$\begin{aligned}
\lambda_1 &= \frac{1}{2}\big(\langle w_1, w_2\rangle + \|w_1\|\|w_2\|\big), \\
\lambda_2 &= \frac{1}{2}\big(\langle w_1, w_2\rangle - \|w_1\|\|w_2\|\big).
\end{aligned}$$

Hence, by Cauchy–Schwarz inequality, $\lambda_1 \geq 0 \geq \lambda_2$ which proves $X \in S_n^{1,1}$. Furthermore, it also shows that the nuclear norm of $X$ is $\|X\|_1 = |\lambda_1| + |\lambda_2| = \|w_1\|\|w_2\|$. $\square$

Now we analyze $V(x,y)$. Parallel to the study of $U(x,y)$ we consider the following quantities:

$$\begin{aligned}
\mu_\varepsilon(x) &:= \inf_{\{y: d(x,y)\leq \varepsilon\}} V(x,y),\\
\mu(x) &:= \liminf_{\{y: d(x,y)\to 0\}} V(x,y) = \liminf_{\varepsilon\to 0} \mu_\varepsilon(x),\\
\mu_0 &:= \inf_x \mu(x),\\
\mu_\infty &:= \inf_{d(x,y)>0} V(x,y),
\end{aligned}$$

as well as the upper bound $\sup_{d_1(x,y)>0} V(x,y)$. By (4.17) we have $|\langle f_j, x\rangle|^2 - |\langle f_j, y\rangle|^2 = \langle F_j, X\rangle$ where $F_j = f_j f_j^T$ and $X = xx^T - yy^T$. Hence

$$V^2(x,y) = \frac{\sum_{j=1}^m |\langle F_j, X\rangle|^2}{\|X\|_1^2}.$$

Set $w_1 = x - y$ and $w_2 = x + y$ and apply Lemma 4.4 we obtain

$$V^2(x,y) = \frac{\sum_{j=1}^m |\langle f_j, w_1\rangle|^2 |\langle f_j, w_2\rangle|^2}{\|w_1\|^2\|w_2\|^2}. \tag{4.18}$$

We can immediately obtain the upper bound:

$$V(x,y) \leq \left(\sup_{\|e_1\|=1,\|e_2\|=1} \sum_{j=1}^m \big|\langle f_j, e_1\rangle\big|^2 \big|\langle f_j, e_2\rangle\big|^2\right)^{1/2} = \left(\max_{\|e\|=1} \sum_{j=1}^m \big|\langle f_j, e\rangle\big|^4\right)^{1/2} =: \Lambda_{\mathcal{F}}{}^2$$

where $\Lambda_{\mathcal{F}}$ denotes the operator norm of the linear analysis operator $T : H \to \mathbb{R}^m$, $T(x) = (\langle x, f_k\rangle)_{k=1}^m$ defined between the Euclidian space $H = \mathbb{R}^n$ and the Banach space $\mathbb{R}^m$ endowed with the $l^4$-norm:

$$\Lambda_{\mathcal{F}} = \left(\max_{\|x\|=1} \sum_{k=1}^m \big|\langle x, f_k\rangle\big|^4\right)^{1/4} \tag{4.19}$$

Note also that

$$\Lambda_{\mathcal{F}}{}^2 = \max_{\|x\|=1} \lambda_{\max}\big(R(x)\big)$$

where $R(x)$ was defined in (2.6). An immediate bound is $\Lambda_{\mathcal{F}} \leq \sqrt{B}\max\|f_k\|$ with $B$ the upper frame bound of $\mathcal{F}$.

Fix $x \neq 0$ and let $d(x,y) \to 0$. Then either $y \to x$ or $y \to -x$. Without loss of generality we assume that $x \to y$. Thus $w_1 = x - y \to 0$ and $w_2 = x + y \to 2x$. However $w_1/\|w_1\|$ can be any unit vector. Thus

$$\mu^2(x) = \frac{1}{\|x\|^2} \inf_{\|u\|=1} \sum_{j=1}^m \big|\langle f_j, x\rangle\big|^2 \big|\langle f_j, u\rangle\big|^2 = \frac{1}{\|x\|^2} \inf_{\|u\|=1} \big\langle R(x)u, u\big\rangle = \frac{1}{\|x\|^2}\lambda_{\min}\big(R(x)\big)$$

where $R(x)$ was introduced in (2.6). Thus we obtain

$$\mu^2(x) = \frac{1}{\|x\|^2}\lambda_{\min}\big(R(x)\big), \qquad \mu_0^2 = \min_{\|u\|=1} \lambda_{\min}\big(R(u)\big).$$

On the other hand note

$$\inf_{d(x,y)>0} V^2(x,y) = \inf_{w_1,w_2\neq 0} \frac{\sum_{j=1}^m |\langle f_j, w_1\rangle|^2 |\langle f_j, w_2\rangle|^2}{\|w_1\|^2\|w_2\|^2} = \min_{\|u\|=1} \lambda_{\min}\big(R(u)\big) = a_0^2,$$

where $a_0$ was introduced in (2.4). Thus we proved:

**Theorem 4.5.** *Assume the frame $\mathcal{F}$ is phase retrievable. Then*

$$\mu(x) = \frac{1}{\|x\|}\sqrt{\lambda_{\min}\big(R(x)\big)}, \tag{4.20}$$

$$\mu_\infty = \mu_0 = \min_{u:\|u\|=1} \sqrt{\lambda_{\min}\big(R(u)\big)} = \sqrt{a_0}. \tag{4.21}$$

*Furthermore ${\alpha_{\mathcal{F}}}^2$ is bi-Lipschitz with upper Lipschitz bound ${\Lambda_{\mathcal{F}}}^2$ and lower Lipschitz bound $\sqrt{a_0}$:*

$$\sqrt{a_0} d_1(x,y) \leq \big\|{\alpha_{\mathcal{F}}}^2(x) - {\alpha_{\mathcal{F}}}^2(y)\big\| \leq {\Lambda_{\mathcal{F}}}^2 d_1(x,y)$$

*where $a_0$ is the same positive constant used in Theorems 2.1 and 3.1, and $\Lambda_{\mathcal{F}}$ is the norm of the analysis operator defined between the Euclidian space $H$ and $l^4(\{1,2,\dots,m\})$.*

**Remark.** Note that the distance $d(.,.)$ is not equivalent to $d_1(.,.)$. Theorem 4.5 now also implies that ${\alpha_{\mathcal{F}}}^2$ is not bi-Lipschitz with respect to the distance $d(.,.)$ on $\hat{H}$. This fact was pointed out in [9].

## 5. Robustness and size of redundancy

Previous sections establish results on the robustness of phaseless reconstruction for the worst case scenario. A natural question is to ask: can "reasonable" robustness be achieved for a given frame, and in particular with small number of samples? We shall examine how $q_\infty$ scales as the dimension $n$ increases.

Consider the case where $m = 2n-1$. This is the minimal redundancy required for phaseless reconstruction. In this case any frame $\mathcal{F}$ would have $\Delta = \omega$. Hence we have $\min\{1/\omega, 1/\varepsilon\} \leq q_\varepsilon = 1/\omega$. The stability of the reconstruction is thus mostly controlled by the size of $1/\omega$. The question is: how big is $\omega$, especially as $n$ increases?

Assume that the frame elements of $\mathcal{F}$ are all bounded by $L$, $\|f_j\| \leq L$ for all $f_j \in \mathcal{F}$. Consider the $n+1$ elements $\{f_j : j = 1,\dots,n+1\}$. They are linearly dependent so we can find $c_j \in \mathbb{R}$ such that $\sum_{j=1}^{n+1} c_j f_j = 0$. Without loss of generality we may assume $|c_{n+1}| = \min\{|c_j|\}$. Set $v = [c_1, c_2, \dots, c_n]^T$. Let $G = [f_1, \dots, f_n]$. Then $Gv = \sum_{j=1}^n c_j f_j = -c_{n+1} f_{n+1}$. Now all $|c_j| \geq |c_{n+1}|$ so $\|v\| \geq \sqrt{n}|c_{n+1}|$. Thus

$$\|Gv\| = |c_{n+1}|\|f_{n+1}\| \leq \frac{L}{\sqrt{n}}\|v\|.$$

It follows that $\sigma_n(G) \leq \frac{L}{\sqrt{n}}$, and hence

$$\omega \leq \frac{L}{\sqrt{n}}. \tag{5.1}$$

Note that here we have considered only the first $n+1$ vectors of the frame $\mathcal{F}$. The actual value of $\omega$ will likely decay much faster as $n$ increases. In a preliminary work we are able to establish the bound $\omega \leq CL/\sqrt{n^3}$ where $C$ is independent of $n$ [18]. But even this estimate is likely far from optimal.

**Conjecture 5.1.** *Let $m = 2n-1$ and $\|f_j\| \leq L$ for all $f_j \in \mathcal{F}$. Then there exist constants $C > 0$ and $0 < \beta < 1$ independent of $n$ such that*

$$\omega \leq CL\beta^n.$$

A related problem is as follows: Consider an $n \times (n+k)$ matrix $F = [g_1, g_2, \ldots, g_{n+k}]$. Let $\tau = \min\{\sigma_n(F_S) : \ S \subset \{1, \ldots, n+k\}, |S| = n\}$. Assume that all $\|g_j\| \leq 1$. How large can $\tau$ be? For $k = 1$ we have already seen that it is bounded from above by $C/\sqrt{n}$. The preliminary work [18] shows that for $k = 1$ it is bounded from above by $C/n^{\frac{3}{2}}$.

**Conjecture 5.2.** *There exists a constant $C = C(k, n)$ such that*

$$\tau \leq \frac{C}{n^{k-\frac{1}{2}}},$$

*where $C(k, n) = O_k(\log^{q_k} n)$ for some $q_k > 0$. Here $O_k$ denotes the dependence on $k$.*

Thus in the minimal setting with $m = 2n-1$ it is impossible to achieve scale independent stability for phaseless reconstruction. The same arguments can be used to show that even when $m = 2n + k_0$ for some fixed $k_0$ scale independent stability is not possible. A natural question is whether scale independent stability is possible when we increase the redundancy of the frame. As it turns out this is possible via a recent work by Wang [17]. More precisely, the following result follows from the main results in [17]:

**Theorem 5.3.** *Let $r_0 > 2$ and let $F = \frac{1}{\sqrt{n}} G$ where $G$ is an $n \times m$ random matrix whose elements are i.i.d. normal $N(0, 1)$ random variables such that $m/n = r_0$. Then there exist constants $0 < \Delta_0 \leq \omega_0$ dependent only on $r_0$ and not on $n$ such that with high probability we have*

$$\Delta \geq \Delta_0, \qquad \omega \geq \omega_0.$$

**Proof.** Theorem 1.1 and Theorem 3.1 of [17] proves the following result: Let $\lambda > \Delta > 1$ be fixed. Assume that $A = \frac{1}{\sqrt{n}} B$ where $B$ is an $n \times N$ random Gaussian matrix with i.i.d. $N(0, 1)$ entries such that $N/n = \lambda$. Then there exists a constant $c > 0$ depending only on $\tau_0$, $\lambda$ and $\Delta$ such that

$$\min_{S \subseteq \{1, \ldots, N\}, |S| \geq \Delta n} \sigma_n(A_S) \geq c$$

with probability at least $1 - 3e^{-\tau_0 n}$. The value $c$ was explicitly estimated in terms of $\tau_0$, $\lambda$ and $\Delta$ in the proof of Theorem 3.1 in [17].

The theorem now readily follows. Observe that because $r_0 > 2$, in the expression for $\Delta$ we may choose $\lambda = r_0$ $\Delta = \frac{r_0}{2} > 1$ and clearly we have

$$\Delta \geq \min_{S \subseteq \{1, \ldots, N\}, |S| \geq \Delta n} \sigma_n(F_S) \geq \Delta_0,$$

for some $\Delta_0 > 0$ independent of $n$. For $\omega$ we may choose $\lambda = r_0$ and $\Delta = r_0 - 1 > 1$. Again the theorem of [17] implies that

$$\omega \geq \min_{S \subseteq \{1, \ldots, N\}, |S| \geq \Delta n} \sigma_n(F_S) \geq \omega_0. \qquad \square$$

In the theorem the values $\Delta_0$ and $\omega_0$ can be estimated explicitly using the estimates in [17]. Here with high probability is in the standard sense that the probability is at least $1 - c_0 e^{-\beta n}$ for some $c_0, \beta > 0$. Thus

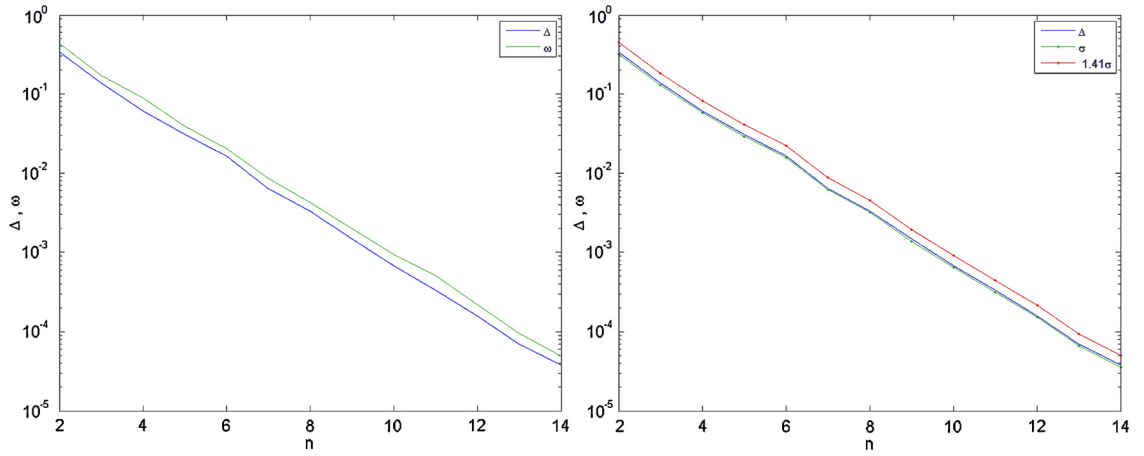


**Fig. 1.** Plots of sample medians of $\Delta$ and $\omega$ (left plot) and $\Delta$ and $\sigma, \sqrt{2}\sigma$ (right plot) for randomly generated frames of size $m = 2n$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

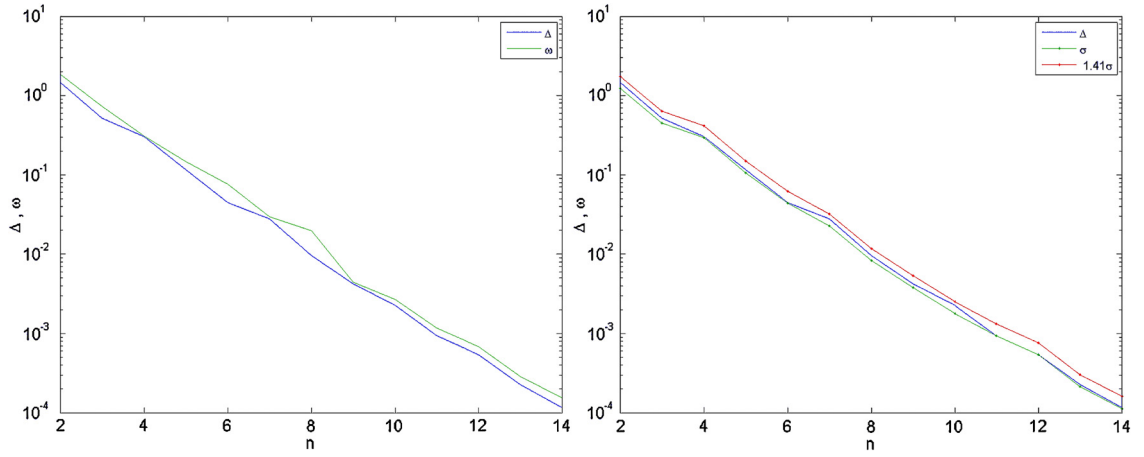


**Fig. 2.** Plots of largest sample value of $\Delta$ and $\omega$ (left plot) and $\Delta$ and $\sigma, \sqrt{2}\sigma$ (right plot) for randomly generated frames of size $m = 2n$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

scale independent stable phaseless reconstruction is possible whenever the redundancy is greater than $2+\Delta$, $\Delta > 0$, at least for random Gaussian matrices.

## 6. Numerical examples

In this section we present two numerical studies of the stability bounds derived earlier.

**1**. First consider the following setup. For each $n$ between 2 and 14 we generate 100 realizations of random frames of $m = 2n$ vectors where each entry is i.i.d. normally distributed with zero mean and unit variance. For each realization we compute $\Delta$, $\omega$ and $\sigma$. For each fixed $n$ we compute the sample median, the largest sample value and the smallest sample value of these random variables.

Fig. 1 contains the plots of sample medians of $\Delta$, $\omega$ and $\sigma$'s for each value of $n$. The left plot contains only $\Delta$ (the lower Lipschitz constant) and $\omega$ (the lower Lipschitz constant for small perturbations); the right plot contains $\Delta$ and the two bounds $\sigma$ and $\sqrt{2}\sigma$ as obtained in [9]. Similar statistics are plotted in Fig. 2 where sample medians are replaced by the largest sample values, and in Fig. 3 where sample medians are replaced by smallest sample values.

Note there is about 1–2 orders of magnitude spread between the largest and the smallest sample value of these random variables.

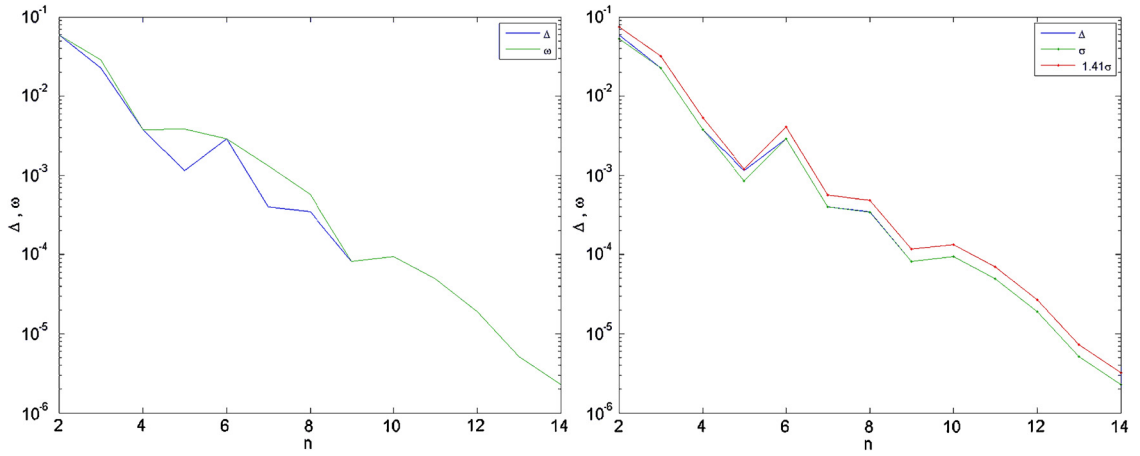


**Fig. 3.** Plots of largest sample value of $\Delta$ and $\omega$ (left plot) and $\Delta$ and $\sigma, \sqrt{2}\sigma$ (right plot) for randomly generated frames of size $m = 2n$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

**2**. Next we consider the following specific example. $H = \mathbb{R}^2$, $m = 4$ and the frame containing

$$f_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad f_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad f_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad f_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

which is a tight frame of bounds $A = B = 3$. The frame is full spark hence phase retrievable. The bounds $\Delta$ and $\omega$ defined by (4.4) and (4.5) are given by

$$\Delta = \sqrt{3 - \sqrt{5}} = 0.874032, \quad \omega = 1$$

which corresponds to choices $S = \{1, 3\}$ and $S = \{1, 2, 3\}$, respectively. The parameters $\boldsymbol{\sigma}$ introduced in (4.16) is given by

$$\boldsymbol{\sigma} = \sqrt{\frac{3 - \sqrt{5}}{2}} = 0.618034$$

and corresponds to $S = \{1, 3\}$. The parameter $\tau$ introduced in the statement of Theorem 4.2 is given by the same expression, $\tau = \boldsymbol{\sigma} = \sqrt{\frac{3-\sqrt{5}}{2}} = 0.618034$ and corresponds to the same selection $S = \{1, 3\}$.

Tedious algebra can provide closed form expressions for $\rho_\varepsilon(x)$ as function of radius $\varepsilon$. Because of scaling relation $\rho_{c\varepsilon}(cx) = \rho_\varepsilon(x)$ for all $c > 0$ it follows that only the direction of $x$ describes this function. For instance for $x^{(1)} = (1, 0)$ we obtain the following expression:

$$\rho_\varepsilon\big(x^{(1)}\big) = \begin{cases} \sqrt{3}, & \varepsilon < \frac{1}{\sqrt{2}} \\ \sqrt{3 - \frac{4\sqrt{2}}{\varepsilon} + \frac{4}{\varepsilon^2}}, & \frac{1}{\sqrt{2}} \le \varepsilon < \sqrt{2} \\ 1, & \sqrt{2} \le \varepsilon \end{cases}$$

For $x^{(2)} = (1, 1)$ we obtain:

$$\rho_\varepsilon\big(x^{(2)}\big) = \begin{cases} \sqrt{3}, & \varepsilon < 1 \\ \sqrt{3 - \frac{4}{\varepsilon} + \frac{4}{\varepsilon^2}}, & 1 \le \varepsilon < 2 \\ \sqrt{2}, & 2 \le \varepsilon \end{cases}$$

The plots of these two functions are depicted in Fig. 4.

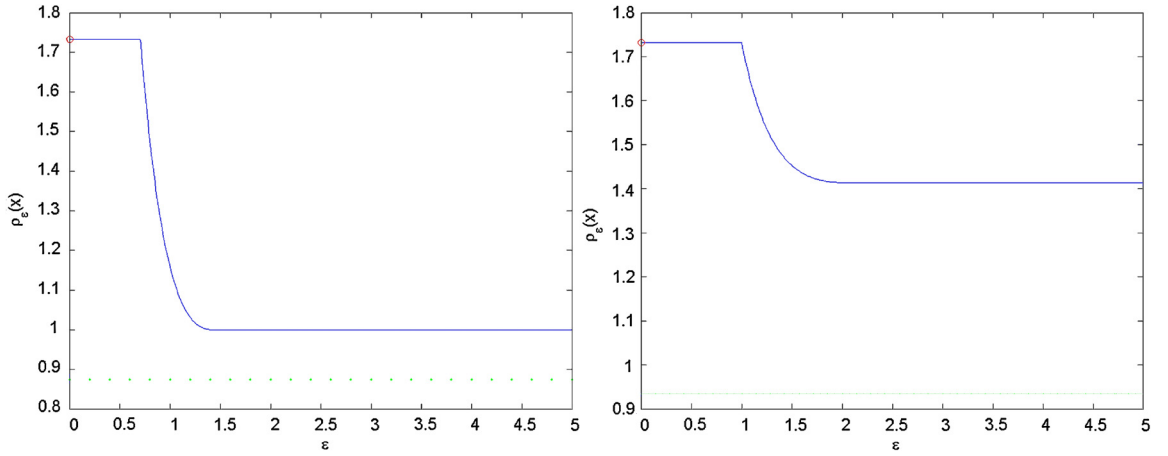


**Fig. 4.** Plots of $\rho_\varepsilon(x^{(1)})$ (left plot) and $\rho_\varepsilon(x^{(2)})$ (right plot) as function of radius $\varepsilon$. The red circle is at $\sqrt{A} = \sqrt{3}$. The horizontal dotted line is the lower bound $\Delta = 0.874$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

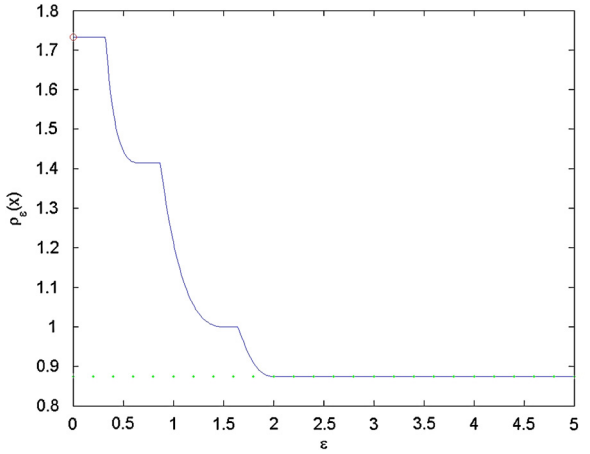


**Fig. 5.** Plots of $\rho_\varepsilon(x^{(3)})$ as function of radius $\varepsilon$. The red circle is at $\sqrt{A} = \sqrt{3}$. The horizontal dotted line is the lower bound $\Delta = 0.874$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Following the proof of Theorem 4.3 it follows the critical direction that achieves the lower bound $\sqrt{\Delta}$ is given by $x = u + v$ where $u$ and $v$ are the two normalized eigenvectors associated to the lowest eigenvalue (i.e. the lower frame bound) for $\{f_1, f_3\}$ and $\{f_2, f_4\}$ respectively. The lowest eigenvalue is given by $\frac{3-\sqrt{5}}{2}$ and the eigenvectors are

$$u = \begin{bmatrix} -\sqrt{\frac{2}{5+\sqrt{5}}} \\ \frac{1+\sqrt{5}}{\sqrt{2(5+\sqrt{5})}} \end{bmatrix}, \qquad v = \begin{bmatrix} -\sqrt{\frac{2}{5-\sqrt{5}}} \\ \frac{1-\sqrt{5}}{\sqrt{2(5-\sqrt{5})}} \end{bmatrix}$$

and thus the critical vector is

$$x^{(3)} = u + v = \begin{bmatrix} -\sqrt{\frac{2}{5+\sqrt{5}}} - \sqrt{\frac{2}{5-\sqrt{5}}} \\ \frac{1+\sqrt{5}}{\sqrt{2(5+\sqrt{5})}} \frac{1-\sqrt{5}}{\sqrt{2(5-\sqrt{5})}} \end{bmatrix} = \begin{bmatrix} -1.3764 \\ 0.3249 \end{bmatrix}$$

The function $\rho_\varepsilon(x^{(3)})$ is computed numerically and is plotted in Fig. 5. For reference we pictured a circle at $\sqrt{A} = \sqrt{3}$ and we plotted a dotted line at $\Delta = 0.874$. We remark in all three cases, the limit

$\lim_{\varepsilon\to 0}\rho_\varepsilon(x) = \sqrt{A} = \rho_0$ as predicted by Theorem 4.3. Furthermore, $\min_{\varepsilon>0,x}\rho_\varepsilon(x) = \Delta = \rho_\infty$ as proved in same Theorem 4.3.

## Acknowledgments

The authors would like to thank Matt Fickus, Dustin Mixon and Jeffrey Schenker for very helpful discussions.

R. Balan was supported in part by NSF DMS-1109498. Y. Wang was supported in part by NSF DMS-1043032, and by AFOSR FA9550-12-1-0455.

## References

[1] B. Alexeev, A.S. Bandeira, M. Fickus, D.G. Mixon, Phase retrieval with polarization, arXiv:1210.7752v1 [cs.IT], 2012.
[2] R. Balan, A nonlinear reconstruction algorithm from absolute value of frame coefficients for low redundancy frames, in: Proceedings of SampTA Conference, Marseille, France, May 2009.
[3] R. Balan, On signal reconstruction from its spectrogram, in: Proceedings of the CISS Conference, Princeton, NJ, May 2010.
[4] R. Balan, Reconstruction of signals from magnitudes of redundant representations, arXiv:1207.1134v1 [math.FA], 2012.
[5] R. Balan, Reconstruction of signals from magnitudes of redundant representations: the complex case, Available online arXiv:1304.1839v1 [math.FA], 2013.
[6] R. Balan, P. Casazza, D. Edidin, On signal reconstruction without phase, Appl. Comput. Harmon. Anal. 20 (2006) 345–356.
[7] R. Balan, P. Casazza, D. Edidin, Equivalence of reconstruction from the absolute value of the frame coefficients to a sparse representation problem, IEEE Signal Process. Lett. 14 (5) (2007) 341–343.
[8] R. Balan, B. Bodmann, P. Casazza, D. Edidin, Painless reconstruction from magnitudes of frame coefficients, J. Fourier Anal. Appl. 15 (4) (2009) 488–501.
[9] A.S. Bandeira, J. Cahill, D.G. Mixon, A.A. Nelson, Saving phase: injectivity and stability for phase retrieval, arXiv:1302.4618v2 [math.FA], 2013.
[10] E. Candés, T. Strohmer, V. Voroninski, PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming, Comm. Pure Appl. Math. 66 (8) (2013) 1241–1274.
[11] E. Candés, Y. Eldar, T. Strohmer, V. Voroninski, Phase retrieval via matrix completion problem, SIAM J. Imag. Sci. 6 (1) (2013) 199–225.
[12] P. Casazza, The art of frame theory, Taiwanese J. Math. (2) 4 (2000) 129–202.
[13] D.A. Depireux, M. Elhilali, Handbook of Modern Techniques in Auditory Cortex (Otolaryngology Research Advances), Nova Science Publishers, 2014.
[14] Y.C. Eldar, S. Mendelson, Phase retrieval: stability and recovery guarantees, Available online arXiv:1211.0872.
[15] S.M. Kay, Fundamentals of Statistical Signal Processing. I. Estimation Theory, Prentice Hall PTR, 2010, 18th printing.
[16] I. Waldspurger, A. d'Aspremont, S. Mallat, Phase recovery, MaxCut and complex semidefinite programming, Available online, arXiv:1206.0102.
[17] Y. Wang, Random matrices and erasure robust frames, E-print http://arxiv.org/abs/1403.5969.
[18] Y. Wang, Worst case condition number for matrices with erasure, preliminary report.