Available online at www.sciencedirect.com
SCIENCE DIRECT®
Appl. Comput. Harmon. Anal. 20 (2006) 345–356
Applied and Computational Harmonic Analysis
www.elsevier.com/locate/acha

On signal reconstruction without phase ☆

Radu Balan [a], Pete Casazza [b,*], Dan Edidin [b]

[a] *Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA*
[b] *Department of Mathematics, University of Missouri, Columbia, MO 65211, USA*

Received 15 December 2004; revised 22 June 2005; accepted 10 July 2005

Available online 19 August 2005

Communicated by Charles K. Chui

Abstract

We will construct new classes of Parseval frames for a Hilbert space which allow signal reconstruction from the absolute value of the frame coefficients. As a consequence, signal reconstruction can be done without using phase or its estimation. This verifies a longstanding conjecture of the speech processing community.

Keywords: Frame; Signal reconstruction; Phase; Speech recognition

1. Introduction

Reconstruction of a signal using noisy phase or its estimation can be a critical problem in speech recognition technology. But, for many years, engineers have believed that speech recognition should be independent of phase. By constructing new classes of Parseval frames for a Hilbert space, we will show that this allows reconstruction of a signal without using noisy phase or its estimation. This verifies the longstanding conjecture of the speech processing community.

Frames are redundant systems of vectors in a Hilbert spaces. They satisfy the well-known property of perfect reconstruction, in that any vector of the Hilbert space can be synthesized back from its inner

☆ The second author was supported by NSF DMS 0405376 and the third author was supported by NSA MDA 904-03-1-0040.
* Corresponding author.
E-mail addresses: radu.balan@siemens.com (R. Balan), pete@math.missouri.edu (P. Casazza), edidin@math.missouri.edu (D. Edidin).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2005.07.001

products with the frame vectors. More precisely, the linear transformation from the initial Hilbert space to the space of coefficients obtained by taking the inner product of a vector with the frame vectors is injective and hence admits a left inverse. This property has been successfully used in a broad spectrum of applications, including Internet coding, multiple antenna coding, optics, quantum information theory, signal/image processing, and much more. The purpose of this paper is to study what kind of reconstruction is possible if we only have knowledge of the absolute values of the frame coefficients.

In this paper we consider only finite-dimensional frames the reason being their direct link to practical applications. Since the same question can be raised for infinite-dimensional frames, we state the problem in the setting of abstract frames.

Consider a Hilbert space H with scalar product $\langle\,,\rangle$. A finite or countable set of vectors $\mathcal{F} = \{f_i;\ i \in \mathbb{I}\}$ of H is called a *frame* if there are two positive constants $A, B > 0$ such that for every vector $x \in H$,

$$A\|x\|^2 \leqslant \sum_{i\in I} \big|\langle x, f_i\rangle\big|^2 \leqslant B\|x\|^2. \tag{1.1}$$

The frame is *tight* when the constants can be chosen equal to one another, $A = B$. For $A = B = 1$, $\mathcal{F}$ is called a *Parseval frame*. The numbers $\langle x, f_i\rangle$ are called *frame coefficients*.

To a frame $\mathcal{F}$ we associate the *analysis* and *synthesis operators* defined by

$$T : H \to l^2(\mathbb{I}), \qquad T(x) = \big\{\langle x, f_i\rangle\big\}_{i\in\mathbb{I}}, \tag{1.2}$$

$$T^* : l^2(\mathbb{I}) \to H, \qquad T^*(c) = \sum_{i\in I} c_i f_i, \tag{1.3}$$

which are well defined due to (1.1), and are adjoint to one another. The range of T in $l^2(\mathbb{I})$ is called the *range of coefficients*. The *frame operator* defined by $S = T^*T : H \to H$ is invertible by (1.1) and provides the perfect reconstruction formula:

$$x = \sum_{i\in\mathbb{I}} \langle x, f_i\rangle S^{-1} f_i. \tag{1.4}$$

For more information on frames we refer the reader to [6].

Consider now the nonlinear mapping

$$\mathbb{M}_a : H \to l^2(\mathbb{I}), \qquad \mathbb{M}_a(x) = \big\{\big|\langle x, f_i\rangle\big|\big\}_{i\in\mathbb{I}} \tag{1.5}$$

obtained by taking the absolute value entrywise of the analysis operator. Let us denote by H_r the quotient space $H_r = H/\sim$ obtained by identifying two vectors that differ by a constant phase factor: $x \sim y$ if there is a scalar c with $|c| = 1$ so that $y = cx$. For real Hilbert spaces c can only be $+1$ or -1, and thus $H_r = H/\{\pm 1\}$. For complex Hilbert spaces c can be any complex number of modulus one, $c = e^{i\varphi}$, and then $H_r = H/\mathbb{T}^1$, where $\mathbb{T}^1$ is the complex unit circle. In quantum mechanics these projective rays define quantum states (see [16]). Clearly two vectors of H in the same ray would have the same image through $\mathbb{M}_a$. Thus the nonlinear mapping $\mathbb{M}_a$ extends to H_r as

$$\mathbb{M} : H_r \to l^2(\mathbb{I}), \qquad \mathbb{M}(\hat{x}) = \big\{\big|\langle x, f_i\rangle\big|\big\}_{i\in\mathbb{I}}, \quad x \in \hat{x}. \tag{1.6}$$

The problem we study in this paper is the injectivity of the map $\mathbb{M}$. When it is injective, $\mathbb{M}$ admits a left inverse, meaning that any vector (signal) in H can be reconstructed up to a constant phase factor from the modulus of its frame coefficients.

The motivation for this problem comes from two applications in signal processing, one concerning noise reduction, and the other regarding speech recognition. There is also a connection with a problem in optics that we describe later.

The traditional method of signal enhancement consists of three steps: first, the input signal is linearly transformed from its input domain (e.g., time, or space) into a transformed domain (e.g., time–frequency, time–scale, space–scale, etc.); second, a (nonlinear) estimation operator is applied in this representation domain; third, a (left) inverse of the linear transformation at step one is applied to the signal obtained at step two in order to synthesize the estimated signal in the input domain. Several linear transformations have been proposed in the literature and are used in practice, e.g., windowed Fourier transform, wavelet filterbanks, local cosine basis, etc. (see [9,18]). Likewise, many signal estimators have been proposed and studied in the literature, some of them statistically motivated, e.g., Wiener (MMSE) filter, maximum a posteriori (MAP), maximum likelihood (ML), etc., others having a rather ad hoc motivation, e.g., spectral subtraction, psychoacoustically motivated audio and video estimators, etc. For more details see [1,8,17] and many other books on this topic. By way of an example let us consider the Ephraim–Malah noise reduction method [7] of speech signals. Let $\{x(t),\ t = 1, 2, \ldots, T\}$ be the samples of a speech signal. These samples are first transformed into the time–frequency domain through the *fast Fourier transform*,

$$X(k,\omega) = \sum_{t=0}^{M-1} g(t)x(t+kN)e^{-2\pi i\omega\frac{t}{M}}, \quad k = 0, 1, \ldots, \frac{T-M}{N}, \tag{1.7}$$

$\omega \in \{0, 1, \ldots, M-1\}$, where g is the analysis window, and M, N are respectively the window size, and the time step. Next a complicated nonlinear transformation is applied to $|X(k,\omega)|$ to produce the MMSE estimate of the short-time spectral amplitude

$$\begin{aligned} Y(k,\omega) = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v(k,\omega)}}{\gamma(k,\omega)}\exp\left(-\frac{v(k,\omega)}{2}\right)\left[\left(1+v(k,\omega)\right)I_0\left(\frac{v(k,\omega)}{2}\right)\right. \\ \left. + v(k,\omega)I_1\left(\frac{v(k,\omega)}{2}\right)\right]\left|X(k,\omega)\right|, \end{aligned} \tag{1.8}$$

where I_0, I_1 are modified Bessel functions of zero and first order, and $v(k,\omega), \gamma(k,\omega)$ are estimates of certain signal-to-noise ratios. The speech signal windowed Fourier coefficients are estimated simply by

$$\hat{X}(k,\omega) = Y(k,\omega)\frac{X(k,\omega)}{|X(k,\omega)|} \tag{1.9}$$

and then are transformed back into time domain through an overlap-add procedure

$$\hat{x}(t) = \sum_{k}\sum_{\omega=0}^{M-1} \hat{X}(k,\omega)e^{2\pi i\omega\frac{t-kN}{M}}h(t-kN), \tag{1.10}$$

where h is the synthesis window. This example illustrates a feature that is common to most signal enhancement algorithms: the nonlinear estimation in the representation domain modifies only the amplitude of the transformed signal, and keeps its noisy phase. In some applications, such as speech recognition, reconstruction with noisy phase is a critical problem. The optimal solution to this problem would occur if we do not need the phase at all to perform reconstruction into the input domain. This paper addresses exactly this issue.

Consider now the problem of automatic speech recognition (ASR) systems. Given a voice signal $\{x(t),\ t = 1, 2, \ldots, T\}$, the ASR outputs a sequence of recognized phonemes from an alphabet. Most ASR systems use different kinds of *cepstral* coefficient statistics (see [5,15]) as described next. The voice signal is transformed into the time–frequency domain by the same discrete windowed Fourier transform (1.7). The (real) cepstral coefficients $C_x(k, \omega)$ are defined as the logarithm of the modulus of $X(k, \omega)$:

$$C_x(k, \omega) = \log\big(\big|X(k, \omega)\big|\big). \tag{1.11}$$

There are two rationales for using this object. First note the recorded signal $x(t)$ is a convolution of the voice signal $s(t)$ with the source-to-microphone (channel) impulse response h. In the time–frequency domain, convolution becomes (almost) multiplication, and the cepstral coefficients decouple,

$$C_x(k, \omega) = \log\big(\big|H(\omega)\big|\big) + C_s(k, \omega), \tag{1.12}$$

where $H(\omega)$ is the channel transfer function, and C_s is the voice signal cepstral coefficient. Since the channel transfer function is time-invariant, by subtracting the time average we obtain

$$F_x(k, \omega) = C_x(k, \omega) - \mathcal{E}\big[C_x(\cdot, \omega)\big] = C_s(k, \omega) - \mathcal{E}\big[C_s(\cdot, \omega)\big], \tag{1.13}$$

where $\mathcal{E}$ is the time average operator. Thus F_x encodes information about the speech signal alone, independent of the reverberant environment.

The second reason for using C_x, and thus F_x, is the widespread belief in the speech processing community that phase does not matter in speech recognition. Hence, by taking the modulus in (1.11) one does not lose information about the message (nor the messenger, as in some speaker identification algorithms).

Returning to the ASR system, the corrected cepstral coefficients F_x are fed into several hidden Markov models (HMMs), one HMM for each phoneme. The outputs of these HMMs give the utterance likelihood of a particular phoneme. The ASR system then chooses the phoneme with the largest likelihood.

In the two classes of signal processing algorithms described above the transformed domain signal either has a secondary role, or has none whatsoever. This observation led us to consider the information loss introduced by taking the modulus of a redundant representation. Clearly a constant phase is always lost, however is this the only loss of information with respect to the original signal? This is the problem we analyze in this paper.

There is also a closely connected problem in optics with applications to X-ray, crystallography, electron microscopy, and coherence theory see [4,10,11,14]. This problem is to reconstruct a discrete signal from the modulus of its Fourier transform under constraints in both the original and the Fourier domain. For finite signals the approach uses the Fourier transform with redundancy 2. All signals with the same modulus of the Fourier transform satisfy a polynomial factorization equation. In dimension one this factorization has an exponential number of possible solutions. In higher dimensions the factorization is shown to have generically a unique solution (see [13]).

The organization of the paper is as follows. Section 2 presents the analysis of real frames; Section 3 analyzes the case of complex frames.

2. Analysis of $\mathbb{M}$ for real frames

Consider the case $H = \mathbb{R}^N$, and the index set $\mathbb{I}$ has cardinality M, $\mathbb{I} = \{1, 2, \ldots, M\}$. Then $l^2(\mathbb{I}) \simeq \mathbb{R}^M$.

The set $Gr(N, M; \mathbb{R})$ of N-dimensional linear subspaces of $\mathbb{R}^M$ has the structure of an $N(M - N)$-dimensional manifold called the *Grassman manifold* [19, p. 129]. The *frame bundle* $F(N, M; \mathbb{R})$ is the

$GL(N,\mathbb{R})$-bundle over $Gr(N,M)$ defined as follows: the fiber of $F(N,M;\mathbb{R})$ over a point of $GL(N,\mathbb{R})$ corresponding to an N-dimensional linear subspace $W\subset\mathbb{R}^M$ is the set of all possible bases for W.

For a frame $\mathcal{F}=\{f_1,\dots,f_M\}$ of $\mathbb{R}^N$ we denote by T the analysis operator,

$$T:\mathbb{R}^N\to\mathbb{R}^M,\qquad T(x)=\sum_{k=1}^{M}\langle x,f_k\rangle e_k, \tag{2.1}$$

where $\{e_1,\dots,e_M\}$ is the canonical basis of $\mathbb{R}^M$. We let W denote the range of the analysis map $T(\mathbb{R}^N)$. It is an N-dimensional linear subspace of $\mathbb{R}^M$ and thus corresponds to a point of the Grassman manifold $Gr(N,M)$. Two frames $\{f_i\}_{i\in I}$ and $\{g_i\}_{i\in I}$ are *equivalent* if there is an invertible operator T on H with $T(f_i)=g_i$, for all $i\in I$. It is known that two frames are equivalent if and only if their associated analysis operators have the same range (see [2,12]). We deduce that M-element frames on $\mathbb{R}^N$ are parametrized by the fiber bundle $F(N,M;\mathbb{R})$.

Recall the nonlinear map we are interested in is

$$\mathbb{M}^{\mathcal{F}}:\mathbb{R}^N/\{\pm1\}\to\mathbb{R}^M,\qquad \mathbb{M}^{\mathcal{F}}(\hat{x})=\sum_{k=1}^{M}\big|\langle x,f_k\rangle\big|e_k,\quad x\in\hat{x}. \tag{2.2}$$

When there is no danger of confusion, we shall drop $\mathcal{F}$ from the notation.

First we reduce our analysis to equivalent classes of frames:

Proposition 2.1. *For any two frames $\mathcal{F}$ and $\mathcal{G}$ that have the same range of coefficients, $\mathbb{M}^{\mathcal{F}}$ is injective if and only if $\mathbb{M}^{\mathcal{G}}$ is injective.*

Proof. Any two frames $\mathcal{F}=\{f_k\}$ and $\mathcal{G}=\{g_k\}$ that have the same range of coefficients are equivalent, i.e., there is an invertible $R:\mathbb{R}^N\to\mathbb{R}^N$ so that $g_k=Rf_k$, $1\leqslant k\leqslant M$. Their associated nonlinear maps $\mathbb{M}^{\mathcal{F}}$, and respectively $\mathbb{M}^{\mathcal{G}}$, satisfy $\mathbb{M}^{\mathcal{G}}(x)=\mathbb{M}^{\mathcal{F}}(R^*x)$. This shows that $\mathbb{M}^{\mathcal{F}}$ is injective if and only if $\mathbb{M}^{\mathcal{G}}$ is injective. Consequently the property of injectivity of $\mathbb{M}$ depends only on the subspace of coefficients W in $Gr(N,M)$. □

This result says that for two frames corresponding to two points in the same fiber of $F(N,M;\mathbb{R})$, the injectivity of their associated nonlinear maps would jointly hold true or fail. Because of this result we shall always assume the induced topology by the base manifold $Gr(N,M)$ of the fiber bundle $F(N,M;\mathbb{R})$ into the set of M-element frames of $\mathbb{R}^N$.

If $\{f_i\}_{i\in I}$ is a frame with frame operator S then $\{S^{-1/2}f_i\}_{i\in I}$ is a Parseval frame which is equivalent to $\{f_i\}_{i\in I}$ and called the *canonical Parseval frame* associated to $\{f_i\}_{i\in I}$. Also, $\{S^{-1}f_i\}_{i\in I}$ is a frame equivalent to $\{f_i\}_{i\in I}$ and is called the *canonical dual frame* associated to $\{f_i\}_{i\in I}$. Proposition 2.1 shows that when the nonlinear map $\mathbb{M}^{\mathcal{F}}$ is injective then the same property holds for the canonical dual frame and the canonical Parseval frame.

Given $\phi\subset\{1,\dots,M\}$, let $\phi(i)$ denote the characteristic function of ϕ defined by the rule that $\phi(i)=1$ if $i\in\phi$ and $\phi(i)=0$ if $i\notin\phi$. Define a map $\sigma_\phi:\mathbb{R}^M\to\mathbb{R}^M$ by the formula

$$\sigma_\phi(a_1,\dots,a_M)=\big((-1)^{\phi(1)}a_1,\dots,(-1)^{\phi(M)}a_M\big).$$

Clearly $\sigma_\phi^2=id$ and $\sigma_{\phi^{\complement}}=-\sigma_\phi$, where $\phi^{\complement}$ is the complement of ϕ. Let L^ϕ denote the $|\phi|$-dimensional linear subspace of $\mathbb{R}^M$ where $L^\phi=\{(a_1,\dots,a_M)\mid a_i=0,\ i\in\phi\}$, and let $P_\phi:\mathbb{R}^M\to L^\phi$ denote the

orthogonal projection onto this subspace. Thus $(P_\phi(u))_i = 0$ if $i \in \phi$, and $(P_\phi(u))_i = u_i$ if $i \in \phi^{\complement}$. For every vector $u \in \mathbb{R}^M$, $\sigma_\phi(u) = u$ iff $u \in L^\phi$. Likewise $\sigma_\phi(u) = -u$ iff $u \in L^{\phi^{\complement}}$. Note

$$P_\phi(u) = \frac{1}{2}\big(u + \sigma_\phi(u)\big), \qquad P_{\phi^{\complement}}(u) = \frac{1}{2}\big(u - \sigma_\phi(u)\big).$$

Theorem 2.2 *(Real frames). If $M \geqslant 2N - 1$ then for a generic frame $\mathcal{F}$, $\mathbb{M}$ is injective.*

By *generic* we mean an open dense subset of the set of all M-element frames in $\mathbb{R}^N$.

Proof. Suppose that x and x' have the same image under $\mathbb{M} = \mathbb{M}^{\mathcal{F}}$. Let $a_1, \ldots, a_M$ be the frame coefficients of x and $a'_1, \ldots, a'_M$ the frame coefficients for x'. Then $a'_i = \pm a_i$ for each i. In particular there is a subset $\phi \subset \{1, \ldots, M\}$ of indices such that $a'_i = (-1)^{\phi(i)} a_i$. Then two vectors x, x' have the same image under $\mathbb{M}$ if and only there is a subset $\phi \subset \{1, \ldots, M\}$ such that $(a_1, \ldots, a_M)$ and $((-1)^{\phi(1)} a_1, \ldots, (-1)^{\phi(M)} a_M)$ are both in W the range of coefficients associated to $\mathcal{F}$.

To finish the proof we will show that when $M \geqslant 2N - 1$ such a condition is impossible for a generic subspace $W \subset \mathbb{R}^N$. This means that the set of such W's is a dense (Zariski) open set in the Grassmanian $Gr(N, M)$. In particular the probability that a randomly chosen W will satisfy this condition is 0.

To finish the proof of the theorem we need the following lemma.

Lemma 2.3. *If $M \geqslant 2N - 1$ then the following holds for a generic N-dimensional subspace $W \subset \mathbb{R}^M$. Given $u \in W$, then $\sigma_\phi(u) \in W$ iff $\sigma_\phi(u) = \pm u$.*

Proof of the lemma. Suppose $u \in W$ and $\sigma_\phi(u) \neq \pm u$ but $\sigma_\phi(u) \in W$. Since σ_ϕ is an involution, $u + \sigma_\phi(u)$ is fixed by σ_ϕ and is nonzero. Thus $W \cap L^\phi \neq 0$. Likewise

$$0 \neq u - \sigma_\phi(u) = u + \sigma_{\phi^{\complement}}(u).$$

Hence $W \cap L^{\phi^{\complement}} \neq 0$.

Now L^ϕ and L^{ϕ^c} are fixed linear subspaces of dimension $M - |\phi|$ and $|\phi|$. If $M \geqslant 2N - 1$ then one of these subspaces has codimension greater than or equal to N. However a generic linear subspace W of dimension N has 0 intersection with a fixed linear subspace of codimension greater than or equal to N. Therefore, if W is generic and $x, \sigma_\phi(x) \in W$ then $\sigma_\phi(x) = \pm x$ which ends the proof of lemma. □

The proof of the theorem now follows from the fact that if W is in the intersection of generic conditions imposed by the proposition for each subset $\phi \subset \{1, \ldots, M\}$ then W satisfies the conclusion of the theorem. □

Note what the above proof actually shows:

Corollary 2.4. *The map $\mathbb{M}$ is injective if and only if whenever there is a nonzero element $u \in W \subset \mathbb{R}^M$ with $u \in L^\phi$, then $W \cap L^{\phi^{\complement}} = \{0\}$.*

Next we observe that this result is best possible.

Proposition 2.5. *If $M \leqslant 2N - 2$, then the result fails for all M-element frames.*

Proof. Since $M \leqslant 2N-2$, we have that $2M-2N+2 \leqslant M$. Let $(e_i)_{i=1}^M$ be the canonical orthonormal basis of $\mathbb{R}^M$. We can write $(e_i)_{i=1}^M = (e_i)_{i=1}^k \cup (e_i)_{i=k+1}^M$, where both k and $M-k$ are $\geqslant M-N+1$.

Let W be any N-dimensional subspace of $\mathbb{R}^M$. Since $\dim W^\perp = M-N$, there exists a nonzero vector $u \in \operatorname{span}\{e_i\}_{i=1}^k$ so that $u \perp W^\perp$, hence $u \in W$. Similarly, there is a nonzero vector v in $\operatorname{span}\{e_i\}_{i=k+1}^M$ with $v \perp W^\perp$, that is $v \in W$. By the above corollary, $\mathbb{M}$ cannot be injective. In fact $\mathbb{M}(u+v) = \mathbb{M}(u-v)$. □

The next result gives an easy way for frames to satisfy the condition above.

Corollary 2.6. *If $\mathcal{F}$ is an M-element frame for $\mathbb{R}^N$ with $M \geqslant 2N-1$ having the property that every N-element subset of the frame is linearly independent, then $\mathbb{M}$ is injective.*

Proof. Given the conditions, it follows that W has no elements which are zero in N coordinates and so the corollary holds. □

Corollary 2.7. (1) *If $M = 2N-1$, then the condition given in Corollary* 2.6 *is also necessary.*
(2) *If $M \geqslant 2N$, this condition is no longer necessary.*

Proof. (1) For the first part we will prove the contrapositive. Let $M = 2N-1$ and assume there is an N-element subset $(f_i)_{i\in\phi}$ of $\mathcal{F}$ which is not linearly independent. Then there is a nonzero $x \in (\operatorname{span}(f_i)_{i\in\phi})^\perp \subset \mathbb{R}^N$. Hence, $0 \neq u = T(x) \in L^\phi \cap W$. On the other hand, since $\dim(\operatorname{span}(f_i)_{i\in\phi^c}) \leqslant N-1$, there is a nonzero $y \in (\operatorname{span}(f_i)_{i\in\phi^c})^\perp \subset \mathbb{R}^N$ so that $0 \neq v = T(y) \in L^{\phi^c} \cap W$. Now, by Corollary 2.4, $\mathbb{M}$ is not injective.

(2) If $M \geqslant 2N$ we construct an M-element frame for $\mathbb{R}^N$ that has an N-element linearly dependent subset. Let $\mathcal{F}' = \{f_1, \ldots, f_{2N-1}\}$ be a frame for $\mathbb{R}^N$ so that any N-element subset is linearly independent. By Corollary 2.4, the map $\mathbb{M}^{\mathcal{F}'}$ is injective. Now extend this frame to $\mathcal{F} = \{f_1, \ldots, f_M\}$ by $f_{2N} = \cdots = f_M = f_{2N-1}$. The map $\mathbb{M}^{\mathcal{F}}$ extends $\mathbb{M}^{\mathcal{F}'}$ and therefore remains injective, whereas clearly any N-element subset that contains two vectors from $\{f_{2N-1}, f_{2N}, \ldots, f_M\}$ is no longer linearly independent. □

Remark. The frames above can easily be constructed "by hand." Start with an orthonormal basis for $\mathbb{R}^N$, say $(f_i)_{i=1}^N$. Assume we have constructed sets of vectors $(f_i)_{i=1}^M$ such that every subset of N vectors is linearly independent. Look at the span of all of the $(N-1)$-element subsets of $(f_i)_{i=1}^M$. Pick f_{M+1} not in the span of any of these subsets. Then $(f_i)_{i=1}^{M+1}$ has the property that every N-element subset is linearly independent.

Now we will give a slightly different proof of this result which gives necessary and sufficient conditions for a frame to have the required properties.

Theorem 2.8. *Let $(f_i)_{i=1}^M$ be a frame for $\mathbb{R}^N$. The following are equivalent*:

(1) *The map $\mathbb{M}$ is injective.*
(2) *For every subset $\phi \subset \{1, 2, \ldots, M\}$, either $\{f_i\}_{i\in\phi}$ spans $\mathbb{R}^N$ or $\{f_i\}_{i\in\phi^c}$ spans $\mathbb{R}^N$.*

Proof. (1) $\Rightarrow$ (2) We prove the contrapositive. So assume that there is a subset $\phi \subset \{1, 2, \ldots, M\}$ so that neither $\{f_i;\ i \in \phi\}$ nor $\{f_i;\ i \in \{\phi^c\}$ spans $\mathbb{R}^N$. Hence there are nonzero vectors $x, y \in \mathbb{R}^N$ so that

$x \perp \operatorname{span}(f_i)_{i\in\phi}$ and $y \perp \operatorname{span}(f_i)_{i\in\phi^{\complement}}$. Then $0 \neq T(x) \in L^S \cap W$ and $0 \neq T(y) \in L^{\phi^{\complement}} \cap W$. Now by Corollary 2.4 we have that $\mathbb{M}$ cannot be injective.

(2) $\Rightarrow$ (1) Suppose $\mathbb{M}(\hat{x}) = \mathbb{M}(\hat{y})$ for some $\hat{x}, \hat{y} \in \mathbb{R}^N/\{\pm 1\}$. This means that for every $1 \leqslant j \leqslant M$, $|\langle x, f_j\rangle| = |\langle y, f_j\rangle|$, where $x \in \hat{x}$ and $y \in \hat{y}$. Let

$$\phi = \big\{j\colon \langle x, f_j\rangle = -\langle y, f_j\rangle\big\}. \tag{2.3}$$

Note

$$\phi^{\complement} = \big\{j\colon \langle x, f_j\rangle = \langle y, f_j\rangle\big\}. \tag{2.4}$$

Now, $x + y \perp \operatorname{span}(f_i)_{i\in\phi}$ and $x - y \perp \operatorname{span}(f_i)_{i\in\phi^{\complement}}$. Assume that $\{f_i;\ i \in \phi\}$ spans $\mathbb{R}^N$. Then $x + y = 0$ and thus $\hat{x} = \hat{y}$. If $\{f_i;\ i \in \phi^{\complement}\}$ spans $\mathbb{R}^N$ then $x - y = 0$ and again $\hat{x} = \hat{y}$. Either way $\hat{x} = \hat{y}$ which proves $\mathbb{M}$ is injective. □

For $M < 2N - 1$ there are plenty of frames for which $\mathbb{M}$ is not injective. However, for a generic frame, we can show the set of rays that can be reconstructed from the image under $\mathbb{M}$ is open dense in $\mathbb{R}^N/\{\pm 1\}$.

Theorem 2.9. *Assume $M > N$. Then for a generic frame $\mathcal{F} \in \mathcal{F}[N, M; \mathbb{R}]$, the set of vectors $x \in \mathbb{R}^N$ so that $(\mathbb{M}^{\mathcal{F}})^{-1}(\mathbb{M}_a^{\mathcal{F}}(x))$ consists of one point in $\mathbb{R}^N/\{\pm 1\}$ has dense interior in $\mathbb{R}^N$.*

Proof. Let $\mathcal{F}$ be a M-element frame in $\mathbb{R}^N$. Then $\mathcal{F}$ is similar to a frame $\mathcal{G}$ which consists of the union of the canonical basis of $\mathbb{R}^N$, $\{d_1, \ldots, d_N\}$, with some other set of $M - N$ vectors. Let $\mathcal{G} = \{g_k;\ 1 \leqslant k \leqslant M\}$. Thus $g_{k_j} = d_j$, $1 \leqslant j \leqslant N$, for some N elements $\{k_1, k_2, \ldots, k_N\}$ of $\{1, 2, \ldots, M\}$. Consider now the set $\mathbb{B}$ of frames $\mathcal{F}$ so that its similar frame $\mathcal{G}$ constructed above has a vector g_k with all entries nonzero,

$$\mathbb{B} = \left\{\mathcal{F} \in \mathcal{F}[N, M; \mathbb{R}] \mid \mathcal{F} \sim \mathcal{G} = \{g_k\},\ \{d_1, \ldots, d_N\} \subset \mathcal{G},\ \prod_{j=1}^{N} \langle g_{k_0}, d_j\rangle \neq 0 \text{ for some } k_0\right\}.$$

Clearly $\mathbb{B}$ is open dense in $\mathcal{F}[N, M; \mathbb{R}]$. Thus generically $\mathcal{F} \in \mathbb{B}$. Let $\mathcal{G}$ be its similar frame satisfying the condition above. We want to prove the set $X = X^{\mathcal{F}}$ of vectors $x \in \mathbb{R}^N$ so that $(\mathbb{M}^{\mathcal{G}})^{-1}(\mathbb{M}_a^{\mathcal{G}}(x))$ has more than one point is *thin*, i.e., it is included in a set whose complement is open and dense in $\mathbb{R}^N$. We claim $X \subset \bigcup_{\phi}(V_{\phi}^{+} \cup V_{\phi}^{-})$, where $(V_{\phi}^{\pm})_{\phi\subset\{1,2,\ldots,N\}}$ are linear subspaces of $\mathbb{R}^N$ of codimension 1 indexed by subsets ϕ of $\{1, 2, \ldots, N\}$. This claim will conclude the proof of theorem.

To verify the claim, let $x, y \in \mathbb{R}^N$ be so that $\mathbb{M}_a^{\mathcal{G}}(x) = \mathbb{M}_a^{\mathcal{G}}(y)$ and yet $x \neq y$, nor $x \neq -y$. Since $\mathcal{G}$ contains the canonical basis of $\mathbb{R}^N$, $|x_k| = |y_k|$ for all $1 \leqslant k \leqslant N$. Then there is a subset $\phi \subset \{1, 2, \ldots, N\}$ so that $y_k = (-1)^{\phi(k)}x_k$. Note $\phi \neq \emptyset$, nor $\phi \neq \{1, 2, \ldots, N\}$. Denote by D_ϕ the diagonal $N \times N$ matrix $(D_\phi)_{kk} = (-1)^{\phi(k)}$. Thus $y = D_\phi x$, and yet $D_\phi \neq \pm I$. Let $g_{k_0} \in \mathcal{G}$ be so that none of its entries vanishes. Then $|\langle x, g_{k_0}\rangle| = |\langle y, g_{k_0}\rangle|$ implies

$$\big\langle x, (I \pm D_\phi)g_{k_0}\big\rangle = 0.$$

This proves the set $X^{\mathcal{G}}$ is included in the union of $2(2^N - 2)$ linear subspaces of codimension 1,

$$\bigcup_{\phi\neq\emptyset,\phi^{\complement}\neq\emptyset} \big\{(I - D_\phi)g_{k_0}\big\}^{\perp} \cup \big\{(I + D_S)g_{k_0}\big\}^{\perp}.$$

Since $\mathcal{F}$ is similar to $\mathcal{G}$, $X^{\mathcal{F}}$ is included in the image of the above set through a linear invertible map, which proves the claim. □

3. Analysis of $\mathbb{M}$ for complex frames

In this section the Hilbert space is $\mathbb{C}^N$. For an M-element frame $\mathcal{F} = \{f_1, \dots, f_M\}$ of $\mathbb{C}^N$ the analysis operator is defined by (2.1), where the scalar product is $\langle x, y\rangle = \sum_{k=1}^N x_k \overline{y_k}$. The range of coefficients, i.e., the range of the analysis operator, is a complex N-dimensional subspace of $\mathbb{C}^M$ that we denote again by W. Thus a frame determines a point of the complex Grassmanian $Gr(N, M)^{\mathbb{C}}$ parametrizing N-dimensional complex subspaces of $\mathbb{C}^N$. As in the real case, the set of M-frames of $\mathbb{C}^N$ is parametrized by points of the fiber bundle $F(N, M; \mathbb{C})$, the $GL(N, \mathbb{C})$-bundle over $Gr(N, M)^{\mathbb{C}}$.

The nonlinear map we are studying is given by

$$\mathbb{M}^{\mathcal{F}} : \mathbb{C}^N/\mathbb{T}^1 \to \mathbb{C}^M, \qquad \mathbb{M}^{\mathcal{F}}(\hat{x}) = \sum_{k=1}^M \big|\langle x, f_k\rangle\big| e_k, \quad x \in \hat{x}, \tag{3.1}$$

where two vectors $x, y \in \hat{x}$ if there is a scalar $c \in \mathbb{C}$ with $|c| = 1$ so that $y = cx$.

Proposition 2.1 holds true for complex frames as well. Thus without loss of generality we shall work with the topology induced by the base manifold of $F(N, M; \mathbb{C})$ into the set of M-element frames of $\mathbb{C}^N$.

As in the real case we reduce the question about M-element frames in $\mathbb{C}^N$ to a question about the Grassmanian of N-planes in $\mathbb{C}^M$.

First we prove the following:

Theorem 3.1. *If $M \geqslant 4N-2$ then the generic N-plane W in $\mathbb{C}^M$ has the property that if $v = (v_1, \dots, v_M)$ and $w = (w_1, \dots, w_M)$ are vectors in W such that $|v_i| = |w_i|$ for all i, then $v = \lambda w$ for some complex number λ of modulus* 1.

Proof. We will say that an N-plane W has *property* $(*)$ if there are nonparallel vectors v, w in W such that $|v_i| = |w_i|$ for all i. Recall two vectors x, y are *parallel* if there is a scalar $c \in \mathbb{C}$ so that $y = cx$.

Given an N-plane W, we may assume, after reordering the coordinates on $\mathbb{C}^M$, that W is the span of the rows of an $N \times M$ matrix of the form

$$\begin{bmatrix} 1 & 0 & \dots & 0 & u_{N+1,1} & \dots & u_{M,1} \\ 0 & 1 & \dots & 0 & u_{N+1,2} & \dots & u_{M,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & u_{N+1,N} & \dots & u_{M,N} \end{bmatrix},$$

where the $N(M-N)$ entries $\{u_{i,j}\}$ are viewed as indeterminates. Thus $Gr(N, M)^{\mathbb{C}}$ is isomorphic to $\mathbb{C}^{N(M-N)}$ in a neighborhood of W.

Now suppose that W satisfies $(*)$ and v and w are two nonparallel vectors whose entries have the same modulus. Our choice of basis for W ensures that one of the first N entries in v (and hence w) are nonzero. Since we only care about these vectors up to rescaling we may assume, after reordering, that $v_1 = w_1 = 1$. Also the vectors are assumed nonparallel so we may assume that $v_i \neq w_i \neq 0$ for some $i \leqslant N$. After yet again reordering we can assume that $v_2 \neq w_2 \neq 0$.

Set $\lambda_1 = 1$. By assumption there are numbers $\lambda_2, \dots, \lambda_M \in \mathbb{T}^1$ with $\lambda_2 \neq 1$ such that $w_i = \lambda_i v_i$ for $i = 1, \dots, M$. Expanding in terms of the basis for W we have for $i > N$, $v_i = \sum_{j=1}^N v_j u_{i,j}$ and $w_i =$

$\sum_{j=1}^{N} \lambda_j v_j u_{i,j}$. Thus if W satisfies $(*)$ there must be $\lambda_2, \ldots, \lambda_N \in \mathbb{T}^1$ (with $\lambda_2 \neq 1$) and $v_2, \ldots, v_N \in \mathbb{C}$ such that for all $N+1 \leqslant i \leqslant M$ we have

$$\left| \sum_{j=1}^{N} v_j u_{i,j} \right| = \left| \sum_{j=1}^{N} \lambda_j v_j u_{i,j} \right|. \tag{3.2}$$

Consider the variety Y of all tuples

$$(W, v_2, \ldots, v_N, \lambda_2, \ldots, \lambda_N)$$

as above. Since $v_2 \neq 0$ and $\lambda_2 \neq 1$ this variety is locally isomorphic to the real $(2N(M-N)+3N-3)$-dimensional variety

$$\mathbb{C}^{N(M-N)} \times \left(\mathbb{C} \setminus \{0\}\right) \times (\mathbb{C})^{N-2} \times \left(\mathbb{T}^1 \setminus \{1\}\right) \times \left(\mathbb{T}^1\right)^{N-2}.$$

The locus in $Gr(N, M)^{\mathbb{C}}$ of planes satisfying property $(*)$ is denoted by X. This variety is the image under projection to the first factor of Y cut out by the $M-N$ equations (3.2) for $N+1 \leqslant i \leqslant M$. The analysis of these equations is summarized by the following result.

Lemma 3.2. *The $M-N$ equations in* (3.2) *are independent. Hence X is a variety of real dimension at most $2N(M-N)+3N-3-(M-N)$.*

Proof of Lemma 3.2. For any choice of $0 \neq v_2, v_3, \ldots, v_N$ and $1 \neq \lambda_2, \lambda_3, \ldots, \lambda_N$ the equation

$$\left| \sum_{j=1}^{M} v_j u_{i,j} \right|^2 = \left| \sum_{j=1}^{M} \lambda_j v_j u_{i,j} \right|^2$$

is nondegenerate. Since the variables $u_{i,1}, \ldots, u_{i,N}$ appear in exactly one equation, these equations (for fixed $v_2, v_3, \ldots, v_N, \lambda_2, \ldots, \lambda_N$) define a subspace of $\mathbb{C}^{N(M-N)}$ of real codimension at least $M-N$. Since this is true for all choices, it follows that the equations are independent. □

From this lemma it follows that the locus of N-planes satisfying $(*)$ has (local) real dimension $2N(M-N)+3N-3-(M-N)$. Therefore if $3N-3-(M-N) < 0$, i.e., if $M \geqslant 4N-2$, this locus cannot be all of $Gr(N, M)^{\mathbb{C}}$. This ends the proof of Theorem 3.1. □

The main result in the complex case then follows from Theorem 3.1.

Theorem 3.3 *(Complex frames). If $M \geqslant 4N-2$ then $\mathbb{M}^{\mathcal{F}}$ is injective for a generic frame $\mathcal{F} = \{f_1, \ldots, f_N\}$.*

Lemma 3.2 yields the following result.

Theorem 3.4. *If $M \geqslant 2N$ then for a generic frame $\mathcal{F} \in \mathcal{F}[N, M; \mathbb{C}]$ the set of vectors $x \in \mathbb{C}^N$ such that $(\mathbb{M}^{\mathcal{F}})^{-1}(\mathbb{M}_a^{\mathcal{F}}(x))$ has one point in $\mathbb{C}^N/\mathbb{T}^1$ has dense interior in $\mathbb{C}^N$.*

Proof. By Lemma 3.2, for a generic frame the $M-N$ equations (3.2) in $2(N-1)$ indeterminates $(v_2, \ldots, v_N, \lambda_2, \ldots, \lambda_N)$ are independent. Note there are $3(N-1)$ real valued unknowns and $M-N$

equations. Hence the set of $\{(v_2, \ldots, v_N)\}$ in $\mathbb{C}^{N-1}$ for which there are $(\lambda_2, \ldots, \lambda_N)$ such that (3.2) has solution in $(\mathbb{C} \setminus \{0\}) \times (\mathbb{C})^{N-2} \times (\mathbb{T}^1 \setminus \{1\}) \times (\mathbb{T}^1)^{N-2}$ has real dimension at most $3(N-1) - (M-N) = 4N - 3 - M$. For $M \geqslant 2N$ it follows $3(N-1) - (M-N) < 2(N-1)$ which shows the set of $v = (v_1, \ldots, v_N)$ such that $(\mathbb{M}^{\mathcal{F}})^{-1}(\mathbb{M}_a^{\mathcal{F}}(v))$ has more than one point is thin in $\mathbb{C}^N$, i.e., its complement has dense interior. □

We do not know the precise optimal bound for the complex case but we believe it is $4N - 2$. However, this case is different from the real case in that complex frames with only $2N - 1$ elements cannot have $\mathbb{M}^{\mathcal{F}}$ injective. To see this we observe that the proof of Theorem 2.8 (1) ⇒ (2) does not use the fact that the frames are real. So in the complex case we have:

Proposition 3.5. *If $\{f_j\}_{j\in I}$ is a complex frame and $\mathbb{M}^{\mathcal{F}}$ is injective, then for every $\phi \subset \{1, 2, \ldots, M\}$ if $L^{\phi} \cap W \neq \{0\}$ then $L^{\phi^c} \cap W = \{0\}$. Hence, for every such ϕ, either $\{f_j\}_{j\in\phi}$ or $\{f_j\}_{j\in\phi^c}$ spans H.*

Now we can show that complex frames must contain at least $2N$ elements for $\mathbb{M}^{\mathcal{F}}$ to be injective.

Proposition 3.6 *(Complex frames). If $\mathbb{M}^{\mathcal{F}}$ is injective then $M \geqslant 2N$.*

Proof. We assume that $M = 2N - 1$ and show that in this case $\mathbb{M}^{\mathcal{F}}$ is not injective. Let $\{z_j\}_{j=1}^N$ be a basis for W and let P be the orthogonal projection onto the first $N - 1$ unit vectors in $\mathbb{C}^M$. Then $\{Pz_j\}_{j=1}^N$ sits in an $(N-1)$-dimensional space and so there are complex scalars $\{a_j\}_{j=1}^{N-1}$, not all zeros, so that $\sum a_j Pz_j = 0$. That is, there is a vector $0 \neq y \in W$ with support $y \subset \{N, N+1, \ldots, 2N-1\}$. Similarly, there is a vector $0 \neq x \in W$ with support $x \subset \{1, 2, \ldots, N\}$. If $x(N) = 0$ or $y(N) = 0$ we contradict Proposition 3.4. Also, if $x(i) = 0$ for all $i < N$, then $(y - cx)(N) = 0$ for $c = y(N)\frac{\overline{x(N)}}{|x(N)|^2}$. Now, x, $y - cx$ are in W and have disjoint support so our map is not injective. Otherwise, let

$$z = \frac{\overline{x(N)}}{|x(N)|^2}, \qquad w = i\frac{\overline{y(N)}}{|y(N)|^2}.$$

Now, $z, w \in W$ and $z(N) = 1$ and $w(N) = i$. Hence, $|z + w| = |z - w|$. It follows that there is a complex number $|c| = 1$ so that $z + w = c(z - w)$. Since $z_i \neq 0$ for some $i < N$ we have that $c = 1$ and $w = 0$ which is a contradiction. □

4. Implementation of these results

For these results to be widely applied they need to run on existing software with only trivial modifications. So there are two critical issues that need to be addressed for implementation of signal reconstruction without phase. (1) Find Gabor frames which work in this setting—so we can use the fast Fourier transform for digitalizing the signal. (2) Find efficient reconstruction algorithms—preferably algorithms which are close to the inverse fast Fourier transform. These two problems are the focus of current research on this topic [3]. It appears at this time that small frames near the threshold of our results ($(2N-1)$ elements in the real case and $(4N-2)$ elements in the complex case) may require exponential time for reconstruction. However, it is shown in [3] that generic frames with N^2-elements give polynomial time reconstruction (on the order of at most N^6 calculations). In [3] there are some special classes of frames

with N^2 elements which have extremely efficient algorithms for reconstruction in N calculations ($2N$ in the complex case).

References

[1] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice Hall, Englewood Cliffs, NJ, 1979.
[2] R. Balan, Equivalence relations and distances between Hilbert frames, Proc. Amer. Math. Soc. 127 (8) (1999) 2353–2366.
[3] R. Balan, P.G. Casazza, D. Edidin, Algorithms for reconstruction without phase, in preparation.
[4] R.H. Bates, D. Mnyama, The status of practical Fourier phase retrieval, in: W.H. Hawkes (Ed.), Advances in Electronics and Electron Physics, vol. 67, 1986, pp. 1–64.
[5] C. Becchetti, L.P. Ricotti, Speech Recognition. Theory and C++ Implementation, Wiley, New York, 1999.
[6] O. Christensen, An Introduction to Frames and Riesz Bases, Birkhäuser, Boston, 2003.
[7] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. 32 (6) (1984) 1109–1121.
[8] J.G. Proakis, et al., Discrete-Time Processing of Speech Signals, IEEE Press, New York, 2000.
[9] J.G. Proakis, et al., Algorithms for Statistical Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 2002.
[10] J.R. Fienup, Reconstruction of an object from the modulus of its Fourier transform, Opt. Lett. 3 (1978) 27–29.
[11] J.R. Fienup, Phase retrieval algorithms: A comparison, Appl. Opt. 21 (15) (1982) 2758–2768.
[12] D. Han, D. Larson, Frames, bases and group representations, Mem. Amer. Math. Soc. 147 (697) (2000).
[13] M.H. Hayes, The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform, IEEE Trans. ASSP 30 (2) (1982) 140–154.
[14] G. Liu, Fourier phase retrieval algorithm with noise constraints, Signal Process. 21 (4) (1990) 339–347.
[15] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice Hall Signal Processing Series, Prentice Hall, Englewood Cliffs, NJ, 1993.
[16] R.F. Streater, A.S. Wightman, PCT, Spin and Statistics and All That, Landmarks in Mathematics and Physics, Princeton Univ. Press, Princeton, NJ, 2000.
[17] H.L. van Trees, Optimum Array Processing, Wiley, New York, 2002.
[18] S.V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, Wiley, New York, 2000.
[19] F.W. Warner, Foundations of Differential Manifolds and Lie Groups, Graduate Texts in Mathematics, vol. 94, Springer-Verlag, Berlin, 1983.