Randomized Signal Processing with Continuous Frames

Ron Levie[1] · **Haim Avron[2]**

Abstract

This paper focuses on signal processing tasks in which the signal is transformed from the signal space to a higher dimensional coefficient space (also called phase space) using a continuous frame, processed in the coefficient space, and synthesized to an output signal. We show how to approximate such methods, termed phase space signal processing methods, using a Monte Carlo method. As opposed to standard discretizations of continuous frames, based on sampling discrete frames from the continuous system, the proposed Monte Carlo method is directly a quadrature approximation of the continuous frame. We show that the Monte Carlo method allows working with highly redundant continuous frames, since the number of samples required for a certain accuracy is proportional to the dimension of the signal space, and not to the dimension of the phase space. Moreover, even though the continuous frame is highly redundant, the Monte Carlo samples are spread uniformly, and hence represent the coefficient space more faithfully than standard frame discretizations.

Keywords Signal processing · Continuous frames · Stochastic methods · Time–frequency analysis · Phase vocoder

Mathematics Subject Classification 42C15 · 42C40 · 65C05 · 65R10 · 65T60 · 46B15

✉ Ron Levie
    levie@math.lmu.de

    Haim Avron
    haimav@tauex.tau.ac.il

[1] Department of Mathematics, Ludwig Maximilian University of Munich, Munich, Germany

[2] Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel

1 Introduction

We consider signal processing tasks based on continuous frames [1,46]. In the general setting, an input signal s, from the Hilbert space of signals $\mathcal{H}$, is first analyzed into a coefficient space representation $V_f[s]$ via the frame analysis operator V_f. Then, $V_f[s]$ is manipulated in the coefficient space by first applying a pointwise nonlinearity $r \circ V_f[s]$ and then a linear operator T, to produce $T(r \circ V_f[s])$, which is finally synthesized back to the signal space via V_f^*. The end-to-end pipeline reads

$$\mathcal{H} \ni s \mapsto V_f^* T(r \circ V_f[s]). \tag{1}$$

We call pipelines of the form (1) *phase space signal processing*. The linear operator T models a global change of the signal in the feature space, while the nonlinearity r allows modifying the feature coefficients term-by-term with respect to their values.

Some examples of continuous frames are the 1D continuous wavelet transform—CWT [13,26], the short time Fourier transform - STFT [25], the Shearlet transform [28] and the Curvelet transform [7]. Signal processing tasks of the form (1) are used in a multitude of applications. In multipliers [2,3,37,40,47], T is a multiplicative operator and the nonlinearity is trivial $r(x) = x$. Multipliers have applications, for example, in audio analysis [4] and increasing signal-to-noise ratio [36]. In signal denoising, e.g., wavelet shrinkage denoising [14,15], and Shearlet denoising [29], the linear operator is trivial $T = I$ and r is a nonlinearity that attenuates low values. The same is true in shearlet based image enhancement [21, Section 4]. In phase vocoder, T is a dilation operator and r is a so-called phase correction nonlinearity [12,16,32,34,41, 44,45,52]. In analysis-based iterative thresholding algorithms for inverse problems, the sparsification step in each iteration can be written as (1) with $T = I$ and r a thresholding nonlinearity [22,30].

We note that the theory presented in this paper works also when the analysis and synthesis in (1) are done by two different frames sharing the same feature space. An example application is shearlet or curevelet based Radon transform inversion [9, 11,20], where T is a multiplicative operator, r is thresholding, analysis is done by the curvelet/shearlet frame, and synthesis is done by some modified curvelet/shearlet frame. For simplicity of the presentation, we stick to the same frame for analysis and synthesis. More accurately, in (5) and (6) we extend (1) to a pipeline based on a frame and its canonical dual.

In this paper we study quadrature discretizations of (1) based on random samples.

1.1 Quadrature Versus Discrete Frame Discretizations of Continuous Frames

An analysis operator of a continuous frame V_f has the form

$$\mathcal{H} \ni s \mapsto V_f[s] = \langle s, f_{(\cdot)} \rangle \in L^2(G). \tag{2}$$

where G is a measure space called *phase space*, that usually has some physical interpretation (e.g., in the STFT G is the time-frequency plane), $\{f_g\}_{g \in G}$ is the continuous

frame, and $L^2(G)$ is called the coefficient space. Accordingly, the synthesis operator V_f^* has the form [46, Theorem 2.6]

$$V_f^*(F) = \int_G F(g) f_g \, dg. \tag{3}$$

As evident from the above description, phase space signal processing involves integrals, and thus some form of discretization is required. The common approach is to sample points from G to construct a discrete frame version $\{f_{g^k}\}_{k=1}^{\infty}$ of the continuous frame (e.g., as in [13,25]). For the discrete frame, the synthesis operator reads, for $(F_k)_k \in l^2$,

$$V_{\{f_{g^k}\}_n}^* \{F_k\}_k = \sum_{k=1}^{\infty} f_{g^k} F_k. \tag{4}$$

Note that (4) looks like a quadrature approximation of (3). However, in the standard continuous-to-discrete frame approach, the points g^k are not chosen for approximating (3). Rather, g^k are chosen so that the discrete system $\{f_{g^k}\}_{n=1}^{\infty}$ satisfies the discrete frame inequality. Hence, the discrete frame is related to the continuous one by the fact that f_{g^k} are sampled from $f_{(\cdot)}$, not by some approximation requirement. In this paper we take a different route to discretization, requiring that (4) approximates (3). Let us call the latter discretization approach the *quadrature approach*.

There is an advantage in the quadrature approach over the discrete frame approach when working with highly redundant continuous frames. To illustrate the idea, consider the STFT, where $G = \mathbb{R}^2$ is the time-frequency plane. Suppose that we extend G by adding to the time and frequency axes t, ω a third axis c that controls the time width of the window. To discretize the resulting continuous frame $f_{t,\omega,c}$ to a discrete frame, one may simply choose to fix the window width axis to one single value $c = c^0$, and sample a time-frequency 2D grid $(t^n, \omega^m, c^0)_{n,m}$. Indeed, such an approach would result in a standard discrete STFT, which is a discrete frame. However, the information along the third axis is lost in this discretization. In fact, nothing in the continuous-to-discrete frame approach forces the discretization to faithfully represent the whole continuous phase space. In the quadrature approach, our goal is to discretize the continuous frames more faithfully, sampling uniformly all the feature directions in phase space.

1.2 Randomized Quadrature Approximations of Continuous Frames

Our motivation is hence to discretize highly redundant continuous frames in the quadrature approach. In such situations, the dimensionality of phase space is higher than that of the signal space, and hence using a randomized discretization makes sense. Our approach is motivated by randomized methods in finite-dimensional numerical linear algebra, which are a prominent approach for dealing with high dimensional data [10,17,18,35,50,51]. The goal in this paper is to develop a similar randomized theory in an infinite dimensional setting in general separable Hilbert spaces, namely, in the phase space signal processing setting.

Randomized algorithms in a context of continuous frames were presented in the past. *Relevant sampling* is a line of work in which integral transforms are randomly discretized [5,24,42,49]. While the goal in our approach is to approximate the continuous frame with a quadrature sum, the goal in relevant sampling is to sample discrete frames from continuous frames.

We summarize our construction as follows.

Signal processing in phase space Let $V_f : \mathcal{H} \to L^2(G)$ be the analysis operator of a continuous frame, and $S_f = V_f^* V_f$ be the frame operator. Since $S_f^{-1} V_f^* V_f = V_f^* V_f S_f^{-1}$ is the identity I, we consider the following two formulations of signal processing in phase space. *Synthesis phase space signal processing* is defined by the pipeline

$$s \mapsto V_f^* Tr \circ \left(V_f[S_f^{-1} s] \right), \tag{5}$$

and *analysis phase space signal processing* is defined by

$$s \mapsto S_f^{-1} V_f^* Tr \circ \left(V_f[s] \right). \tag{6}$$

Here, T is a bounded operator in $L^2(G)$ and $r : \mathbb{C} \to \mathbb{C}$ is a nonlinearity. The pipelines (5) and (6) can be seen as working with the canonical dual frame $S_f^{-1} f$ [46] either in the analysis or in the synthesis step. We suppose that S_f has an efficient discretization in the signal space $\mathcal{H}$. Hence, we would like to find an efficient discretization of the rest of the pipeline, namely, of $V_f^* Tr \circ \left(V_f[s] \right)$.

Mote Carlo signal processing in phase space We study a Monte Carlo approximation of signal processing in phase space based on the pipelines (5) and (6). We first consider a Monte Carlo approximation of synthesis. For $F \in L^2(G)$, under certain assumptions given in Sect. 3.5, we consider the approximation of the synthesis operator by

$$V_f^* F = \int_G F(g) f_g dg \approx \frac{C}{K} \sum_{k=1}^{K} F(g^k) f_{g^k}. \tag{7}$$

Here, $\{g^k\}_{k=1}^K \subset G$ is a finite set of independent random sample points, and C is a normalization constant.

Using this approximation, in Sect. 4 we study the approximation rate of stochastic signal processing in phase space (5) and (6). The first version of the approximation reads, for the synthesis and analysis formulations respectively,

$$s \mapsto \frac{C}{K} \sum_{k=1}^{K} \left(T(r \circ V_f[S_f^{-1} s]) \right)(g^k) f_{g^k}, \tag{8}$$

$$s \mapsto \frac{C}{K} S_f^{-1} \sum_{k=1}^{K} \left(T(r \circ V_f[s]) \right)(g^k) f_{g^k}. \tag{9}$$

Under some general assumptions, we also approximate the signal processing pipelines (5) and (6) when T in an integral operator defined by

$$T F(g) = \int_G R(g, g') F(g') dg',$$

where $R : G^2 \to \mathbb{C}$ (Definition 6). The synthesis and analysis approximations in this case read respectively

$$s \mapsto \frac{C}{KL} \sum_{k=1}^{K} \sum_{m=1}^{L} R(g^k, y^m) r\left(V_f[S_f^{-1}s](f_{y^m})\right) f_{g^k}, \qquad (10)$$

$$s \mapsto \frac{C}{KL} S_f^{-1} \sum_{k=1}^{K} \sum_{m=1}^{L} R(g^k, y^m) r\left(V_f[s](f_{y^m})\right) f_{g^k}. \qquad (11)$$

Here, $\{g^k\}_{k=1}^{K}$, $\{y^m\}_{m=1}^{L} \subset G$ are two finite sets of independent random sample points.

Methods (10) and (11) are useful for integral operators. Methods (8) and (9) are useful when the samples $(T r \circ V_f[s])(g^k)$ can be computed using some other samples $V_f[s](y^k)$ of $V_f[s]$, which is the case for multiplicative and diffeomorphism operators (see Sect. 6.2).

Summary of our main results

– We prove the convergence of the Monte Carlo methods (8)–(11) as the number as samples increase, and also introduce non-asymptotic error bounds. When considering discrete signals of resolution/dimension M, embedded in the continuous signal space, the error in the stochastic method is of order $O(\sqrt{M/K})$, where K is the number of Monte Carlo samples.
– The computational complexity of our method does not depend on the dimension of phase space for a rich class of signal processing pipelines. This allows approximating highly redundant continuous frames efficiently using sample points which are well spread in all directions in phase space.
– As a toy application of the theory, we show how to increase the expressive capacity of the 2D time-frequency phase space by adding a third axis, and use the construction in a phase vocoder scheme.

The proofs in this paper are inspired by the constructions in standard Monte Carlo theory (see, e.g., [6, Section 2]), adapted to infinite dimensional Hilbert spaces.

2 Background: Harmonic Analysis in Phase Space

In this section we review the theory of continuous frames, and give the two examples: STFT and CWT. By convention, all Hilbert spaces in this paper are assumed to be separable. The Fourier transform $\mathcal{F}$ is defined with the following normalization

$$[\mathcal{F}s](\omega) = \hat{s}(\omega) = \int_{\mathbb{R}} s(t) e^{-2\pi i \omega t} dt, \quad [\mathcal{F}^{-1}\hat{s}](t) = \int_{\mathbb{R}} \hat{s}(\omega) e^{2\pi i \omega t} d\omega. \quad (12)$$

We denote the norm of a vector v in a Banach space $\mathcal{B}$ by $\|v\|_{\mathcal{B}}$. For a measure space $\{G, \mu\}$, we denote interchangeably by

$$\|f\|_p = \|f\|_{L^p(G)} = \left(\int_G |f(g)|^p \, d\mu(g) \right)^{1/p}$$

the p norm of the signal $f \in L^p(G)$, where $1 \leq p < \infty$, and denote interchangeably

$$\|f\|_\infty = \|f\|_{L^\infty(G)} = \operatorname{ess\,sup}_{g \in G} |f(g)|$$

for $f \in L^\infty(G)$. We note that an equality between two L^p functions is by definition an almost-everywhere equality. We denote the induced operator norm of operators in Banach spaces using the subscript of the Banach space, e.g., for a bounded linear operator $T : L^p(G) \to L^p(G)$, we denote $\|T\|_p$. When we want to emphasize that the norm is an operator norm, we also denote $\|T\|_{p \to p}$.

2.1 Continuous Frames

The following definitions and claims are from [46] and [23, Chapter 2.2], with notation adopted from the later.

Definition 1 Let $\mathcal{H}$ be a Hilbert space, and $(G, \mathcal{B}, \mu)$ a locally compact topological space with Borel sets $\mathcal{B}$, and σ-finite Borel measure μ. Let $f : G \to \mathcal{H}$ be a weakly measurable mapping, namely for every $s \in \mathcal{H}$

$$g \mapsto \langle s, f_g \rangle$$

is a measurable function $G \to \mathbb{C}$. For any $s \in \mathcal{H}$, we define the *coefficient function*

$$V_f[s] : G \to \mathbb{C} \quad , \quad V_f[s](g) = \langle s, f_g \rangle_{\mathcal{H}}. \tag{13}$$

1. We call f a *continuous frame*, if $V_f[s] \in L^2(G)$ for every $s \in \mathcal{H}$, and there exist constants $0 < A \leq B < \infty$ such that

$$A \|s\|_{\mathcal{H}}^2 \leq \|V_f[s]\|_2^2 \leq B \|s\|_{\mathcal{H}}^2 \tag{14}$$

   for every $s \in \mathcal{H}$.
2. If it is possible to choose $A = B$, f is called a *tight frame*.
3. We call $\mathcal{H}$ the *signal space*, G *phase space*, V_f the *analysis operator*, and V_f^* the *synthesis operator*.
4. We call the frame f *bounded*, if there exist a constant $0 < C \in \mathbb{R}$ such that

$$\forall g \in G , \|f_g\|_{\mathcal{H}} \leq C.$$

5. We call $S_f = V_f^* V_f$ the *frame operator*, and $Q_f = V_f V_f^*$ the *Gramian operator*.
6. We call f a *Parseval* continuous frame, if V_f is an isometry between $\mathcal{H}$ and $L^2(G)$.

Remark 2 A frame is Parseval if and only if the frame bounds cane be chosen as $A = B = 1$.

For the closed form formula of the synthesis operator V_f^*, we recall the notion of weak vector integrals, also called Pettis integral, introduced in [43].

Definition 3 Let $\mathcal{H}$ be a separable Hilbert space, and G a measure space. Let $v : G \to \mathcal{H}$ be a mapping such that the mapping $s \mapsto \int_G \langle s, v(g) \rangle \, dg$ is continuous in $s \in \mathcal{H}$. Then the *weak vector integral* (or weak $\mathcal{H}$ integral) is defined to be the vector $\int_G^{\mathrm{w}} v(g) dg \in \mathcal{H}$ such that

$$\forall s \in \mathcal{H} \quad \int_G \langle s, v(g) \rangle \, dg = \left\langle s, \int_G^{\mathrm{w}} v(g) dg \right\rangle.$$

The existence of such a vector is guaranteed by Riesz representation theorem. In this case, v is called a *weakly integrable function*.

Given a continuous frame, the synthesis operator can be written by [46, Theorem 2.6]

$$V_f^*[F] = \int_G^{\mathrm{w}} F(g) f_g dg. \tag{15}$$

Definition 4 The *frame kernel* $K_f : G^2 \to \mathbb{C}$ is defined by

$$K_f(g, g') = \langle f_g, f_{g'} \rangle = V_f[f_g](g'). \tag{16}$$

The following result is taken from [23, Proposition 2.12].

Proposition 5 *The Gramian operator Q_f is an integral operator with kernel K. Namely, for every $F \in L^2(G)$*

$$[Q_f F](g') = \int_G F(g) K_f(g, g') dg. \tag{17}$$

For a Parseval frame, the image space $V_f[\mathcal{H}]$ is a reproducing kernel Hilbert space, with kernel $K_f(g, \cdot)$, and the orthogonal projection upon $V_f[\mathcal{H}]$ is given by the Gramian operator $Q_f = V_f V_f^*$.

2.2 Examples

An important class of Parseval frames are wavelet transforms based on square integrable representations, which we call in this paper simply wavelet transforms. We refer the reader to [23, Chapters 2.3–2.5], and the classical papers [19,27]. The wavelet system in the general theory is generated by fixing one signal $f \in \mathcal{H}$, that is typically called the *mother wavelet* or the *window function*, and applying on it a set of transformations $\{\pi(g) f \mid g \in G\}$, parameterized by a locally compact topological group

G. The Haar measure is taken in G, and $\pi : G \to \mathcal{U}(\mathcal{H})$ is assumed to be a *square integrable representation*, where $\mathcal{U}(\mathcal{H})$ is the group of unitary operators in $\mathcal{H}$.

The wavelet transform is defined by

$$V_f : \mathcal{H} \to L^2(G) \quad , \quad V_f[s](g) = \langle s, \pi(g)f \rangle .$$

For any two mother wavelets f_1 and f_2, the reconstruction formula of the wavelet transform is given by

$$s = \frac{1}{\langle Af_2, Af_1 \rangle} V_{f_2}^* V_{f_1}(s) = \frac{1}{\langle Af_2, Af_1 \rangle} \int_G^w V_{f_1}[s](g)\pi(g)f_2 \, dg.$$

Here, A is a special positive operator in $\mathcal{H}$, called the *Duflo-Moore operator*, uniquely defined for every square integrable representation π, that determines the normalization of windows.

The short time Fourier transform The following construction is taken from [25]. Consider the signal space $L^2(\mathbb{R})$. Let $\mathcal{T} : \mathbb{R} \to \mathcal{U}(L^2(\mathbb{R}))$ be the translation in $L^2(\mathbb{R})$, defined for $x \in \mathbb{R}$ and $f \in L^2(\mathbb{R})$ by $[\mathcal{T}(x)f](t) = f(t - x)$. Let $\mathcal{M} : \mathbb{R} \to \mathcal{U}(L^2(\mathbb{R}))$ be the modulation in $L^2(\mathbb{R})$, defined for $\omega \in \mathbb{R}$ and $f \in L^2(\mathbb{R})$ by $[\mathcal{M}(\omega)f](t) = e^{2\pi i \omega t} f(t)$. Denote $\pi(x, \omega) = \mathcal{T}(x)\mathcal{M}(\omega)$. For a normalized window f, the mapping

$$\mathbb{R}^2 \ni (x, \omega) \mapsto \pi(x, \omega)f$$

is a Parseval continuous frame, with the standard Lebesgue measure of the phase space $\mathbb{R}^2$. The resulting transform $V_f[s](x, \omega) = \langle s, \pi(x, \omega)f \rangle$ is called the *Short Time Fourier Transform* (STFT).

2.2.1 The 1D Continuous Wavelet Transform

The following construction is taken from [13,26]. Consider the signal space $L^2(\mathbb{R})$, and the translation $\mathcal{T}$ as in the STFT. Let $\mathcal{D} : \mathbb{R} \setminus \{0\} \to \mathcal{U}(L^2(\mathbb{R}))$ be the dilation in $L^2(\mathbb{R})$, defined for $\tau \in \mathbb{R} \setminus \{0\}$ and $f \in L^2(\mathbb{R})$ by $[\mathcal{D}(\tau)f](t) = \frac{1}{\sqrt{|\tau|}} f(\frac{t}{\tau})$. The set of transformations

$$\mathcal{A} = \{\mathcal{T}(x)\mathcal{D}(\tau) \mid (x, \tau) \in \mathbb{R} \times (\mathbb{R} \setminus \{0\})\} \tag{18}$$

is closed under compositions. We can treat $\mathcal{A}$ as a group of tuples $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$, with group product derived from the compositions of operators in (18). The group $\mathcal{A}$ is called the 1D *affine group*. The mapping

$$\pi(x, \tau) = \mathcal{T}(x)\mathcal{D}(\tau)$$

is a square integrable representation, with Dulfo-Moore operator A defined by $[\mathcal{F}A\mathcal{F}^*\hat{f}](z) = \frac{1}{\sqrt{|z|}} \hat{f}(z)$, where $\mathcal{F}$ is the Fourier transform. The resulting wavelet transform is called the *Continuous Wavelet Transform* (CWT).

Next, we show how the CWT atoms are interpreted as time-frequency atoms, and the CWT is interpreted as a time-frequency transform. Here, by changing variable $\omega = \frac{1}{\tau}$, we obtain the Parseval frame

$$\{\pi'(x, \omega)f\}_{(x,\omega)\in\mathbb{R}\times(\mathbb{R}\setminus\{0\})}$$

based on the representation $\pi'(x, \omega) = \mathcal{T}(x)\mathcal{D}(\omega^{-1})$. The parameter ω is interpreted as frequency. The mapping π' is a representation of the 1D affine group with the new parameterization $\omega = \frac{1}{\tau}$, in which the Haar measure is the standard Lebesgue measure of $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$.

3 Elements of Stochastic Signal Processing in Phase Space

In this section we develop basic approximation results that will be used later in the paper to bound the approximation error between the signal processing pipelines (5) and (6), and their stochastic approximations (7)–(11).

3.1 Phase Space Operators

We start by defining integral operators in the coefficient space.

Definition 6 Let T be a bounded linear operator in $L^2(G)$, where G is a locally compact topological space with σ-finite Borel measures.

1. We call T a *phase space integral operator (PSI operator)* if there exists a measurable function $R : G \times G \to \mathbb{C}$ with $R(\cdot, g) \in L^2(G)$ for almost every $g \in G$, such that for every $F \in L^2(G)$

$$TF = \int_G R(\cdot, g)F(g)dg. \qquad (19)$$

2. A phase space integral operator T is called *uniformly square integrable*, if there is a constant $D > 0$ such that for almost every $g \in G$

$$\|R(\cdot, g)\|_{L^2(G)} = \sqrt{\int_G |R(g', g)|^2\, dg'} \le D. \qquad (20)$$

Example 7 The Gramian operator Q_f of a continuous frame is a phase space operator by Proposition 5, with $\|Q_f\|_2 \le B$. If f is bounded, with bound $\|f_g\|_{\mathcal{H}} \le C$, then Q_f is uniformly square integrable with bound

$$\|K_f(g, \cdot)\|_2 = \sqrt{\int_G |K_f(g, g')|^2\, dg'} = \sqrt{\int_G |V_f[f_g](g')|^2\, dg'} \le B^{1/2}C.$$

3.2 Sampling in Phase Space

Let $F \in L^2(G)$, and let f be a continuous frame. The phase space G in general does not have finite measure, and thus uniform sampling is not defined on G. However, when G has infinite measure, functions $F \in L^2(G)$ must decay in some sense "at infinity", so it is possible to restrict our sampling to a compact domain in G, in which F has most of its energy. More accurately, since G is σ-finite, it is the disjoint union of at most countably many sets of finite measure. Namely, there are disjoint measurable sets X_n of finite measure, with $\bigcup_{n \in \mathbb{N}} X_n = G$, such that for every $F \in L^2(G)$

$$\|F\|_{L^2(G)}^2 = \int_G |F(g)|^2 \, dg = \sum_{n \in \mathbb{N}} \int_{X_n} |F(g)|^2 \, dg = \sum_{n \in \mathbb{N}} \|F\|_{L^2(X_n)}^2 . \quad (21)$$

Denote $G_n = \bigcup_{j=1}^n X_n$, and note that $\bigcup_{n \in \mathbb{N}} G_n = G$. Now, (21) is equivalent to

$$\|F\|_{L^2(G)}^2 = \lim_{n \to \infty} \|F\|_{L^2(G_n)}^2 .$$

Thus, for every $\epsilon > 0$, there exists an indicator function ψ_ϵ (that depends on F) of a measurable set of finite measure $\|\psi_\epsilon\|_1$, such that

$$\|\psi_\epsilon F - F\|_2 < \epsilon. \quad (22)$$

In our analysis, we allow more general forms of *envelopes* ψ_ϵ.

Definition 8 An *envelope* is a positive $\psi \in L^1(G) \cap L^\infty(G)$ satisfying $\|\psi\|_\infty \leq 1$.

Given an envelope ψ, samples can be drawn from G according to the probability density $\frac{\psi(g)}{\|\psi\|_1}$. In the following analysis we fix an envelope $\psi_\epsilon = \psi$ independently of a specific function $F \in L^2(G)$. This is the common approach in classical signal processing, where a compact frequency band $[a, b]$ is predefined independently of a specific signal. It is implicit that we can only treat signals having most of their frequency energy in $[a, b]$. Any frequency information outside of $[a, b]$ is lost or projected into the band. In Sect. 5, and specifically Definition 25, we study the support of ψ required to capture most of the energy of discrete signals.

3.3 Input Sampling in Phase Space Operators

Given a PSI operator T with kernel R, in this subsection we sample the input variable g of $R(g', g)$, and keep the output variable g' continuous. In Sect. 4 we show that sampling the output variable g' is a special case of the framework developed in this subsection. Let ψ be an envelope, and $g \in G$ be a random sample according to the probability distribution $\frac{\psi(g)}{\|\psi\|_1}$. Define the random rank one operator $T^{\psi,1}$, applied on $F \in L^2(G)$, by

$$g \mapsto (T^{\psi,1}F)(g) = \|\psi\|_1 R(\cdot, g) F(g).$$

We also denote $T^{\psi,1}F(g';g) = \|\psi\|_1 R(g',g)F(g)$, where g' is the variable of the output function $T^{\psi,1}F$. Next, we define the Monte Carlo approximation of TF as a sum of independent $T^{\psi,1}F$ vectors.

Definition 9 (Input Monte Carlo phase space operator) Let T be a PSI operator in $L^2(G)$ (Definition 6), ψ an envelope, $F \in L^2(G)$, and $K \in \mathbb{N}$. Let $G_k = G$, $k = 1, \ldots, K$, be K copies of G, and let $\{g^k\}_{k=1}^K$ denote a random sample from $G_1 \times \ldots \times G_K$ with the probability distribution $\prod_{k=1}^K \frac{\psi(g^k)}{\|\psi\|_1}$. Let $T_k^{\psi,1}F : G_k \to L^2(G)$ be the random vectors defined for $g^k \in G_k$ by $[T_k^{\psi,1}F](g^k) = [T^{\psi,1}F](g^k)$, $k = 1, \ldots, K$. Define the random vector $T^{\psi,K}F : G_1 \times \ldots \times G_K \to L^2(G)$ by

$$[T^{\psi,K}F](g';g^1,\ldots,g^K) := \frac{1}{K}\sum_{k=1}^K [T_k^{\psi,1}F](g';g^k) = \frac{\|\psi\|_1}{K}\sum_{k=1}^K R(g',g^k)F(g^k) \quad (23)$$

We call $T^{\psi,K}F$ the *Monte Carlo phase space integral operator* applied on F and based on K samples, approximating TF.

When the envelope ψ is fixed throughout the analysis, we often denote interchangeably $T^K F = T^{\psi,K}F$. In the following we fix an envelope ψ.

Remark 10 Note that $T^k F$ is a random variable – a function with $(\{g^k\}_{k=1}^K, g')$ as the variable. Thus, we can sample $L^2(G)$ vectors in (23), which are equivalence classes of functions, without requiring any continuity assumption on $R(\cdot,\cdot\cdot)F(\cdot\cdot)$. Indeed, $T^k F$ is defined up to a set of tuples $(\{g^k\}_{k=1}^K, g')$ of measure zero.

The expected value $\mathbb{E}(T^K F) \in L^2(G)$ of $T^K F$ is a function in $L^2(G)$, defined by

$$\mathbb{E}(T^K F) = \int_{G^K} [T^K F]\big((\cdot);g^1,\ldots,g^K\big)\frac{\psi(g^1)}{\|\psi\|_1}dg^1 \ldots \frac{\psi(g^K)}{\|\psi\|_1}dg^K.$$

We define the variance $\mathbb{V}(T^K F)$ as the integral

$$\mathbb{V}(T^K F) = \int_{G^K} \left| [T^K F]\big((\cdot);g^1,\ldots,g^K\big) - \mathbb{E}^{\mathrm{w}}(T^K F)\right|^2 \frac{\psi(g^1)}{\|\psi\|_1}dg^1 \ldots \frac{\psi(g^K)}{\|\psi\|_1}dg^K.$$

Given an envelope ψ, by abuse of notation, we also denote by ψ the multiplicative operator

$$\psi : L^2(G) \to L^2(G) \quad , \quad [\psi F](g) = \psi(g)F(g).$$

Proposition 11 *Let T be a PSI operator, and $F \in L^2(G)$. Then,*

1. the expected value $\mathbb{E}(T^K F)$ is in $L^2(G)$ and satisfies

$$\mathbb{E}(T^K F) = T(\psi F), \quad (24)$$

2. *if T is a uniformly square integrable PSI operator, with bound D, then*

$$\mathbb{V}(T^K F) = \frac{1}{K}\mathbb{V}(T^1 F) \in L^1(G), \tag{25}$$

and $\left\| \mathbb{V}(T^K F)\right\|_1 \le \frac{1}{K}\left\|\psi\right\|_1 D^2 \left\|F\right\|_2^2.$

The proof of Item 1 of Proposition 11 follows directly from the definition of PSI operators (Definition 6). Indeed, for $T^1 F$,

$$\mathbb{E}(T^1 F)(g') = \int_G \left\|\psi\right\|_1 R(g', g)F(g)\frac{\psi(g)}{\left\|\psi\right\|_1}dg = T(\psi F)(g'),$$

and for T^K we use linearity. The proof of Item 2 of Proposition 11 is in the next subsection.

We next bound the average square error in approximating $T[\psi F]$ by $T^k F$.

Proposition 12 *Let f be a continuous frame, and T a uniformly square integrable PSI operator with bound D. Then*

$$\mathbb{E}\left(\left\| T^K F - T(\psi F)\right\|_2^2 \right) \le \frac{\left\|\psi\right\|_1}{K}D^2 \left\|F\right\|_2^2.$$

Proof By the Fubini–Tonelli theorem, and Proposition 11,

$$\mathbb{E}\left(\left\| T^K F - T(\psi F)\right\|_2^2 \right)$$
$$= \int_G \int_G \left| [T^K F](g'; g^1, \ldots, g^K) - T(\psi F)(g')\right|^2$$
$$dg'\psi(g^1)/\left\|\psi\right\|_1^1 \cdots \psi(g^K)/\left\|\psi\right\|_1^K dg^1 \ldots dg^K$$
$$= \left\| \mathbb{V}(T^k F)(g')\right\|_1 \le \frac{\left\|\psi\right\|_1}{K}D^2 \left\|F\right\|_2^2.$$

$\square$

The expected error in Proposition 12 is pointwise in F. We note that an operator expected error bound of the form "$\mathbb{E}\left\| T^K - T(\psi(\cdot))\right\|_2^2 = O(\frac{\left\|\psi\right\|_1}{K})$" like in the finite dimensional matrix operator case [48] is not possible. Indeed, for any sample set $\{g^k\}_{k=1}^K$ there is a normalized function $F \in L^2(G)$ supported in $G \setminus \{g^k\}_{k=1}^K$, so $T^K F = 0$ and $\left\| T^K - T\left(\psi(\cdot)\right)\right\|_2 \ge \left\| T^K F - T(\psi F)\right\|_2 = \left\| T(\psi F)\right\|_2$. We thus focus in this paper on pointwise error estimates.

The following is an important special case of Propositions 11 and 12.

Corollary 13 *Let $T = Q_f = V_f V_f^*$ be the Gramian operator of a bounded continuous frame with $\left\| f_g\right\|_{\mathcal{H}} \le C$ and upper frame bound B. Let K_f be the frame kernel. Then, we have*

$$Q_f^K F = \frac{\left\|\psi\right\|_1}{K}\sum_{k=1}^K F(g^k)K_f(g^k, \cdot), \tag{26}$$

and

$$\mathbb{E}(Q_f^K F) = Q_f(\psi F).$$

Moreover, Q_f is a uniformly square integrable PSI operator with bound $B^{1/2}C$, and

$$\mathbb{E}\Big(\left\| Q_f^K F - Q_f(\psi F) \right\|_2^2 \Big) \le \frac{\|\psi\|_1}{K} BC^2 \|F\|_2^2.$$

Proof By Example 7, Q_f is a uniformly square integrable PSI operator with bound $B^{1/2}C$. The rest of the results follow from Propositions 11 and 12.   □

3.4 Proof of Proposition 11

The proof is based on the following lemma.

Lemma 14 *Let T be a PSI operator with kernel $R(g', g)$, let ψ be an envelope, and let $F \in L^2(G)$. Then the following holds.*

1. *The expected value of $T^1 F$ satisfies*

$$\mathbb{E}(T^1 F) = T(\psi F). \tag{27}$$

2. *If T is a uniformly square integrable PSI operator, then $\mathbb{V}(T^1 F) \in L^1(G)$, and*

$$\mathbb{V}(T^1 F) = \mathbb{E}\Big(\left| T^1 F \right|^2 \Big) - \left| \mathbb{E}(T^1 F) \right|^2. \tag{28}$$

Here, $\left| T^1 F \right|^2$ is the function $(g; g') \mapsto \left| \|\psi\|_1 R(g', g) F(g) \right|^2$, and expected value is with respect to the random variable g.

3. *If T is a uniformly square integrable PSI operator, with bound D, then*

$$\left\| \mathbb{V}(T^1 F) \right\|_1 \le \|\psi\|_1 D^2 \|F\|_2^2. \tag{29}$$

Proof Part 1 was shown in the discussion below Proposition 11. For parts 2 and 3, we write

$$\begin{aligned}
\mathbb{V}(T^1 F)(g') = & \int_G \left| [T^1 F](g'; g) \right|^2 \frac{\psi(g)}{\|\psi\|_1} dg \\
& - 2\text{Real} \int_G [T^1 F](g'; g) \overline{\mathbb{E}(T^1 F)(g')} \frac{\psi(g)}{\|\psi\|_1} dg \\
& + \int_G \left| \mathbb{E}(T^1 F)(g') \right|^2 \frac{\psi(g)}{\|\psi\|_1} dg.
\end{aligned} \tag{30}$$

We first use the Fubini–Tonelli theorem to prove integrability with respect to g' of the first term of (30). By the fact that T is uniformly square integrable (Definition 6.2),

and by $\|\psi\|_\infty \le 1$,

$$\int_G \int_G \left|[T^1 F](g'; g)\right|^2 dg' \frac{\psi(g)}{\|\psi\|_1} dg = \|\psi\|_1 \int_G \int_G |R(g', g)|^2 dg' |F(g)|^2 \psi(g) dg$$
$$\le \|\psi\|_1 D^2 \|F\|_2^2, \tag{31}$$

so $\int_G \left|[T^1 F](g'; g)\right|^2 \frac{\psi(g)}{\|\psi\|_1} dg \in L^1(G)$ with respect to g'. For the second term of (30), we have

$$\int_G \int_G [T^1 F](g'; g)\overline{\mathbb{E}(T^1 F)(g')} \frac{\psi(g)}{\|\psi\|_1} dg dg'$$
$$= \int_G \int_G R(g', g) F(g) \psi(g) dg \overline{T(\psi F)(g')} dg' = \left\|\mathbb{E}(T^1 F)\right\|_2^2.$$

This leads to (28).   By the non-negativity of the integrand in the definition of $\mathbb{V}(T^1 F)(g')$, we can write $\left\|\mathbb{V}(T^1 F)\right\|_1 = \int_G \mathbb{V}(T^1 F)(g') dg'$, so by (28) and (31), we get (29). $\square$

Proof of Proposition 11 Part 1 was shown in the discussion below Proposition 11. Next we show Part 2. By the Fubini–Tonelli theorem, we have

$$\left\|\mathbb{V}(T^K F)\right\|_1 \tag{32}$$
$$= \int_{G^K} \int_G \left|\frac{1}{K} \sum_{k=1}^K [T_k^1 F](g') - \mathbb{E}(T^K F)(g')\right|^2 dg' \|\psi\|_1^{-K} \psi(g^1) dg^1 \psi(g^K) dg^K$$

When expanding the product in (32), we have the term

$$\int_{G^K} \int_G \frac{1}{K^2} \sum_{k=1}^K \left|[T_k^1 F](g') - \mathbb{E}(T^K F)(g')\right|^2 dg' \|\psi\|_1^{-K} \psi(g^1) dg^1 \dots \psi(g^K) dg^K,$$

and mixed terms, for $k \ne k'$,

$$\int_{G^K} \int_G \frac{1}{K^2} \left([T_k^1 F](g^k; g') - \mathbb{E}(T^K F)(g')\right) \times$$
$$\overline{\left([T_{k'}^1 F](g^{k'}; g') - \mathbb{E}(T^K F)(g')\right)} dg' \|\psi\|_1^{-K} \psi(g^1) dg^1 \dots \psi(g^K) dg^K$$
$$= \int_G \int_G \overline{\left([T_k^1 F](g^{k'}; g') - \mathbb{E}(T^K F)(g')\right)} \frac{1}{K^2} \times$$
$$\int_G \left([T_k^1 F](g^k; g') - \mathbb{E}(T^K F)(g')\right) \|\psi\|_1^{-1} \psi(g^k) dg^k \, dg' \, \|\psi\|_1^{-1} \psi(g^{k'}) dg^{k'},$$

which are equal to zero, since

$$
\int_G \left([T_k^1 F](g^k; g') - \mathbb{E}(T^K F)(g') \right) \|\psi\|_1^{-1} \psi(g^k) dg^k
$$
$$
= \int_G [T_k^1 F](g^k; g') \|\psi\|_1^{-1} \psi(g^k) dg^k - \mathbb{E}(T_k^1 F)(g') = 0.
$$

Here, the fact that $\overline{\left([T_{k'}^1 F](g^{k'}; (\cdot)) - \mathbb{E}(T^K F)(\cdot) \right)} \in L^2(G)$ for a.e. $g^{k'}$, and the fact that $\|\psi\|_1^{-1} \psi(g^k) dg^k$ and $\|\psi\|_1^{-1} \psi(g^{k'}) dg^{k'}$ are probability measures, justify the above use of Fubini's theorem.

We thus have, by part 3 of Lemma 14,

$$
\left\| \mathbb{V}(T^K F) \right\|_1 = \int_{G^K} \int_G \frac{1}{K^2} \sum_{k=1}^K \left| \left([T_k^1 F](g^k; g') - \mathbb{E}(T^K F)(g') \right) \right|^2
$$
$$
dg' \|\psi\|_1^{-K} \psi(g^1) dg^1 \dots \psi(g^K) dg^K
$$
$$
= \frac{1}{K^2} \sum_{k=1}^K \int_{g^k} \int_{g'} \left| \left([T_k^1 F](g^k; g') - \mathbb{E}(T^K F)(g') \right) \right|^2 dg' \|\psi\|_1^{-1} \psi(g^k) dg^k
$$
$$
= \frac{1}{K^2} \sum_{k=1}^K \mathbb{V}(T_k^1 F) = \frac{1}{K} \mathbb{V}(T_k^1 F) \le \frac{1}{K} \|\psi\|_1 D^2 \|F\|_2^2.
$$

$\square$

3.5 Monte Carlo Synthesis

In this subsection, we use the results of Sect. 3.3 to define and analyze the Monte Carlo approximation of synthesis.

Definition 15 Let ψ be an envelope and $\{g^k\}_{k=1}^K$ random samples as in Definition 9. Given $F \in L^2(G)$, the *Monte Carlo synthesis* $V_f^{*\psi,K} F$ is the random variable $G^K \to \mathcal{H}$ defined as

$$
V_f^{*\psi,K} F = \frac{\|\psi\|_1}{K} \sum_{k=1}^K F(g^k) f_{g^k}.
$$

When the envelope ψ is constant in the analysis, we often denote the Monte Carlo synthesis in short by V_f^{*K}. The following proposition formulates the Monte Carlo synthesis using the Monte Carlo PSI operator Q_f^K approximating the Gramian operator Q_f (Corollary 13), and the frame operator S_f.

Proposition 16 $V_f^{*K} F = S_f^{-1} V_f^* Q_f^K F.$

Proof By linearity, it is enough to prove for $K = 1$. By the fact that $S_f = V_f^* V_f$,

$$
S_f^{-1} V_f^* Q_f^1 F = S_f^{-1} V_f^* \|\psi\|_1 F(g) K(g, \cdot)
$$
$$
= \|\psi\|_1 F(g) S_f^{-1} V_f^* V_f(f_g) = \|\psi\|_1 F(g) f_g = V_f^{*1} F.
$$

$\square$

Next, we show that $V_f^{*K} F$ approximates $V_f^*[\psi F]$.

Proposition 17 *(Synthesis Monte Carlo approximation rate)* *Let f be a bounded continuous frame with frame bounds A, B, and $\|f_g\|_{\mathcal{H}} \leq C$. Let $\psi \in L^1(G)$ be an envelope. Then*

$$
\mathbb{E}\Big(\left\| V_f^{*K} F - V_f^*[\psi F] \right\|_{\mathcal{H}}^2 \Big) \leq \frac{\|\psi\|_1}{K} \frac{B}{A} C^2 \|F\|_2^2.
$$

Proof By Proposition 16 and Lemma 35 of Appendix B

$$
\begin{aligned}
\left\| V_f^{*K} F - V_f^*[\psi F] \right\|_{\mathcal{H}} &= \left\| S_f^{-1} V_f^* Q_f^K F - S_f^{-1} V_f^* Q_f[\psi F] \right\|_{\mathcal{H}} \\
&= \left\| V_f^+ \Big(Q_f^K F - Q_f[\psi F] \Big) \right\|_{\mathcal{H}} \\
&\leq \left\| V_f^+ \right\| \left\| Q_f^K F - Q_f[\psi F] \right\|_2 \leq A^{-1/2} \left\| Q_f^K F - Q_f[\psi F] \right\|_2.
\end{aligned}
\tag{33}
$$

Indeed, by the frame bound $A^{1/2} \|s\|_{\mathcal{H}} \leq \left\| V_f[s] \right\|_2$ for $s = V_f^+ F$,

$$
\left\| V_f^+ F \right\|_{\mathcal{H}} \leq A^{-1/2} \left\| V_f V_f^+ F \right\|_2 = A^{-1/2} \left\| P_{V_f(\mathcal{H})} F \right\|_2 \leq A^{-1/2} \|F\|_2.
$$

Now, the result follows from Corollary 13.

$\square$

4 Stochastic Phase Space Signal Processing of Continuous Signals

In this subsection we formulate and analyze the Monte Carlo approximations (8)–(11) of the signal processing in phase space pipelines (5) and (6). In Sect. 4.2 we bound the expected value of the error, and in Sect. 4.3 we bound the concentration of measure of the error.

4.1 Definition of Stochastic Phase Space Signal Processing

For Parseval frames (Definition 1.6), the frame operator is $S_f = I$, and hence signal processing in phase space takes the form $\mathcal{P}_{f,T,r} := V_f^* Tr \circ V_f$. We call $\mathcal{P}_{f,T,r}$ *Parseval signal processing in phase space* even if f is non-Parseval. For non-Parseval frames, synthesis and analysis signal processing in phase space involve the multiplications $\mathcal{P}_{f,T,r} S_f^{-1}$ and $S_f^{-1} \mathcal{P}_{f,T,r}$ respectively. Since $\left\| S_f^{-1} \right\|_2 \leq A^{-1} < \infty$, it is enough

to bound the error entailed by randomly approximating $\mathcal{P}_{f,T,r}$, and then multiply the bound by $\left\| S_f^{-1} \right\|_2$ for non-Parseval pipelines. We hence focus only on Parseval signal processing in our analysis.

As discussed in Sect. 3.2, sampling in phase space requires enveloping. We hence formulate the following list of signal processing pipelines.

Definition 18 Let f be a continuous frame over the phase space G, T be a bounded operator in $L^2(G)$, $r : \mathbb{C} \to \mathbb{C}$, and $\psi, \eta \in L^1(G)$ two envelopes. Let $s \in \mathcal{H}$ denote a generic signal.

1. A *signal processing pipeline* is defined by

$$\mathcal{P}_{f,T,r}s = V_f^* T (r \circ V_f[s]). \tag{34}$$

2. An *output enveloped signal processing pipeline* is defined by

$$\mathcal{P}_{f,T,r}^{\psi}s = V_f^* \psi T (r \circ V_f[s]). \tag{35}$$

3. An *input-output enveloped signal processing pipeline* is defined by

$$\mathcal{P}_{f,T,r}^{\psi;\eta}s = V_f^* \psi T \left(\eta r \circ V_f[s] \right). \tag{36}$$

The following list of Monte Carlo approximations correspond to the pipelines of Definition 18.

Definition 19 Let $\mathcal{P}_{f,T,r}$ be a signal processing pipeline, ψ and η two envelopes, and $K, L \in \mathbb{N}$. Let $s \in \mathcal{H}$ denote a generic signal.

1. The *output stochastic signal processing pipeline* is defined by

$$[\mathcal{P}s]_{f,T,r}^{\psi,K} = V_f^{*\psi,K} \left(T(r \circ V_f[s]) \right) \tag{37}$$

2. For a phase space integral operator T, the *input-output stochastic signal processing pipeline* is defined by

$$[\mathcal{P}s]_{f,T,r}^{\psi,K;\eta,L} = V_f^{*\psi,K} \left(T^{\eta,L}(r \circ V_f[s]) \right) \tag{38}$$

We typically fix f, T, and r, in which case we omit them from the pipeline notation and denote $\mathcal{P}$, $\mathcal{P}^{\psi}$, $\mathcal{P}^{\psi,K}$ etc. Equations (8)–(11) give explicit formulas for the synthesis and analysis formulations of $[\mathcal{P}s]^{\psi,K;\eta,L}$ and $[\mathcal{P}s]^{\psi,K}$, based on the samples in phase space. As noted in Sect. 1.2, the pipeline (37) is useful for multipliers, shrinkage, and phase vocoder, and the pipeline (38) is useful for PSI operators.

4.2 Expected Error in Stochastic Phase Space Signal Processing

In the following, we estimate the error of the stochastic methods.

Theorem 20 *Let f be a bounded continuous frame with bounds A, B, and $\left\| f_g \right\|_{\mathcal{H}} \leq C$. Let T be a bounded operator in $L^2(G)$, and $r : \mathbb{C} \to \mathbb{C}$ satisfy $|r(x)| \leq E\,|x|$ for some $E \geq 0$ and every $x \in \mathbb{C}$. Let ψ and η be two envelopes, and $K, L \in \mathbb{N}$. Then, for every signal $s \in \mathcal{H}$, the following two properties hold.*

1. *Output enveloped signal processing stochastic approximation:*

$$\mathbb{E}\Big(\left\| [\mathcal{P}s]^{\psi,K} - [\mathcal{P}s]^{\psi} \right\|_{\mathcal{H}}^2 \Big) \leq \frac{\|\psi\|_1}{K} A^{-1} B^2 C^2 E^2 \, \|T\|_2^2 \, \|s\|_{\mathcal{H}}^2 .$$

2. *Input-output enveloped signal processing stochastic approximation: if T is a uniformly bounded PSI operator with bound D, then*

$$
\begin{aligned}
\mathbb{E}&\Big(\left\| [\mathcal{P}s]^{\psi,K;\eta,L} - [\mathcal{P}s]^{\psi;\eta} \right\|_{\mathcal{H}}^2 \Big) \\
&\leq 4 \frac{\|\eta\|_1}{L} D^2 B^2 E^2 \, \|s\|_{\mathcal{H}}^2 + 16 \frac{\|\psi\|_1 \|\eta\|_1}{KL} A^{-1} B^2 C^2 D^2 E^2 \, \|s\|_{\mathcal{H}}^2 \\
&\quad + 16 \frac{\|\psi\|_1}{K} A^{-1} B^2 C^2 E^2 \, \|T\|_2^2 \, \|s\|_{\mathcal{H}}^2 \\
&= O\left(\frac{\|\eta\|_1}{L} + \frac{\|\psi\|_1}{K} + \frac{\|\psi\|_1 \|\eta\|_1}{KL} \right) \|s\|_{\mathcal{H}}^2 .
\end{aligned}
$$

We use the following simple observation to prove Theorem 20.

Lemma 21 *Let Z_1, Z_2, Z_3 be non-negative real-valued random variables such that $Z_1 \leq Z_2 + Z_3$ pointwise in the sample set. Then,*

$$\mathbb{E}(Z_1^2) \leq 4\mathbb{E}(Z_2^2) + 4\mathbb{E}(Z_3^2). \tag{39}$$

Proof We have $Z_1 \leq 2\max\{Z_2, Z_3\}$, where the maximum is pointwise in the sample space. Therefore, $Z_1^2 \leq 4\max\{Z_2^2, Z_3^2\} \leq 4Z_2^2 + 4Z_3^2$, and (39) follows.   □

Proof of Theorem 20 We prove 2, and note that 1 is simpler and uses similar techniques. Denote by $\mathbf{g} = \{g^1, \ldots, g^K\}$ the output samples underlying $V_f^{*\psi,K}$, and by $\mathbf{y} = \{y^1, \ldots, y^L\}$ the input samples underlying $T^{\eta,L}$. Denote $F = r(V_f[s]) \in L^2(G)$. By the triangle inequality, and by the fact that $\|V_f\| = \left\| V_f^* \right\| \leq B^{1/2}$ and $0 \leq \psi(g) \leq 1$,

$$
\begin{aligned}
&\left\| V_f^{*\psi,K} T^{\eta,L} F - V_f^* \psi T[\eta F] \right\|_{\mathcal{H}} \\
&\leq \left\| V_f^{*\psi,K} T^{\eta,L} F - V_f^* \psi T^{\eta,L} F \right\|_{\mathcal{H}} + B^{1/2} \left\| T^{\eta,L} F - T[\eta F] \right\|_2 .
\end{aligned}
\tag{40}
$$

When calculating the conditional expected value of $\left\| V_f^{*\psi,K} T^{\eta,L} F - V_f^* \psi T[\eta F] \right\|_{\mathcal{H}}^2$, with respect to a fixed $\mathbf{g}$ (denoted here by $\mathbb{E}(\cdot \,|\mathbf{g})$), we use Lemma 21 and Proposition 12

to get

$$\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T[\eta F]\Big\|_{\mathcal{H}}^2 \Big|\mathbf{g}\Big)$$
$$\leq 4\frac{\|\eta\|_1}{L}D^2 B\,\|F\|_2^2 + 4\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T^{\eta,L}F\Big\|_{\mathcal{H}}^2 \Big|\mathbf{g}\Big).$$

Thus

$$\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T[\eta F]\Big\|_{\mathcal{H}}^2\Big)$$
$$\leq 4\frac{\|\eta\|_1}{L}D^2 B\,\|F\|_2^2 + 4\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T^{\eta,L}F\Big\|_{\mathcal{H}}^2\Big).$$

Note that Fubini–Tonelli theorem is satisfied in the computation of $\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T^{\eta,L}F\Big\|_{\mathcal{H}}^2\Big)$ as a repeated integral of $\mathbf{y}$ and $\mathbf{g}$, since the integrand is positive and the measure is σ-finite.

Next, by Proposition 17,

$$\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T^{\eta,L}F\Big\|_{\mathcal{H}}^2 \Big|\mathbf{y}\Big) \leq \frac{\|\psi\|_1}{K}A^{-1}BC^2\Big\|T^{\eta,L}F\Big\|_2^2.$$

Now,

$$\Big\|T^{\eta,L}F\Big\|_2 \leq \Big\|T^{\eta,L}F - T[\eta F]\Big\|_2 + \|T[\eta F]\|_2\,,$$

so by Lemma 21, by Proposition 12, and by the fact that $0 \leq \eta(y) \leq 1$,

$$\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T^{\eta,L}F\Big\|_{\mathcal{H}}^2\Big)$$
$$\leq \frac{\|\psi\|_1}{K}A^{-1}BC^2\Big(4\frac{\|\eta\|_1}{L}D^2\,\|F\|_2^2 + 4\,\|T\|_2^2\,\|F\|_2^2\Big).$$

Altogether,

$$\mathbb{E}\Big(\Big\|V_f^{*\psi,K}T^{\eta,L}F - V_f^*\psi T[\eta F]\Big\|_{\mathcal{H}}^2\Big)$$
$$\leq 4\frac{\|\eta\|_1}{L}D^2 B\,\|F\|_2^2 + 16\frac{\|\psi\|_1}{K}A^{-1}BC^2\frac{\|\eta\|_1}{L}D^2\,\|F\|_2^2$$
$$+ 16\frac{\|\psi\|_1}{K}A^{-1}BC^2\,\|T\|_2^2\,\|F\|_2^2\,.$$

The claim now follows from $\|F\|_2^2 = \big\|r \circ V_f[s]\big\|_2^2 \leq E^2 B\,\|s\|_{\mathcal{H}}^2$.   □

4.3 Concentration of Error in Stochastic Phase Space Signal Processing

Propositions 12 and 17, and Theorem 20 estimate the average square error of the stochastic approximations. In this subsection, we formulate the approximation results as bounds on the error that hold in high probability. We show how to apply the classical concentration of measure estimates, Markov's inequality, and Bernstein's inequality, in our setting.

4.3.1 A Bernstein Inequality in Hilbert Spaces

In the following version of Bernstein's inequality, we define expected values of weakly integrable random vectors v over the sample set G using the weak integral (Definition 3) as $\mathbb{E}^{\mathrm{w}}(v) = \int_G^{\mathrm{w}} v(g) d\mu(g)$. The following version of Bernstein's inequality is a direct result of [8, Theorem 2.6], and is proved in Appendix A.

Theorem 22 *(Hilbert space Bernstein's inequality) Let $\mathcal{H}$ be a separable Hilbert space, and G a probability space. Let $\{v_k\}_{k=1}^K : G^K \to \mathcal{H}^K$ be a finite sequence of independent random weakly integrable vectors. Suppose that for every $k = 1, \dots, K$, $\mathbb{E}^{\mathrm{w}}(v_k) = 0$ and $\|v_k\|_{\mathcal{H}} \leq B$ a.s. and assume that $\rho_K^2 > \sum_{k=1}^K \mathbb{E} \|v_k\|_{\mathcal{H}}^2$ for some constant $\rho_K \in \mathbb{R}$. Then, for every $0 \leq t \leq \rho_K^2 / B$,*

$$P\left(\left\|\sum_{k=1}^K v_k\right\|_{\mathcal{H}} \geq t\right) \leq \exp\left(-\frac{t^2}{8\rho_K^2} + \frac{1}{4}\right). \tag{41}$$

We note that existing variants of Bernstein's inequality in infinite dimensional Hilbert spaces are not adequate for us. For example, the operator Bernstein's inequality of [38] is limited to trace class operators, and thus does not even include the identity.

4.3.2 Concentration of Error Results

Markov type concentration of error results can be derived from Propositions 12 and 17, and Theorem 20, by multiplying the error bound by δ^{-1}, and replacing the expected value with an event that has probability at least $(1 - \delta)$. For example, the Markov type concentration of error version of Proposition 17 reads

$$\left\|V_f^{*K} F - V_f^*[\psi F]\right\|_{\mathcal{H}}^2 \leq \frac{\|\psi\|_1}{K} \frac{B}{A} C^2 \|F\|_2^2 \delta^{-1}$$

in probability more that $(1 - \delta)$.

The following proposition summarizes the Markov and Bernstein types concentration of error bounds in output stochastic signal processing.

Theorem 23 *(Output signal processing concentration of error) Let f be a bounded continuous frame with frame bounds A and B, and with $\|f_g\|_{\mathcal{H}} \leq C$, let ψ be an envelope, and $K \in \mathbb{N}$. Let T be a bounded operator in $L^2(G)$, and $r : \mathbb{C} \to \mathbb{C}$ satisfy*

$|r(x)| \leq E\,|r(x)|$ *for every* $x \in \mathbb{C}$, *where* $E > 0$. *Let* $s \in \mathcal{H}$ *and* $0 < \delta < 1$. *Then, with probability more than* $1 - \delta$, *we have*

$$\left\| [\mathcal{P}s]^{\psi,K} - [\mathcal{P}s]^{\psi} \right\|_{\mathcal{H}} \leq \frac{\sqrt{\|\psi\|_1}}{\sqrt{K}} A^{-1/2} BCE \|T\|_2 \|s\|_{\mathcal{H}} \kappa(\delta), \qquad (42)$$

where $\kappa(\delta)$ *can be chosen as one of the following two options.*

1. Markov type error bound: $\kappa(\delta) = \delta^{-\frac{1}{2}}$.
2. Bernstein type error bound: $\kappa(\delta) = 2\sqrt{2}\sqrt{\ln\left(\frac{1}{\delta}\right) + \frac{1}{4}}$ *in case* $\|T\|_\infty < \infty$ *and* K
   satisfies

$$K \geq \|\psi\|_1 \left(\frac{C}{B^{1/2}} \frac{\|T\|_\infty}{\|T\|_2} + \frac{B^{1/2}}{C\|\psi\|_1} \right)^2 \kappa(\delta)^2. \qquad (43)$$

Proof We prove 2 and note that 1 is simpler and based on Markov's inequality. Denote $F = Tr(V_f[s])$. Below, we use the following bounds

$$\|F\|_\infty = \left\| Tr(V_f[s]) \right\|_\infty \leq \|T\|_\infty\, EC\, \|s\|_{\mathcal{H}}, \qquad (44)$$

$$\|F\|_2 = \left\| Tr(V_f[s]) \right\|_2 \leq \|T\|_2\, EB^{1/2}\, \|s\|_{\mathcal{H}}. \qquad (45)$$

and

$$B^{1/2}C\,\|F\|_\infty + \frac{1}{\|\psi\|_1}B\,\|F\|_2 \leq J, \qquad (46)$$

where

$$J = B^{1/2}C^2 E\,\|T\|_\infty\,\|s\|_{\mathcal{H}} + \frac{1}{\|\psi\|_1}B^{1.5}E\,\|T\|_2\,\|s\|_{\mathcal{H}}.$$

We use Theorem 22 as follows. Define the independent random vectors

$$v_k : G^k \to L^2(G), \quad v_k(\mathbf{g}) = \frac{1}{K}\left(Q_f^1(g^k)F - Q_f\psi F\right), \quad k = 1, \ldots, K$$

where the sample set is $\{G^k ; \prod_{k=1}^{K} \frac{\psi(g^k)}{\|\psi\|_1} dg^k\}$. By Corollary 13, $\mathbb{E}^{w}(v_k) = \mathbb{E}(v_k) = 0$, and $\mathbb{E}(\|v_k\|_2^2) \leq \frac{\|\psi\|_1}{K^2}BC^2\|F\|_2^2$. Therefore, by (45),

$$\sum_{k=1}^{K} \mathbb{E}(\|v_k\|_2^2) \leq \frac{\|\psi\|_1}{K}BC^2\|F\|_2^2 \leq \frac{\|\psi\|_1}{K}B^2C^2E^2\|T\|_2^2\|s\|_{\mathcal{H}}^2.$$

Moreover, by Proposition 5, Example 7, and (46), for every $g^k \in G$

$$\|v_k\|_2 \leq \frac{1}{K} \left(\left\| \|\psi\|_1 K(g^k, \cdot) F(g^k) \right\|_2 + \left\| Q_f \psi F \right\|_2 \right)$$
$$\leq \frac{\|\psi\|_1}{K} \left(B^{1/2} C \|F\|_\infty + \frac{1}{\|\psi\|_1} B \|F\|_2 \right) \leq \frac{\|\psi\|_1}{K} J.$$

Hence, by Theorem 22, for every $0 \leq t \leq B^2 C^2 E^2 \|T\|_2^2 \|s\|_{\mathcal{H}}^2 / J$

$$P\left(\left\| Q_f^K F - Q_f(\psi F) \right\|_2 \geq t \right) \leq \exp\left(-\frac{t^2}{8 B^2 C^2 E^2 \|T\|_2^2 \|s\|_{\mathcal{H}}^2} \frac{K}{\|\psi\|_1} + \frac{1}{4} \right). \qquad (47)$$

Now, set

$$\delta = \exp\left(-\frac{t^2}{8 B^2 C^2 E^2 \|T\|_2^2 \|s\|_{\mathcal{H}}^2} \frac{K}{\|\psi\|_1} + \frac{1}{4} \right),$$

or equivalently

$$t = \sqrt{8} \sqrt{-\ln(\delta) + \frac{1}{4}} BCE \|T\|_2 \|s\|_{\mathcal{H}} \frac{\sqrt{\|\psi\|_1}}{\sqrt{K}},$$

and demand $0 \leq t \leq B^2 C^2 E^2 \|T\|_2^2 \|s\|_{\mathcal{H}}^2 / J$, namely,

$$\sqrt{8} \sqrt{-\ln(\delta) + \frac{1}{4}} BCE \|T\|_2 \|s\|_{\mathcal{H}} \frac{\sqrt{\|\psi\|_1}}{\sqrt{K}} \leq B^2 C^2 E^2 \|T\|_2^2 \|s\|_{\mathcal{H}}^2 / J.$$

This gives, in probability at least $(1 - \delta)$,

$$\left\| Q_f^K F - Q_f \psi F \right\|_{\mathcal{H}} \leq BCE \|T\|_2 \|s\|_{\mathcal{H}} \frac{\sqrt{\|\psi\|_1}}{\sqrt{K}} \kappa(\delta),$$

whenever k satisfies (43). Last, using Proposition 16 and (33), we get

$$\left\| [\mathcal{P}s]^{\psi, K} - [\mathcal{P}s]^\psi \right\|_{\mathcal{H}} = \left\| S_f^{-1} V_f^* Q_f^K F - S_f^{-1} V_f^* Q_f \psi F \right\|_{\mathcal{H}}$$
$$\leq A^{-1/2} \left\| Q_f^K F - Q_f \psi F \right\|_{\mathcal{H}}$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

5 Stochastic Phase Space Signal Processing of Discrete Signals

In previous sections we showed how to randomly discretize phase space. In Theorems 20 and 23, when the number of samples satisfy $K, L = Z \max\{\|\psi\|_1, \|\eta\|_1\}$, for

$Z > 0$, the approximation errors are of order $O(Z^{-1/2})$. In this section, we additionally discretize the signal space $\mathcal{H}$ to a finite dimensional subspace $V_M \subset \mathcal{H}$ of dimension/resolution $M \in \mathbb{N}$. The main goal is to relate the choices of $\|\psi\|_1$ and $\|\eta\|_1$ to the resolution M. We introduce a class of frames, called *linear volume discretizable (LVD) frames*, for which there are envelopes ψ_M and η_M with $\|\psi_M\|_1, \|\eta_M\|_1 = O(M)$ that contain most of the energy of $V_f[s_M]$ for every $s_M \in V_M$. Thus, a stochastic signal processing method for LVD frames requires $K, L = ZM$ samples, with $Z > 0$, for the approximation error to be $O(\frac{1}{\sqrt{Z}})$.

5.1 Discrete Signals and Linear Volume Discretization of Continuous Frames

We treat discrete signals as embedded in the Hilbert space of signals $\mathcal{H}$. A discrete signal is an element of a finite dimensional subspace of $\mathcal{H}$. On the one hand, we can analyze discrete signals directly in $\mathcal{H}$. On the other hand, discrete signals are determined by a finite number of scalars, so they are well adapted to numerical analysis. In our analysis, we sometimes restrict ourselves to a class of signals $\mathcal{R} \subset \mathcal{H}$ which need not be a linear space. We typically consider $\mathcal{R}$ defined by imposing a restriction on signals in $\mathcal{H}$ which is natural for real life signals of some type.

Definition 24 Let $\mathcal{H}$ be a Hilbert space that we call the *signal space*. A *class of signals* $\mathcal{R} \subset \mathcal{H}$ is a (possibly non-linear) subset of $\mathcal{H}$. A *sequence of discretizations* of $\mathcal{R}$ is a sequence of (generally non-linear) subspaces $\{V_M \subset \mathcal{H}\}_{M=1}^\infty$ that satisfies the following condition: for every $s \in \mathcal{R}$ there is a sequence $\{s_M \in V_M\}_{M=1}^\infty$ such that

$$\lim_{M \to \infty} \|s_M - s\|_{\mathcal{H}} = 0.$$

The *resolution* $\dim(V_M)$ of V_M is defined to be the dimension of $\mathrm{span}\, V_M$.

The idea in discretizing a continuous frame is to find an envelope ψ_M for each discrete space V_M such that for any $s_M \in V_M$, the approximation error of $V_f[s_M]$ by $\psi_M V_f[s_M]$ is controlled. The envelopes ψ_M are interpreted as covering domains $G_M \subset G$ in which most of the energy of functions from $V_f[V_M]$ resides.

Definition 25 Let $f : G \to \mathcal{H}$ be a continuous frame. Let $\mathcal{R} \subset \mathcal{H}$ be a class of signals, and $\{V_M\}_{m=1}^\infty$ a discretization of $\mathcal{R}$.

1. The continuous frame f is called *linear volume discretizable (LVD)* with respect to the class $\mathcal{R}$ and the discretization $\{V_M\}_{M=1}^\infty$, if for every error tolerance $\epsilon > 0$ there is a constant $C_\epsilon > 0$ and $M_0 \in \mathbb{N}$, such that for any $M \geq M_0$ there is an envelope ψ_M with

$$\|\psi_M\|_1 \leq C_\epsilon \dim(V_M) \tag{48}$$

such that for any $s_M \in V_M$,

$$\frac{\left\| V_f[s_M] - \psi_M V_f[s_M] \right\|_2}{\left\| V_f[s_M] \right\|_2} < \epsilon. \tag{49}$$

2. For a linear volume discretizable continuous frame f with respect to $\mathcal{R}$ and $\{V_M\}_{M=1}^{\infty}$, and a fixed tolerance $\epsilon > 0$ with a corresponding fixed C_ϵ and envelope sequence $\{\psi_M\}_{M=1}^{\infty}$ satisfying (48) and (49), we call f together with $\mathcal{R}$, $\{V_M\}_{m=1}^{\infty}$, and $\{\psi_M\}_{M=1}^{\infty}$, an ϵ-*linear volume discretization (ϵ-LVD) of* f.

5.2 Error in Discrete Stochastic Phase Space Signal Processing

Next, we study the error in discrete stochastic phase space signal processing of LVD frames. Since the energy of $V_f[s_M]$ may be shifted after applying an operator T on $V_f[s_M]$, we first introduce the following definition.

Definition 26 Let G be a phase space, T a bounded linear operator in $L^2(G)$, ψ and η two envelopes, and $\epsilon > 0$. We say that T *maps the energy of* η *to* ψ *up to* ϵ, if

$$\|T\eta - \psi T\eta\|_{2\to 2} \le \epsilon. \tag{50}$$

The next theorem summarizes the expected approximation error in stochastic signal processing with ϵ-LVD frames.

Theorem 27 *Let f be a bounded continuous frame with bound $\|f_g\|_{\mathcal{H}} \le C$. Let $r : \mathbb{C} \to \mathbb{C}$ satisfy $|r(x)| \le E|x|$, where $E > 0$. Suppose that f together with the signal class $\mathcal{R}$, the discretization $\{V_M\}_{m=1}^{\infty}$, and the envelopes $\{\eta_M\}_{M=1}^{\infty}$, is an ϵ-LVD of f, with constant C_ϵ. Let $\{\psi_M\}_{M=1}^{\infty}$ be a sequence of envelopes satisfying*

$$\|\psi_M\|_1 \le C_\epsilon \dim(V_M). \tag{51}$$

Let T be a bounded operator on $L^2(G)$ that maps the energy of η_M to ψ_M up to ϵ. Then, the following two bounds are satisfied for every $s_M \in V_M$.

1.

$$\frac{\mathbb{E}\left(\left\| [\mathcal{P}s_M]_{f,T,r}^{\psi_M,K} - [\mathcal{P}s_M]_{f,T,r} \right\|_{\mathcal{H}}^2 \right)}{\|s_M\|_{\mathcal{H}}^2}$$
$$\le 4\frac{C_\epsilon \dim(V_M)}{K} A^{-1} B^2 C^2 E^2 \|T\|_2^2 + 4B^2 E^2 (1 + 2\|T\|_2)^2 \epsilon^2 \tag{52}$$

2. If T is a uniformly square integrable PSI operator with bound D, then

$$\frac{\mathbb{E}\left(\left\| [\mathcal{P}s_M]_{f,T,r}^{\psi_M,K;\eta_M,L} - [\mathcal{P}s_M]_{f,T,r} \right\|_{\mathcal{H}}^2 \right)}{\|s_M\|_{\mathcal{H}}^2}$$
$$\le 16\frac{C_\epsilon \dim(V_M)}{L} D^2 B^2 E^2 + 64\frac{C_\epsilon^2 \dim(V_M)^2}{KL} A^{-1} B^2 C^2 D^2 E^2$$
$$+ 64\frac{C_\epsilon \dim(V_M)}{K} A^{-1} B^2 C^2 E^2 \|T\|_2^2 + 4B^2 E^2 (1 + \|T\|_2^2) \epsilon^2. \tag{53}$$

Proof We first prove (53). By Lemma 21,

$$
\mathbb{E}\left(\left\| [\mathcal{P}s_M]^{\psi_M,K;\eta_M,L} - [\mathcal{P}s_M]_{f,T,r} \right\|_{\mathcal{H}}^2 \right)
$$
$$
\leq 4\mathbb{E}\left(\left\| [\mathcal{P}s_M]^{\psi_M,K;\eta_M,L} - [\mathcal{P}s_M]^{\psi_M;\eta_M} \right\|_{\mathcal{H}}^2 \right) + 4 \left\| [\mathcal{P}s_M]^{\psi_M;\eta_M} - [\mathcal{P}s_M] \right\|_{\mathcal{H}}^2. \quad (54)
$$

Next, we bound the second term of (54).

$$
\left\| [\mathcal{P}s_M]^{\psi_M;\eta_M} - [\mathcal{P}s_M] \right\|_{\mathcal{H}} = \left\| V_f^* \psi_M T \left(\eta_M r \circ V_f[s_M] \right) - V_f^* T \left(r \circ V_f[s_M] \right) \right\|_{\mathcal{H}}
$$
$$
\leq B^{1/2} \left\| \psi_M T \left(\eta_M r \circ V_f[s_M] \right) - T \left(r \circ V_f[s_M] \right) \right\|_{\mathcal{H}}
$$
$$
\leq B^{1/2} \left\| T \eta_M r \circ V_f[s_M] - \psi_M T \eta_M r \circ V_f[s_M] \right\|_2
$$
$$
+ B^{1/2} \left\| T r \circ V_f[s_M] - T \eta_M r \circ V_f[s_M] \right\|_2 \quad (55)
$$
$$
\leq B^{1/2} \left\| T \eta_M - \psi_M T \eta_M \right\|_2 \left\| r \circ V_f[s_M] \right\|_2
$$
$$
+ B^{1/2} \left\| T \right\|_2 \left\| r \circ V_f[s_M] - \eta_M r \circ V_f[s_M] \right\|_2
$$
$$
\leq B \epsilon E \left\| s_M \right\|_{\mathcal{H}} + B^{1/2} \left\| T \right\|_2 \left\| (1 - \eta_M) r \circ V_f[s_M] \right\|_2.
$$

For the second term of the last line of (55), by the LVD property,

$$
\left\| (1 - \eta_M) r \circ V_f[s_M] \right\|_2 \leq E \left\| \eta_M V_f[s_M] - V_f[s_M] \right\|_2 \leq \epsilon E B^{1/2} \left\| s_M \right\|_{\mathcal{H}}. \quad (56)
$$

To conclude, (54) together with (55), (56), and Theorem 20, give (53).

Next, we prove (52). As before,

$$
\mathbb{E}\left(\left\| [\mathcal{P}s_M]_{f,T,r}^{\psi_M,K} - [\mathcal{P}s_M]_{f,T,r} \right\|_{\mathcal{H}}^2 \right)
$$
$$
\leq 4\mathbb{E}\left(\left\| [\mathcal{P}s_M]^{\psi_M,K} - [\mathcal{P}s_M]^{\psi_M} \right\|_{\mathcal{H}}^2 \right) + 4 \left\| [\mathcal{P}s_M]^{\psi_M} - [\mathcal{P}s_M] \right\|_{\mathcal{H}}^2. \quad (57)
$$

We bound the second term of (57) using (50) and (56) by

$$
\left\| [\mathcal{P}s_M]^{\psi_M} - [\mathcal{P}s_M] \right\|_{\mathcal{H}} = \left\| V_f^* \psi_M T r \circ V_f[s_M] - V_f^* T r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
\leq B^{1/2} \left\| T r \circ V_f[s_M] - T \eta_M r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
+ B^{1/2} \left\| T \eta_M r \circ V_f[s_M] - \psi_M T \eta_M r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
+ B^{1/2} \left\| \psi_M T \eta_M r \circ V_f[s_M] - \psi_M T r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
\leq B^{1/2} \left\| T \right\|_2 \left\| r \circ V_f[s_M] - \eta_M r \circ V_f[s_M] \right\|_{\mathcal{H}} \quad (58)
$$
$$
+ B^{1/2} \left\| T \eta_M - \psi_M T \eta_M \right\|_{\mathcal{H}} \left\| r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
+ B^{1/2} \left\| T \right\|_2 \left\| \eta_M r \circ V_f[s_M] - r \circ V_f[s_M] \right\|_{\mathcal{H}}
$$
$$
\leq B^{1/2} \left\| T \right\|_2 \epsilon E B^{1/2} \left\| s_M \right\|_{\mathcal{H}} + B^{1/2} \epsilon E B^{1/2} \left\| s_M \right\|_{\mathcal{H}} + B^{1/2} \left\| T \right\|_2 \epsilon E B^{1/2} \left\| s_M \right\|_{\mathcal{H}}
$$
$$
= B E (2 \left\| T \right\|_2 + 1) \left\| s_M \right\|_{\mathcal{H}}.
$$

This, together with (57) and Theorem 20, leads to (52). □

Next, we formulate concentration of error results for LVD frames. A Markov type concentration of error result can be derived directly from Theorem 27. For a Bernstein type error bound, we offer the following theorem only for the output stochastic signal processing pipeline (Definition 19.1).

Theorem 28 *Consider the setting of Theorem 27.1, and suppose that $\|T\|_\infty < \infty$. Let $\delta > 0$, $\kappa(\delta) = 2\sqrt{2}\sqrt{\ln\left(\frac{1}{\delta}\right) + \frac{1}{4}}$, and K satisfy (43). Then, in probability more than $(1 - \delta)$,*

$$
\frac{\left\|[\mathcal{P}s_M]_{f,T,r}^{\psi_M,K} - [\mathcal{P}s_M]_{f,T,r}\right\|_{\mathcal{H}}}{\|s_M\|_{\mathcal{H}}} \tag{59}
$$
$$
\leq \frac{\sqrt{C_\epsilon \dim(V_M)}}{\sqrt{K}} A^{-1/2} BCE \|T\|_2 \kappa(\delta) + BE(2\|T\|_2 + 1)\epsilon.
$$

Proof The proof follows from (58) and Theorem 23.2, similarly to the proof of Theorem 27.1. □

5.3 Discrete Stochastic Time–Frequency Signal Processing

In this subsection we present a discretization under which the STFT is LVD. In the companion paper [33] we present a discretization under which the CWT is linear volume discretizable. We analyze time signals $s : \mathbb{R} \to \mathbb{C}$ by decomposing them to compact time interval sections. Without loss of generality, we suppose that each signal segment is supported in $[-1/2, 1/2]$. Focusing on one segment, we take the signal class $\mathcal{R}$ as $L^2[-1/2, 1/2]$. Let V_M be the space of trigonometric polynomials of order M (namely, finite Fourier series expansions). In the frequency domain, signals $q \in V_M$ are represented by

$$
\hat{q}(z) = \sum_{n=-M}^{M} c_n \mathrm{sinc}(z - n)
$$

where c_n are the Fourier coefficients of q, and sinc is the Fourier transform of the indicator function of $[-1/2, 1/2]$. Consider a window function f supported at the time interval $[-S, S]$ that satisfies the following. There exist constants $C', Y > 0$, and $\kappa > 1/2$, such that for every $z > Y$ or $z < -Y$

$$
\hat{f}(z) \leq C' |z|^{-\kappa}. \tag{60}
$$

Let $W > 0$. For each $M \in \mathbb{N}$, we consider the following phase space domain $G_M \subset G$, where G is the STFT time-frequency plane,

$$
G_M = \left\{(x, \omega) \,\middle|\, -WM < \omega < WM, \ |x| < 1/2 + S\right\}. \tag{61}
$$

The area of G_M in the time-frequency plane is

$$\mu(G_M) = 2WM(1 + 2S). \tag{62}$$

Denote

$$\psi_M(g) = \begin{cases} 1 \,, & g \in G_M \\ 0 \,, & g \notin G_M. \end{cases} \tag{63}$$

Theorem 29 *Under the above setting, the STFT is LVD with respect to the class* $L^2[-1/2, 1/2]$ *and the discretization* $\{V_M\}_{M \in \mathbb{N}}$, *with the envelopes* ψ_M *defined by (63) for large enough* W *that depends only on* ϵ *of Definition 25.*

Proof Let $W > 1$. A direct calculation of the STFT shows

$$\int_{\mathbb{R}} \left| V_f[q](\omega, x) \right|^2 dx = \int_{\mathbb{R}} \left| \hat{q}(z) \right|^2 \left| \hat{f}(z - \omega) \right|^2 dz. \tag{64}$$

We consider $\omega > 0$ and $z > 0$, and note that the other cases are similar. For each value of $\omega > MW$, we decompose the integral (64) along z into the two integrals in $z \in (0, (M + \omega)/2)$ and $z \in ((M + \omega)/2, \infty)$. For $z \in (0, (M + \omega)/2)$, since $\omega \geq MW$ and $z \leq (M + \omega)/2$,

$$z - \omega \leq M - \omega/2 \leq -M(W/2 - 1) < 0,$$

so $|z - \omega|^{-2\kappa}$ obtains its maximum at $z = (M + \omega)/2$. Thus, by (60),

$$\int_0^{(M+\omega)/2} \left| \hat{q}(z) \right|^2 \left| \hat{f}(z - \omega) \right|^2 dz \leq \|q\|_2^2 \max_{0 \leq z \leq (M+\omega)/2} C'^2 |z - \omega|^{-2\kappa}$$

$$= \|q\|_2^2 \, C'^2 \, |(M + \omega)/2 - \omega|^{-2\kappa} = \|q\|_2^2 \, C'^2 \, |(M - \omega)/2|^{-2\kappa} \tag{65}$$

Integrating the bound (65) for $\omega \in (WM, \infty)$ gives

$$\int_{WM}^{\infty} \int_0^{(M+\omega)/2} \left| \hat{q}(z) \right|^2 \left| \hat{f}(z - \omega) \right|^2 dz d\omega = (W - 1)^{-2k+1} M^{-2k+1} \|q\|_2^2 \, O(1)$$

$$= o_W(1) o_M(1) \|q\|_2^2. \tag{66}$$

For $z \in ((M + \omega)/2, \infty)$, $\hat{q}$ decays like $M^{1/2}(z - M)^{-1}$. Indeed, since $z > M$

$$\sum_{n=-M}^{M} c_n \operatorname{sinc}(z - n) \leq \|\{c_n\}\|_2 \sqrt{\sum_{n=-M}^{M} \frac{1}{(z - n)^2}} \leq \|q\|_2 \sqrt{\sum_{n=-M}^{M} \frac{1}{(z - M)^2}}$$

$$\leq 2 \|q\|_2 \sqrt{M}(z - M)^{-1}. \tag{67}$$

Now, by (64) and (67),

$$\int_{(M+\omega)/2}^{\infty} |\hat{q}(z)|^2 \left|\hat{f}(z-\omega)\right|^2 dz \leq 2 \|f\|_2^2 \|q\|_2^2 \max_{(M+\omega)/2 \leq z < \infty} M(z-M)^{-2}$$
$$= \|f\|_2^2 \|q\|_2^2 M\big((\omega-M)/2\big)^{-2}. \tag{68}$$

Integrating the bound (68) for $\omega \in (WM, \infty)$ gives

$$\int_{WM}^{\infty} \int_{(M+\omega)/2}^{\infty} |\hat{q}(z)|^2 \left|\hat{f}(z-\omega)\right|^2 dz d\omega = (W-1)^{-1} \|q\|_2^2 O(1). \tag{69}$$

Last, the bounds (66) and (69) are combined to give $\left\|(I - \psi_M)V_f[q]\right\|_2 = o_W(1) \|q\|_2$, so by the frame inequality

$$\frac{\left\|(I - \psi_M)V_f[q]\right\|_2}{\left\|V_f[q]\right\|_2} = o_W(1).$$

This means that given $\epsilon > 0$, we may choose W large enough to guarantee $\frac{\|(I-\psi_M)V_f[q]\|_2}{\|V_f[q]\|_2} < \epsilon$, and also guarantee that for every $M \in \mathbb{N}$, $\|\psi_M\|_1 \leq C_\epsilon M$, with $C_\epsilon = 2W(1+2S)$ by (62). □

6 Applications of Stochastic Signal Processing of Continuous Signals

In this section, we introduce two applications of the theory developed in this paper: integration of continuous frames and stochastic phase space diffeomorphism.

6.1 Integration of Linear Volume Discretizable Frames

Here, we show how to integrate a set of LVD continuous frames into one continuous LVD frame, while retaining all stochastic approximation bounds of a single LVD frame. We first show how to integrate frames.

Proposition 30 *Let G and U be two topological spaces with σ-finite Borel measures μ_G and μ_U respectively, with $\mu_U(U) = 1$. Let A, B, $C > 0$ and $\mathcal{H}$ be a Hilbert space. For each $u \in U$, let $f_{\cdot, u} : g \mapsto f_{g,u}$ be a bounded continuous frame over the phase space G and the signal space $\mathcal{H}$, with frame constants A, B and bound $\left\|f_{g,u}\right\|_{\mathcal{H}} \leq C$. Suppose that the mapping $f : (g, u) \mapsto f_{g,u}$ is continuous. Then f is a bounded continuous frame over the phase space $G \times U$, with frame constants A, B and bound $\left\|f_{g,u}\right\|_{\mathcal{H}} \leq C$.*

Proof Consider the mapping V_f that maps $s \in \mathcal{H}$ to the function

$$V_f[s] : (g, u) \mapsto \langle s, f_{g,u} \rangle.$$

By continuity of $(g, u) \mapsto f_{g,u}$, $V_f[s]$ is continuous for every $s \in \mathcal{H}$. Indeed

$$\left| V_f[s](g, u) - V_f[s](g', u') \right| = \left| \langle s, f_{g,u} \rangle - \langle s, f_{g',u'} \rangle \right| \leq \|s\|_{\mathcal{H}} \|f_{g,u} - f_{g',u'}\|_{\mathcal{H}}.$$

Thus, for every $s \in \mathcal{H}$, $V_f[s] : G \times U \to \mathbb{C}$ is a measurable function. For each $u \in U$, denote by $V_{f,u}$ the analysis operator corresponding to the continuous frame $f_{\cdot,u}$. By Fubini-Tonelli theorem, for every signal $s \in \mathcal{H}$

$$\|V_f[s]\|_2^2 = \iint_{G \times U} |\langle s, f_{g,u} \rangle|^2 \, d(g, u) = \int_U \int_G |\langle s, f_{g,u} \rangle|^2 \, dg \, du = \int_U \|V_{f,u}[s]\|_2^2 \, du.$$

Therefore,

$$A \|s\|_{\mathcal{H}}^2 = \int_U A \|s\|_{\mathcal{H}}^2 \, du \leq \|V_f[s]\|_2^2 \leq \int_U B \|s\|_{\mathcal{H}}^2 \, du = B \|s\|_{\mathcal{H}}^2,$$

and $\{f_{g,u}\}_{(g,u) \in G \times U}$ is a continuous frame with frame bounds A, B. $\qquad \square$

Next, we show that the LVD property is retained under integration of frames.

Proposition 31 *Consider the setting of Proposition 30. Let $\mathcal{R} \subset \mathcal{H}$ be a signal class, and V_M a discretization of $\mathcal{R}$. Suppose that for every $u \in U$, $f_{\cdot,u}$ is an LVD frame. Let $\epsilon > 0$ and $\{\psi_M \in L^1(G)\}_{M=1}^{\infty}$ a sequence of envelopes. Suppose that for every $u \in U$, $f_{\cdot,u}$ is ϵ-LVD with respect to the envelopes $\{\psi_M \in L^1(G)\}$ and the constant C_ϵ. For each M, denote by $\psi_M \in L^1(G \times U)$ the envelope $(g, u) \mapsto \psi_M(g)$. Then, f is an ϵ-LVD frame with respect to the envelopes $\{\psi_M \in L^1(G \times U)\}$ and the bound C_ϵ.*

Proof By Fubini-Tonelli theorem

$$\left\| V_f[s_M] - \psi_M V_f[s_M] \right\|_2^2 = \int_U \left\| V_{f,u}[s_M] - \psi_M V_{f,u}[s_M] \right\|_2^2 \, du \leq \epsilon^2 \left\| V_f[s] \right\|_2^2. \tag{70}$$

$\qquad \square$

Last, we show how to integrate operators in phase space, and show that mapping the energy between envelopes up to ϵ (Definition 26) is preserved under integration.

Proposition 32 *Consider the setting of Proposition 30. Let $\mathcal{R} \subset \mathcal{H}$ be a signal class, and V_M a discretization of $\mathcal{R}$. Suppose that for every $u \in U$, $f_{\cdot,u}$ is an LVD frame. Let $\epsilon > 0$ and $\{\eta_M \in L^1(G)\}_{M=1}^{\infty}$ and $\{\psi_M \in L^1(G)\}$ two sequences of envelopes. Suppose that for every $u \in U$, $f_{\cdot,u}$ is ϵ-LVD with respect to the envelopes $\{\eta_M \in L^1(G)\}$ and the constant C_ϵ. For each $u \in U$, let T_u be a bounded operator in $L^2(G)$ with $\|T_u\|_{L^2(G)} \leq C_T$. Suppose that for every $M \in \mathbb{N}$ and a.e. $u \in U$, T_u maps the energy of η_M to ψ_M up to ϵ. Let T be the operator in $L^2(G \times U)$ defined for $F \in L^2(G \times U)$ by*

$$T F(g, u) = T_u F(\cdot, u)(g).$$

Then T is bounded with $\|T\|_{L^2(G\times U)} \leq C_T$, and maps the energy of $\eta_M \in L^1(G\times U)$ to $\psi_M \in L^1(G\times U)$ up to ϵ.

Under the assumptions of Proposition 32, f and T satisfy the conditions of Theorems 27 and 28. This means that the number of random samples in the stochastic method in f, required for a given accuracy, is comparable to the number of samples required for each $\{f_{\cdot,u}\}_{u\in U}$. Namely, the addition of the new feature direction U to the phase space G does not entail an increase in computational complexity, and the approximation of the continuous method by the discrete Monte Carlo method is of order $O(\frac{\sqrt{\dim(V_M)}}{\sqrt{K}})$.

The above procedure of integrating continuous frames can be carried out when the definition of a certain continuous frame depends on some free parameters u. For example, in the STFT and the CWT, the window function and mother wavelet are free parameters. Instead of fixing the window function, we may consider a parametric space of window functions, parameterized by u, sharing the same linear volume discretization, and add the parameter u as additional dimensions to phase space. For example, in the CWT we may choose as u the spread of the mother wavelet. Integration of frames is the basis on which we construct the LTFT in Definition 33 below.

6.2 Stochastic Diffeomorphism Operator and Highly Redundant Phase Vocoder

In this subsection, we study the signal processing pipeline when T is a diffeomorphism operator, and propose a potential application in audio signal processing.

6.2.1 Stochastic Signal Processing with Diffeomorphism Operators

Let $f : G \to \mathcal{H}$ be a bounded continuous frame, with bound $\|f_g\|_{\mathcal{H}} \leq C$, and suppose that the phase space G is a Riemannian manifold. Let $\mathcal{R} \subset \mathcal{H}$ be a class of signal, $\{V_M\}_M$ a discretization of $\mathcal{R}$, and $\{\eta_M\}_M$ a sequence of envelopes. Let $\epsilon > 0$, and suppose that $\{f, \mathcal{R}, \{V_M\}_M, \{\eta_M\}_M\}$ is an ϵ-LVD of f.

Let $d : G \to G$ be a diffeomorphism (invertible smooth mapping with smooth inverse), with Jacobian $J_d \in L^\infty(G)$. Consider the diffeomorphism operator T, defined for any $F \in L^2(G)$ by

$$[TF](g) = F(d^{-1}(g)). \tag{71}$$

Note that $\|T\|_2 = \|J_d\|_\infty$. Let $r : \mathbb{C} \to \mathbb{C}$ satisfy $|r(x)| \leq E\,|x|$. The signal processing pipeline based on the diffeomorphism T is defined to be $\mathcal{P}_{f,T,r} S_f^{-1} s$ for the synthesis pipeline, and $S_f^{-1} \mathcal{P}_{f,T,r} s$ for the analysis pipeline. The stochastic approximations of these pipelines are given on $s_M \in V_M$, up to the application of S_f^{-1} from the right or from the left, by

$$[\mathcal{P}s_M]^{\eta_M,K} = \frac{\|\eta_M\|_1}{K} \sum_{k=1}^{K} r\big(V_f[s](g^k)\big) f_{d(g^k)}. \tag{72}$$

In (72), the points $\{g^k\}_{k=1}^K$ are sampled from the envelope η_M. This means that the points $\{d(g^k)\}_{k=1}^K$ are sampled from the envelope $\psi_M(g) = \eta_M\big(d(g)\big)J_d(g)$ with $\|\psi_M\|_1 = \big\|\eta_M\big(d(\cdot)\big)J_d(\cdot)\big\|_1 = \|\eta_M\|_1$. We can use either Theorems 27 or 28 to bound the stochastic approximation error, and in either case we obtain an error of order $O(\frac{\sqrt{\dim(V_M)}}{\sqrt{K}})$.

6.2.2 Integer Time Dilation Phase Vocoder

A time stretching phase vocoder is an audio effect that slows down an audio signal without dilating its frequency content. In the classical definition, G is the time frequency plane, and V_f is the STFT. Phase vocoder can be formulated as phase space signal processing in case the signal is dilated by an integer [52, Section 7.4.3]. For an integer Δ, we consider the diffeomorphism operator T with $d(g_1, g_2) = (\Delta g_1, g_2)$, and consider the nonlinearity r, defined by

$$r(e^{i\theta}a) = e^{i\Delta\theta}a, \tag{73}$$

for $a \in \mathbb{R}_+$ and $\theta \in \mathbb{R}$. The phase vocoder is defined to be $s \mapsto V_f^* T r \circ V_f[s]$. Note that since the STFT is a Parseval frame, there is no difference between analysis and synthesis signal processing.

Next, we replace the STFT frame in the phase vocoder with a highly redundant time-frequency representation, based on a 3D phase space.

6.2.3 The Localizing Time–Frequency Transform

Here, we construct an example redundant time frequency transform based on a combination of CWT atoms and STFT atoms. The CWT is better than the STFT at isolating transient high frequency events, since middle to high frequency wavelet atoms have shorter time supports than LTFT atoms. On the other hand, low frequency events are smeared by the CWT, since low frequency wavelet atoms have large supports. We thus use STFT atoms to represent low frequencies, and CWT atoms to represent middle frequencies. High frequencies are represented again by STFT atoms with narrow time supports. This is done to potentially avoid false positive detection of very short transient events by very short wavelet atoms.

We then add to this 2D time-frequency system a third axis that controls the number of oscillations in the CWT atoms. We motivate this as follows. Time-frequency atoms are subject to the uncertainty principle. The more accurately a time-frequency atom measures frequency, the less accurately it measures time. Different signal features call for a different balance between the time and the frequency measurement accuracy. In polyphonic audio signals we expect a range of such appropriate balances, which means that no choice of window is appropriate for all features. Hence, the addition of the *number of oscillations* axis may be useful for representing a variety of features in polyphonic audio signals.

Consider a non-negative real valued window $h(t)$ supported in $[-1/2, 1/2]$. For example, the *Hann window* is defined to be $h(t) = \big(1 + cos(2\pi t)\big)/2$, and zero for

$t \notin [-1/2, 1/2]$. Consider a parameter τ that controls the *number of oscillations* in the CWT atoms. We denote by $0 < \tau_1 < \tau_2$ the minimal and maximal number of oscillations of the wavelet atoms. The LTFT phase space is defined to be $G = \mathbb{R}^2 \times [\tau_1, \tau_2]$, where the measure μ_3 on $[\tau_1, \tau_2]$ is any weighted Lebesgue measure with $\mu_3([\tau_1, \tau_2]) = 1$. There are two transition frequencies in the LTFT, where the atoms change from STFT to CWT atoms and back. In general, we allow these transition frequencies $0 < a_\tau < b_\tau < \infty$ to depend on τ.

Definition 33 (The localizing time-frequency continuous frame) Consider the above setting. The LTFT atoms are defined for $(x, \omega, \tau) \in \mathbb{R}^2 \times [\tau_1, \tau_2]$, where x represents time, ω frequencies, and τ the number of wavelet oscillations, by

$$f_{x,\omega,\tau}(t) = \begin{cases} \sqrt{\frac{a_\tau}{\tau}} h\left(\frac{a_\tau}{\tau}(t-x)\right) e^{2\pi i \omega(t-x)} & \text{if} \quad |\omega| < a_\tau \\ \sqrt{\frac{\omega}{\tau}} h\left(\frac{\omega}{\tau}(t-x)\right) e^{2\pi i \omega(t-x)} & \text{if } a_\tau \leq |\omega| \leq b_\tau \\ \sqrt{\frac{b}{\tau}} h\left(\frac{b_\tau}{\tau}(t-x)\right) e^{2\pi i \omega(t-x)} & \text{if} \quad b_\tau < |\omega| \end{cases} \qquad (74)$$

In the companion paper [33] we prove that the LTFT is an LVD continuous frame. This is natural in view of Sect. 6.1, since, up to the low and high frequency truncation, the LTFT is based on integrating LVD wavelet transforms.

6.2.4 LTFT-Based Phase Vocoder

In this subsection we offer a toy example application of the LTFT, namely, integer time stretching phase vocoder based on the implementation of Sect. 6.2.2. The integer time dilation phase modification r of (73) is said to preserve the *horizontal phase coherence*, and deals well with slowly varying instantaneous frequencies [39]. When the instantaneous frequency of a component in a signal changes rapidly, the phase modification r is not sufficient, and methods for "locking" the phase to a frequency bin outside the horizontal line are used in modern implementations [31,32]. Nevertheless, in this toy application we consider horizontal phase lock. One potential motivation for using this simplistic model is that phasiness artifacts[1] may be alleviated by using CWT atoms, since CWT atoms have shorter time supports than STFT atoms.

An example implementation of stochastic LTFT phase vocoder (72) is given in https://github.com/RonLevie/LTFT-Phase-Vocoder. In the companion paper [33], we prove that the total number of scalar operations in the phase vocoder LTFT method is $O(ZM + M \log(M))$, when the number of Monte Carlo samples is $K = ZM$. In future work, we will study more modern phase vocoder implementations based on the LTFT, akin to [16,34,41,45].

---

[1] Phasiness [31] is the audible artifact resulting from summing two time-frequency atoms with intersecting time and frequency supports, but with out of sync phases

A Hilbert space Berntein inequality

The proof of Theorem 22 is based on the finite dimensional counterpart from [8, Theorem 2.6]. There, the theorem is formulated for vectors in $\mathbb{R}^n$, but can easily be extended to $\mathbb{C}^n$.

Theorem 34 (*Finite dimensional Bernstein inequality [8]*) *Let* $\{v_k\}_{k=1}^K \subset \mathbb{C}^d$ *be a finite sequence of independent random vectors. Suppose that* $\mathbb{E}(v_k) = 0$ *and* $\|v_k\|_2 \leq B$ *a.s. and assume that* $\rho_K^2 > \sum_{k=1}^K \mathbb{E}\|v_k\|_2^2$. *Then for all* $0 \leq t \leq \rho_K^2/B$,

$$P\left(\left\|\sum_{k=1}^K v_k\right\|_2 \geq t\right) \leq \exp\left(-\frac{t^2}{8\rho_K^2} + \frac{1}{4}\right).$$

In the following proof of Theorem 22, we use the fact that for weakly integrable random vectors $v : G \to \mathcal{H}$ and bounded operators T on $\mathcal{H}$, we have

$$\int_G^w Tv(g)dg = T\int_G^w v(g)dg. \tag{75}$$

Proof of Theorem 22 For a fixed K we denote $\rho = \rho_K$. Let $\{w_l\}_{l=1}^\infty$ be an orthonormal basis in $\mathcal{H}$, and P_j be the orthogonal projection upon span$\{w_l\}_{l=1}^j$. Let us use Theorem 34 on the random vectors $\{v_k^j\}_{k=1}^K = \{P_j v_k\}_{k=1}^K$, as vectors of $\mathbb{C}^{d_j}$, for fixed j. By (75), we have

$$\mathbb{E}^w(v_k^j) = \mathbb{E}^w(P_j v_k) = P_j \mathbb{E}^w(v_k) = \mathbb{E}(v_k) = 0.$$

Next, by the fact that P_j is a projection

$$\left\|v_k^j\right\|_2 = \left\|P_j v_k\right\|_{\mathcal{H}} \leq \|v_k\|_{\mathcal{H}} \leq B.$$

Now, the pointwise bound $\left\|P_j v_k\right\|_{\mathcal{H}}^2 \leq \|v_k\|_{\mathcal{H}}^2$ carries to the integrals in the calculation of the expected values, so

$$\sum_{k=1}^K \mathbb{E}\left\|v_k^j\right\|_2^2 = \sum_{k=1}^K \mathbb{E}\left\|P_j v_k\right\|_{\mathcal{H}}^2 \leq \sum_{k=1}^K \mathbb{E}\|v_k\|_{\mathcal{H}}^2 \leq \rho^2.$$

Thus, Theorem 34 gives

$$\forall\, 0 \le t \le \rho^2/B \, . \quad P\left(\left\| \sum_{k=1}^{K} v_k^j \right\|_2 \ge t \right) \le \exp\left(-\frac{t^2}{8\rho^2} + \frac{1}{4} \right). \tag{76}$$

Next, we show that (76) carries also in the limit as $j \to \infty$. Consider the following functions in the probability space G^K: the characteristic function χ of the set

$$\left\{ \mathbf{g} \mid \left\| \sum_{k=1}^{K} v_k(\mathbf{g}) \right\|_{\mathcal{H}} > t \right\},$$

and characteristic function χ_j of

$$\left\{ \mathbf{g} \mid \left\| \sum_{k=1}^{K} v_k^j(\mathbf{g}) \right\|_2 > t \right\}.$$

By the fact that projections reduce norms, $\chi_j(\mathbf{g}) \le \chi(\mathbf{g})$ for every $\mathbf{g} \in G_0^K$. Moreover, χ_j is a pointwise monotone sequence of measurable functions. By the strong convergence of the projections P_j to I, we have

$$\forall \mathbf{g} \in G_0^K \, . \quad \lim_{j\to\infty} \chi_j(\mathbf{g}) = \chi(\mathbf{g}).$$

This is shown as follows. Let $\mathbf{g} = (g_1, \ldots, g_K)$ be a fixed point. If $\chi(\mathbf{g}) = 0$ then it is trivial to see $\lim_{j\to\infty} \chi_j(\mathbf{g}) = \chi(\mathbf{g})$. Otherwise, for every $\epsilon > 0$ there is a big enough $J_\epsilon \in \mathbb{N}$ such that for every $j > J_\epsilon$ we have

$$\left| \left\| \sum_{k=1}^{K} v_k(g_k) \right\|_{\mathcal{H}} - \left\| \sum_{k=1}^{K} v_k^j(g_k) \right\|_2 \right| < \epsilon.$$

Since $\chi(\mathbf{g}) = 1$, we have

$$\left\| \sum_{k=1}^{K} v_k(g_k) \right\|_{\mathcal{H}} = r > t.$$

Therefore, for $\epsilon < 0.5(r - t)$, and any $j > J_\epsilon$

$$t < r - \epsilon < \left\| \sum_{k=1}^{K} v_k^j \right\|_2 < r + \epsilon$$

so $\chi_j(\mathbf{g}) = 1$, which proves that $\lim_{j\to\infty} \chi_j(\mathbf{g}) = \chi(\mathbf{g})$. Lastly, (41) follows from Beppo Levi's monotone convergence theorem. $\qquad \square$

B Pseudo Inverse of Frame Analysis and Synthesis Operators

For an injective linear operator with close range $B : \mathcal{V} \to \mathcal{W}$ between the Hilbert spaces $\mathcal{V}$ and $\mathcal{W}$, we define the pseudo inverse

$$B^+ : \mathcal{W} \to \mathcal{V}, \quad B^+ = \left(B\big|_{\mathcal{V} \to B\mathcal{V}}\right)^{-1} R_{B\mathcal{V}},$$

where $R_{B\mathcal{V}} : \mathcal{W} \to B\mathcal{V}$ is the surjective operator given by the orthogonal projection from $\mathcal{W}$ to $B\mathcal{V}$ and restriction of the image space to the range $B\mathcal{V}$, and $B\big|_{\mathcal{V} \to B\mathcal{V}}$ is the restriction of the image space of B to its range $B\mathcal{V}$, in which it is invertible. Note that $R_{B\mathcal{V}}^*$ is the operator that takes a vector from $B\mathcal{V}$ and canonically embeds it in $\mathcal{W}$, and $P_{B\mathcal{V}} = R_{B\mathcal{V}}^* R_{B\mathcal{V}} : \mathcal{W} \to \mathcal{W}$ is the orthogonal projection upon $B\mathcal{V}$. In the following we list basic properties of V_f and V_f^+.

Lemma 35 *Let $f : G \to \mathcal{H}$ be a continuous frame. Then the following properties hold.*

1. $V_f^+ V_f = I$.
2. $V_f V_f^+ = P_{V_f[\mathcal{H}]}$.
3. $V_f^+ P_{V_f[\mathcal{H}]} = V_f^+$.
4. *The operator V_f^* is the pseudo inverse of V_f^{+*}, and $V_f^{+*}[\mathcal{H}] = V_f[\mathcal{H}]$.*
5. $(V_f^* V_f)^{-1} = V_f^+ V_f^{+*}$.

References

1. Ali, S., Antoine, J., Gazeau, J.: Continuous frames in Hilbert space. Ann. Phys. **222**(1), 1–37 (1993)
2. Balazs, P.: Basic definition and properties of Bessel multipliers. J. Math. Anal. Appl. **325**(1), 571–585 (2007)
3. Balazs, P., Bayer, D., Rahimi, A.: Multipliers for continuous frames in Hilbert spaces. J. Phys. A **45**(24), 244023 (2012)
4. Balazs, P., Laback, B., Eckel, G., Deutsch, W.A.: Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking. Trans. Audio Speech Lang. Proc. **18**(1), 34–49 (2010). https://doi.org/10.1109/TASL.2009.2023164
5. Bass, R.F., Gröchenig, K.: Relevant sampling of band-limited functions. Ill. J. Math. **57**(1), 43–58 (2013). https://doi.org/10.1215/ijm/1403534485
6. Caflisch, R.E.: Monte Carlo and quasi-Monte Carlo methods. Acta Numer. **7**, 1–49 (1998). https://doi.org/10.1017/S0962492900002804
7. Candes, E., Donoho, D.: Continuous curvelet transform: I. resolution of the wavefront set. Appl. Comput. Harmon. Anal. **19**(2), 162–197 (2005)
8. Candes, E.J., Plan, Y.: A probabilistic and RIPless theory of compressed sensing. IEEE Trans. Inf. Theory **57**(11), 7235–7254 (2011). https://doi.org/10.1109/TIT.2011.2161794
9. Candès, E.J., Donoho, D.L.: Recovering edges in ill-posed inverse problems: optimality of curvelet frames. Ann. Stat. **30**(3), 784–842 (2002). https://doi.org/10.1214/aos/1028674842
10. Clarkson, K.L., Woodruff, D.P.: Low-rank PSD approximation in input-sparsity time. In: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 2061–2072. https://doi.org/10.1137/1.9781611974782.134
11. Colonna, F., Easley, G., Guo, K., Labate, D.: Radon transform inversion using the shearlet representation. Appl. Comput. Harmon. Anal. **29**(2), 232–250 (2010). https://doi.org/10.1016/j.acha.2009.10.005

12. Crochiere, R.: A weighted overlap-add method of short-time fourier analysis/synthesis. IEEE Trans. Acoust. Speech Signal Process. **28**(1), 99–102 (1980). https://doi.org/10.1109/TASSP.1980.1163353
13. Daubechies, I.: Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, SIAM (1992)
14. Donoho, D., Johnstone, J.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994). https://doi.org/10.1093/biomet/81.3.425
15. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D.: Wavelet shrinkage: Asymptopia? J. R. Stat. Soc. Ser. B **57**(2), 301–369 (1995)
16. Driedger, J., Müller, M.: A review of time-scale modification of music signals. Appl. Sci. **12**(2), 57 (2016)
17. Drineas, P., Kannan, R., Mahoney, M.: Fast Monte Carlo algorithms for matrices I: approximating matrix multiplication. SIAM J. Comput. **36**(1), 132–157 (2006). https://doi.org/10.1137/S0097539704442684
18. Drineas, P., Mahoney, M.W.: On the Nyström method for approximating a Gram matrix for improved kernel-based learning. J. Mach. Learn. Res. **6**, 2153–2175 (2005)
19. Duflo, M., Moore, C.: On the regular representation of a nonunimodular locally compact group. J. Funct. Anal. **21**, 209–243 (1976)
20. Easley, G.R., Colonna, F., Labate, D.: Improved radon based imaging using the shearlet transform. In: Szu, H.H., Agee, F.J. (eds.) Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VII, vol. 7343, pp. 58–67. International Society for Optics and Photonics, SPIE (2009)
21. Easley, G.R., Labate, D.: Image Processing Using Shearlets, pp. 283–325. Birkhäuser Boston, Boston (2012). https://doi.org/10.1007/978-0-8176-8316-0_8
22. Elad, M., Starck, J., Querre, P., Donoho, D.: Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). Appl. Comput. Harmon. Anal. **19**, 340–358 (2005)
23. Führ, H.: Abstract Harmonic Analysis of Continuous Wavelet Transforms. Springer, Berlin (2005)
24. Führ, H., Xian, J.: Relevant sampling in finitely generated shift-invariant spaces. J. Approx. Theory **240**, 1–15 (2019). https://doi.org/10.1016/j.jat.2018.09.009
25. Gröchenig, K.: Foundations of Time-Frequency Analysis. Birkhäuser Basel (2001)
26. Grossmann, A., Morlet, J.: Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. **15**(4), 723–736 (1984)
27. Grossmann, A., Morlet, J., Paul, T.: Transforms associated with square integrable group representations I. general results. J. Math. Phys. **26**(10), 2473–2479 (1985)
28. Guo, K., Kutyniok, G., Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators. In: International Conference on the Interaction between Wavelets and Splinesl **Wavelets and Splines** (2005)
29. Guo, Q., Yu, S., Chen, X., Liu, C., Wei, W.: Shearlet-based image denoising using bivariate shrinkage with intra-band and opposite orientation dependencies. In: 2009 International Joint Conference on Computational Sciences and Optimization **1**, 863–866 (2009). https://doi.org/10.1109/CSO.2009.218
30. Kutyniok, G.: Geometric separation by single-pass alternating thresholding. Appl. Comput. Harmon. Anal. **36**(1), 23–50 (2014)
31. Laroche, J., Dolson, M.: Phase-vocoder: about this phasiness business. In: Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, p. 4 (1997)
32. Laroche, J., Dolson, M.: Improved phase vocoder time-scale modification of audio. IEEE Trans. Speech Audio Process. **7**(3), 323–332 (1999)
33. Levie, R., Avron, H.: Randomized continuous frames in time-frequency analysis. arXiv:2009.10525 [math.NA] (2020)
34. Liuni, M., Roebel, A.: Phase vocoder and beyond. Music Technol. **7**, 73–89 (2013)
35. Mahoney, M.W.: Randomized algorithms for matrices and data. Found. Trends Mach. Learn. **3**(2), 123–224 (2011). https://doi.org/10.1561/2200000035
36. Majdak, P., Balázs, P., Kreuzer, W., Dörfler, M.: A time-frequency method for increasing the signal-to-noise ratio in system identification with exponential sweeps. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 3812–3815 (2011)
37. Matz, G., Hlawatsch, F.: Time-frequency transfer function calculus of linear time-varying systems. In: Boashas, B. (ed.) Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, pp. 135–144. Elsevier, Oxford, UK (2003)

38. Minsker, S.: On some extensions of Bernstein's inequality for self-adjoint operators. Stat. Probab. Lett. **127**, 111–119 (2017)
39. Moulines, E., Laroche, J.: Non-parametric techniques for pitch-scale and time-scale modification of speech. Speech Commun. **16**(2), 175–205 (1995). https://doi.org/10.1016/0167-6393(94)00054-E
40. Olivero, A., Torrésani, B., Kronland-Martinet, R.: A class of algorithms for time-frequency multiplier estimation. IEEE Trans. Audio Speech Lang. Process. **21**, 1550–1559 (2013)
41. Ottosen, E.S., Dörfler, M.: A phase vocoder based on nonstationary Gabor frames. IEEE/ACM Trans. Audio Speech Lang. Process. **25**, 2199–2208 (2017)
42. Patel, D., Sampath, S.: Random sampling in reproducing kernel subspaces of lp(rn). J. Math. Anal. Appl. **491**(1), 124270 (2020). https://doi.org/10.1016/j.jmaa.2020.124270
43. Pettis, B.J.: On integration in vector spaces. Trans. Am. Math. Soc. **44**(2), 277–304 (1938)
44. Portnoff, M.: Implementation of the digital phase vocoder using the fast Fourier transform. IEEE Trans. Acoust. Speech Signal Process. **24**(3), 243–248 (1976). https://doi.org/10.1109/TASSP.1976.1162810
45. Průša, Z., Holighaus, N.: Phase vocoder done right. In: Proceedings of 25th European Signal Processing Conference (EUSIPCO-2017), pp. 1006–1010. Kos (2017)
46. Rahimi, A., Najati, A., Dehghan, Y.N.: Continuous frame in Hilbert spaces. Methods Funct. Anal. Topol. **12**(2), 170–182 (2006)
47. Stoeva, D., Balazs, P.: Invertibility of multipliers. Appl. Comput. Harmon. Anal. **33**(2), 292–299 (2012)
48. Tropp, J.A.: An introduction to matrix concentration inequalities. Found. Trends Mach. Learn. **8**(1–2), 1–230 (2015)
49. Velasco, G.A.: Relevant sampling of the short-time fourier transform of time-frequency localized functions. arXiv:1707.09634 [math.FA] (2017)
50. Woodruff, D.P.: Sketching as a tool for numerical linear algebra. Found. Trends Theor. Comput. Sci. **10**(12), 1–157 (2014). https://doi.org/10.1561/0400000060
51. Yang, J., Meng, X., Mahoney, M.W.: Implementing randomized matrix algorithms in parallel and distributed environments. Proc. IEEE **104**(1), 58–92 (2016). https://doi.org/10.1109/JPROC.2015.2494219
52. Zolzer, U.: DAFX: Digital Audio Effects, 2nd edn. Wiley, New York (2011)