Multiscale Chirplets and Near-Optimal Recovery of Chirps

Emmanuel J. Candès
Applied and Computational Mathematics
California Institute of Technology
Pasadena, California 91125

Abstract

This paper considers the model problem of recovering a signal $f(t)$ from noisy sampled measurements. The objects we wish to recover are chirps which are neither smoothly varying nor stationary but rather, which exhibit rapid oscillations and rapid changes in their frequency content.

We introduce a mathematical model to describe classes of chirps of the general form $f(t) = A(t)\cos(\lambda\varphi(t))$ where λ is a (large) base frequency, $\varphi(t)$ is time-varying and $A(t)$ is slowly varying, by imposing some smoothness conditions on the amplitude $A(t)$ and the "instantaneous frequency" $\varphi'(t)$. For example, our models allow the unknown object to oscillate at nearly the sampling/Nyquist rate.

Building on recent advances in computational harmonic analysis, we construct libraries of tight frames of *multiscale chirplets* which are rapidly searchable and with fast algorithms for analysis and synthesis. We show that it is possible to invoke low-complexity algorithms which select a best tight-frame from our library in which simple thresholding achieves nearly minimax mean-squared errors over our classes of chirps. Our methodology is adaptive in the sense that it does not require a-priori knowledge of the degree of smoothness of the amplitude and the instantaneous frequency, and nearly attains the minimax risk over a meaningful range of chirp classes.

Keywords. Minimax Estimation, Chirps, Recursive Partitioning, Time-Frequency Analysis, Local Cosines, Adaptive Estimation, Oracle Inequalities.

Acknowledgments. This research was supported by a National Science Foundation grant DMS-0140540 and by an Alfred P. Sloan Fellowship. The author would like to thank Elisabeth Gassiat and the Department of Statistics at the Université of Paris-Sud for their hospitality. He also acknowledges fruitful conversations with Pascal Massart, David Donoho, and McKay Hyde.

1 Introduction

We consider the problem of removing noise from one-dimensional signals. We suppose that we have noisy sampled data

$$y_i = f_i + z_i, \quad i = 0, 1, \dots, N-1; \tag{1.1}$$

where the z_i's are i.i.d. $N(0, \sigma^2)$, and the unknown coordinates (f_i) are sampled values $f_i = f(i/N)$ of a signal of interest $f(t)$, $t \in [0,1]$. We wish to recover f with small mean-squared error

$$MSE(\hat{f}, f) = E\left(\frac{1}{N}\sum_i (\hat{f}_i - f_i)^2\right). \tag{1.2}$$

This is of course a well-known problem in the literature of statistics and signal processing; classical results would assume that the objects we wish to recover are spatially homogeneous and obey some smoothness conditions, and would develop estimation procedures based on local averages of the noisy data [34].

1.1 Chirps

Our interest in this paper does not concern smooth objects but rather signals which exhibit a highly oscillatory behavior. Such signals are known under the name of *chirps* and take the general form

$$f(t) = A(t)\cos(\lambda\varphi(t)); \tag{1.3}$$

here λ is a (large) base frequency, $\varphi(t)$ is time-varying and the amplitude $A(t)$ is slowly varying. To fix ideas, consider the chirp prototype $f_0(t) = \cos(\pi N t^2/2)$, with $1/N$ being the sampling rate. This example illustrates several characteristics of chirps. First, chirps may oscillate rapidly and this paper will consider signals which may actually change sign at nearly the sampling rate. Second, chirps may span a wide range of frequencies; in fact, the frequency of oscillation of a chirp may be rapidly changing over time so that at different times, a chirp may display very different patterns. It is clear that chirps are very different from the type of smooth and homogeneous objects the literature traditionally assumes. This is the reason why most of the statistical methodology such as smoothing splines, kernel smoothers, etc. would make no sense in this setting.

Chirps are ubiquitous in nature and arise in a number of scientific disciplines. In speech processing for example, voice signals possess chirping components. In the neurosciences, researchers are interested in the study of echolocating animals; in echolocation a series of short, chirping signals are emitted by an animal which are then reflected by objects and surfaces creating an echo; the animal processes this echo which gives it a sense of the surrounding elements. Finally, in contemporary physics, there is an ongoing effort aimed at detecting gravitational waves, many of which are chirping waveforms.

The strong nonstationary character of chirping signals causes Fourier series to be very inefficient for representing chirps. For example, our chirp prototype f_0 does not interact significantly with monofrequency sinusoids and all the Fourier coefficients of f_0 are of a very small size. Hence, it would be highly problematic to use the Fourier transform in a setting where one would wish to detect or recover f_0 from noisy or cluttered data. Those limitations spurred the development of time-frequency analysis which aim at developing a mixed representation of a signal in terms of a double sequence of elementary signals, each of which occupies a certain domain in the time-frequency plane. After decades of research in this field, we now have transforms which provide time-frequency phase portraits of a signal, as well as new orthonormal bases and frames which can efficiently represent certain kinds of time-frequency phenomena: here we mention Wilson Bases [37], Cosine Packets [9], and Gaussian Frames [22].

1.2 Time-Frequency Analysis and Statistical Theory

Many of the problems which arise in time-frequency analysis are statistical in nature. For instance a vast literature [24, 31, 6, 21, 7] is concerned with the following detection problem. We suppose we have noisy data $y(t) = \alpha f(t) + z(t)$ where α is unknown and $f(t)$ is unknown (we may have some partial information about f) and chirping. We wish to know whether $\alpha = 0$ or not; that is, whether there is signal or not. In the field of astrophysics alone, there are hundreds of publications on this subject in connection with the problem of detecting gravitational waves, see [32, 25, 1] and references therein. The same would apply to the estimation problem and in short, a large portion of the Time-Frequency literature is concerned with the analysis of data which are subject to measurement errors.

Despite a great deal of activity in this field, very little is known—if anything at all—about simple statistical questions such as how well one can recover chirps from noisy data. Researchers in time-frequency analysis have successfully assembled a great collection of tools but not much is known about how one can deploy these tools to design optimally sensitive detectors and optimally efficient estimation procedures. From this viewpoint, it seems fair to say that the concepts of statistical theory do not play an important role in the literature of Time-Frequency analysis.

Our last statement is a little provocative but certainly justified in light of the strong interactions between the agendas of statistical theory and other areas of applied harmonic analysis and applied mathematics in general. For instance, there is a long tradition linking classical Fourier analysis with statistical estimation going back to the work of Wiener and the estimation of Gaussian processes, and the work of Kolmogorov and Tikhonov on statistical regularization. More recently, Pinsker [33] has shown how the damping of Fourier coefficients is asymptotically optimal for recovering smooth objects obeying constraints of the form $\|f^{(m)}\|_{L_2}^2 \le R^2$ ($f^{(m)}$ is the mth derivative of f) from Gaussian white noise, see also [20]. In a different area, Wahba and others [36] have established deep connections between nonparametric regression and the theory of splines. And of course, in a series of pathbreaking papers, Donoho and his collaborators [18, 17] have connected the field of wavelet analysis with long-standing problems in nonparametric statistics by proposing the wavelet shrinkage, a simple and—in some sense—universal estimation principle with provably optimal properties over a wide range of classes of inhomogeneous signals.

1.3 Challenge

This paper is an attempt to establish a link between concepts from time-frequency analysis and from statistical theory. We are interested in (1) developing a quantitative approach to time-frequency analysis and (2) developing methods which are flexible, i.e. which can adapt to a variety of situations as opposed to strategies based on restrictive parametric assumptions. For example, we would like to know how well one can recover chirps from noisy measurements and quantify the ultimate limits of performance. And we would like to know whether it is possible to design practical algorithms which would be amenable to rigorous analysis with provably optimal or near optimal properties; that is, which would come close to those limits of performance.

Here we understand the word 'optimal' in a mathematical sense which requires the development of a mathematical model and of quantitative analysis establishing an optimality result for that model as we explain below.

1.4 Classes of Chirps

We define classes of chirps by penalizing the roughness of the amplitude A and instantaneous frequency φ'. To measure this, we introduce the Hölder regularity s defined as follows. For $0 < s \leq 1$, we say that $g(t)$ is in the Hölder class $\text{Hölder}^s(C)$ if $\|g\|_{L_\infty} \leq C$ and

$$|g(t) - g(t')| \leq C \cdot |t - t'|^s, \quad 0 \leq t, t' \leq 1; \tag{1.4}$$

for $s > 1$ and $m < s \leq m + 1$, the Hölder class $\text{Hölder}^s(C)$ is the collection of functions g obeying $\|g\|_{L_\infty} \leq C$, and

$$|g^m(t) - g^m(t')| \leq C \cdot |t - t'|^{s-m} \quad 0 \leq t, t' \leq 1.$$

We now define a class of chirps $\text{Chirp}(s;\, \lambda, R)$ for signals of the form

$$f(t) = A(t)\cos(\lambda\varphi(t)), \quad \text{or} \quad f(t) = A(t)\exp(\mathbf{i}\lambda\varphi(t)).$$

by imposing

$$\text{Chirp}(s;\, \lambda, R) = \{f,\;\; A, \varphi \in \text{Hölder}^s(R) \leq C,\, |\varphi'(t)| \leq \pi\}, \quad 1 \leq \lambda \leq N,\, R > 0. \tag{1.5}$$

The condition $|\varphi'(t)| \leq \pi$ quantifies the maximum oscillation rate of those elements $f \in \text{Chirp}(s;\, \lambda, R)$ since it says that their frequency of oscillation is in some sense bounded by λ. Of special interest is the situation in which $\lambda = N$; in this case, the model $\text{Chirp}(s;\, N, R)$ allows for oscillations at nearly the sampling rate, i.e. the sign of f is allowed to switch at nearly each sampling point. In this case, note that $|\varphi'(t)| \leq \pi$ becomes an identifiability condition, as otherwise the signal may have several cycles in-between sample points.

We would like to point out that our model is somewhat nonclassical. In the literature of statistics, it is standard to define a fixed functional class and study the asymptotic behavior as the sample size increases; i.e. we are gathering more sample about a fixed object of interest. Suppose we take $\lambda = N$ in (1.5). Then our model changes with N and because the underlying objects A and φ can be the same, we speak of a collection with R fixed and N variable as a chirp class; that is, we are really interested in the asymptotics $\lambda_N \to \infty$ as $N \to \infty$. This is a necessary ingredient to develop a meaningful theory as the asymptotics $N \to \infty$ with λ fixed would correspond to classical smoothing setups and would be uninteresting. In any event, since the main results of this paper are nonasymptotic, we can always think about N as fixed (large).

We note that related models were introduced by Donoho and Johnstone [16], and Meyer [23] who proposed the study of the asymptotic properties of chirps (1.3) as $\lambda \to \infty$. Our definition is more precise than that of Chassande and Flandrin [7].

1.5 Quantifying Performance

To make make things concrete, suppose that $\lambda = N$ so that we are interested in the recovery of signals of the form $A(t)\cos(N\varphi(t))$. In this section, we let $\mathcal{F} = \textsc{Chirp}(s; N, R)$ for some $s \geq 2$; that is, we suppose that the "instantaneous frequency" φ' is in some sense differentiable. Then any linear procedure T yielding an estimate $\hat{f} = TY$ would obey

$$\sup_{f \in \mathcal{F}} MSE(\hat{f}, f) \geq A, \tag{1.6}$$

for some constant A; that is, the MSE would not even decay as the sample size increases! In other words, setting $\hat{f} = 0$ would in some sense be just as good as any linear estimator. Next, following the recent literature on nonlinear estimation, we may want to investigate the properties of estimators which seek to exploit the sparsity of the object to recover in a preferred basis. Consider for instance a nice wavelet thresholding estimate $\hat{f}^{Wave}$. Then the risk of such an estimate would also obey

$$\sup_{f \in \mathcal{F}} MSE(\hat{f}^{Wave}, f) \geq A.$$

Moreover, one would essentially obtain the same behavior if instead, we were to threshold the coefficients in a Fourier basis, $\sup_{f \in \mathcal{F}} MSE(\hat{f}^{Fourier}, f) \geq A$. In comparison, consider an estimation procedure based on the thresholding of nice Gabor expansions. Then it is possible to tune the window size as to achieve

$$\sup_{f \in \mathcal{F}} MSE(\hat{f}^{Wave}, f) \sim A \cdot N^{-2/5},$$

which is considerably better but far from the optimal asymptotic behavior.

1.6 Optimal Behavior

Suppose that $\lambda = N$ so that $\mathcal{F} = \textsc{Chirp}(s;\, N, R)$, for some $s \in [2,3]$. A main result of this paper is the construction of estimators $\hat{f}$ with the property

$$\sup_{f \in \mathcal{F}} MSE(\hat{f}, f) = O(\log N) \cdot N^{-\frac{2(s-1)}{2s+1}}. \tag{1.7}$$

The behavior is essentially optimal. No estimator can achieve an essentially better rate uniformly over $\mathcal{F}$. Indeed, let $M^*(N, \mathcal{F})$ be the minimax risk

$$M^*(N, \mathcal{F}) = \inf_{\hat{f}} \sup_{\mathcal{F}} MSE(\hat{f}, f). \tag{1.8}$$

Then the minimax risk obeys

$$M^*(N, \mathcal{F}) \geq c \cdot N^{-\frac{2(s-1)}{2s+1}},$$

which shows that ignoring log-like factors, our estimation procedures are optimal as regards rate of convergence. Moreover, our methodology is adaptive in the sense that it achieves nearly the minimax risk for any value of the unknown degree of regularity of the amplitude and the instantaneous frequency, $2 \leq s \leq 3$. In addition, we will show that one may construct such estimators using low-complexity algorithms, typically of the order of $N^{4/3}$ operations for a signal of size N.

Underlying our methodology is the inspiration of computational harmonic analysis whose aim is to find new types of data representations, fast algorithms to compute these, and apply these tools to practical problems [10, 13].

1.7 Organization of the Paper

In section 2, we introduce a library of multiscale chirplets. Estimation strategies together with the main results of this paper are presented in section 3. In section 4, we study the computational complexity of the proposed algorithms. Sections 5, 6 and 7 are concerned with the proofs and analysis of our results: section 5 develops oracle inequalities which are an essential tool for deriving upper bounds on the mean-squared error of estimation while section 6 develops approximation properties of the multiscale chirplet libraries; section 7 provides a proof of lower bounds. Section 8 opens up the discussion by suggesting areas for future research and proposing a main challenge to the time-frequency research community.

2 Multiscale Chirplets

2.1 Recursive Dyadic Partitions

We begin with some notations and terminology. We let I be a dyadic interval $I = [k2^{-j}, (k+1)2^{-j})$ for $k \in \mathbb{Z}$ and an integer $j \geq 0$. Recall the definition of a Recursive Dyadic Partition (RDP) which is classical in the literature of time-frequency analysis.

Definition 2.1 *The set of recursive dyadic partitions of the unit interval* $[0,1)$ *are those partitions* $\mathcal{P}$ *constructed with the following rules:*

1. *the trivial partition* $\mathcal{P} = \{[0,1)\}$ *is an RDP;*
2. *if* $\mathcal{P} = \{I_1, \ldots, I_m\}$ *is any RDP, then the partition obtained by splitting any interval* I_j *into two adjacent dyadic intervals is also an RDP.*

For instance, the partition $\mathcal{P} = \{[0,1/2), [1/2,3/4), [3/4,1)\}$ is an RDP of the unit interval while the partition $\mathcal{P} = \{[0,1/4), [1/4,3/4), [3/4,1)\}$ is not. This definition may be extended to the real line and we will say that $\mathcal{P}$ is an RDP if it may be obtained from the initial partition $\{[n,n+1),\ n \in \mathbb{Z}\}$ by recursive dyadic partitioning. Among all recursive dyadic partitions, we distinguish those which we call *balanced*:

Definition 2.2 *We say that an RDP is* **balanced** *if any two adjacent dyadic intervals* $|I|, |I'| \in \mathcal{P}$ *obey*

$$1/2 \leq |I|/|I'| \leq 2. \tag{2.1}$$

This definition appears explicitly in [35] where it is argued that although the doubling condition (2.1) is a restriction on the allowed RDP's, two adjacent intervals in an arbitrary RDP can be only of very different sizes at special dyadic locations. The essential point here is that balanced RDP's (BRDP) allow the use of windows w_I which obey a uniform bound on their time-frequency concentration [35].

The construction begins with a smooth nondecreasing cutoff function $\rho \in C^d$ obeying $\rho(t) = 0$ for $t < -1/2$ and $\rho(t) = 1$ for $t > 1/2$. We suppose further that ρ obeys $\rho(t)^2 + \rho(-t)^2 = 1$ in the region $|t| < 1/2$. Let $I = [t_I, t'_I)$ be a dyadic interval and define the window

$$w_I^{\epsilon,\epsilon'}(t) = \rho\left(\frac{t - t_I}{\epsilon}\right) \cdot \rho\left(\frac{t'_I - t}{\epsilon'}\right) \tag{2.2}$$

For example, for $I = [0,1)$ and $\epsilon = \epsilon' = 1$, $w_I^{\epsilon,\epsilon'}(t) = w(t)$ is a smooth window which vanishes outside of the interval $[-1/2, 3/2)$. Suppose we are given a BRDP $\mathcal{P}$. For each ordered pair (I, I') of adjacent intervals at the dyadic segmentation point t'_I, we define the cut-off

$$\epsilon'_I = \min(|I|, |I'|). \tag{2.3}$$

With this choice of cut-off, we introduce a family of windows associated with the partition $\mathcal{P}$

$$w_I^{\epsilon_I,\epsilon'_I}(t) := w^{\alpha_I}\left(\frac{t - t_I}{|I|}\right), \tag{2.4}$$

where w^α, $\alpha = \alpha^\pm$, is the basic window

$$w^\alpha(t) = \rho(2^{\alpha^-} t) \cdot \rho(2^{\alpha^+}(1-t)), \quad \alpha^\pm \in \{0,1\}.$$

For cutoffs given by the rule (2.3), the parameters obey

$$2^{\alpha_I^-} = |I|/\epsilon_I, \quad 2^{\alpha_I^+} = |I|/\epsilon'_I.$$

Note that the length of the support of $w_I^{\alpha_I}$ is at most $2|I|$. In the remainder of this paper, we shall denote by η the pair (ϵ, ϵ') so that we may write our dyadic windows (2.4) as $w_I^{\eta_I}$. At times, we will abuse notations and actually drop the dependence on the gender η.

It follows from the properties of the cut-off function ρ, that the family of windows $(w_I^{\eta_I})_{I \in \mathcal{P}}$ is an orthonormal partition of unity; that is, this collection obeys

$$\sum_{I \in \mathcal{P}} |w_I^{\eta_I}(t)|^2 = 1. \tag{2.5}$$

2.2 Tight Frames of Chirplets

As remarked earlier, the support of each window $w_I^{\eta_I}$ is contained in the interval $\tilde{I} = [(k-1/2)2^{-j}, (k+3/2)2^{-j})$ of size $2|I|$. Since

$$u_{I,n} = e^{\mathbf{i}\pi n t/|I|}/\sqrt{2|I|}, \quad n \in \mathbb{Z}, \tag{2.6}$$

is an orthobasis for $L_2(\tilde{I})$, we have available the following Parseval relation

$$\sum_n |\langle f w_I, u_{I,n}\rangle|^2 = \|f w_I^{\eta_I}\|_{L_2}^2$$

valid for any (real or complex-valued) signal f. Then consider the family of *multiscale chirplets*

$$V_{b,I,n}(t) = \frac{1}{\sqrt{2|I|}} \cdot w_I^{\eta_I}(t) e^{\mathbf{i} b_I t^2/2} e^{\mathbf{i}\pi n t/|I|}, \tag{2.7}$$

for all dyadic intervals I, and choices of cutoffs η_I and sequences $b = (b_I)_I$. For each interval I, the parameter b_I may only take on a discrete set of values; unless specified otherwise, we shall assume a scale-dependent discretization of the form

$$b_I = \ell \cdot 2^j \cdot \delta_j, \quad \ell = 0, \pm 1, \pm 2, \ldots, \quad \text{and } |b_I| \le B. \tag{2.8}$$

Hence, chirplets occur at all possible scales 2^{-j} and at all possible dyadic locations $k2^{-j}$, $k \in \mathbb{Z}$, and assume a wide array of base frequencies $a_I = n/|I|$ and chirping rates b_I.

Some subcollections of the dictionary of multiscale chirplets are, of course, of special interest. Suppose we are given a BRDP $\mathcal{P}$ and an arbitrary sequence of chirp rates $(b_I)_{I \in \mathcal{P}}$. Then for any signal f,

$$\sum_n |\langle f, V_{b,I,n}\rangle|^2 = \int |f(t)|^2 \, |w_I(t)|^2 \, dt,$$

and, therefore, it follows from (2.5) that the family of chirplets $(V_{b,I,n})_{I \in \mathcal{P}, n \in \mathbb{Z}}$ obeys

$$\sum_{I \in \mathcal{P}} \sum_n |\langle f, V_{b,I,n}\rangle|^2 = \|f\|_{L_2}^2. \tag{2.9}$$

This equality says that $(V_{b,I,n})_{I,n}$ is a tight frame and standard arguments give the reproducing formula

$$f = \sum_{I \in \mathcal{P}} \sum_n \langle f, V_{b,I,n}\rangle V_{b,I,n}, \tag{2.10}$$

with equality holding in an L_2 sense.

An important issue with the tight frames of chirplets $(V_{b,I,n})_{I\in\mathcal{P},n\in\mathbb{Z}}$ as defined above is that in general, they do not provide sparse representation of signals of the form $f(t) = A(t)\cos(\lambda\varphi(t))$ with λ a large parameter. It is possible, however, to adapt to this situation by considering the richer class

$$V^+_{b,I,n} = V_{b,I,n}/\sqrt{2}, \quad V^-_{b,I,n} = \overline{V_{b,I,n}}/\sqrt{2}, \tag{2.11}$$

where the bar sign indicates complex conjugation. It then follows from (2.9) that the system $(V^{\pm}_{b,I,n})$ is also a tight frame. The effect of the conjugation is to allow chirplets with opposite chirp-rates $(b_I, -b_I)$ in each tight frame—a feature which would be desirable to efficiently represent objects like

$$A(t)\cos(\lambda\varphi(t)) = A(t)\exp(\mathbf{i}\lambda\varphi(t)) + A(t)\exp(-\mathbf{i}\lambda\varphi(t)).$$

We summarize the results collected so far in the following definition.

Definition 2.3 *We let $\mathcal{D}_{\text{CHIRPLETS}}$ be the dictionary of all chirplets of the form (2.7) and denote by*

- *$\mathcal{L}_{\text{CHIRPLETS}}$ the library of all tight frames of the form $(V_{b,I,n})_{I\in\mathcal{P},n\in\mathbb{Z}}$,*
- *and $\mathcal{L}^+_{\text{CHIRPLETS}}$ the library of all tight frames of the form $(V^{\pm}_{b,I,n})_{I\in\mathcal{P},n\in\mathbb{Z}}$,*

where as before $\mathcal{P}$ ranges over all possible choices of BRDP, and b ranges over all possible sequences of discrete chirp rates.

Haykin and Mann [31] have proposed the so-called chirplet transform of a signal (we would also like to mention [4]): starting from the Gaussian multiparameter collection of linear chirps

$$g_\lambda(t) = g((t-b)/a)e^{\mathbf{i}(\omega t+\delta t^2)}, \quad \lambda = (a, b, \omega, \delta)$$

with g a Gaussian window and $a > 0$, $b, \omega, \delta \in \mathbb{R}$, they define the chirplet transform of a signal f as being the collection of inner products $\langle f, g_\lambda\rangle$. In short, the chirplet transform is a multiscale Gabor type transform with an extra modulation parameter δ. Unlike the wavelet transform or the Gabor transform, however, there is no real formula for synthesis, i.e. for reconstructing a signal from the datum of its coefficients $\langle f, g_\lambda\rangle$. This lack of synthesis rules sets their work apart from ours as we have a library of tight frames with trivial formulae for both analysis and synthesis. In addition, we will address discretization issues below.

2.3 Orthonormal Bases of Chirplets

Following [8], we may introduce a slightly different dictionary of chirplets and consider objects of the form

$$U_{I,n}(t) = \sqrt{\frac{2}{|I|}} \cdot w_I^{\eta_I}(t)e^{\mathbf{i}(a_I t+b_I t^2/2)} \sin\left[\pi(n+1/2)\frac{t-t_I}{|I|}\right], \quad n = 0, 1, 2, \ldots. \tag{2.12}$$

for all dyadic interval I, and choices of cutoffs η_I and phases $a_I t + b_I t^2/2$. Note that [8] does not mention BRDP's and the like. A distinguished feature is that under a special condition, the system $(U_{I,n})$ is an orthonormal basis.

Theorem 2.4 *Suppose we are given a BRDP $\mathcal{P}$ and a family of phases such that the piecewise affine function $a_I + b_I t$, $t \in I$, be continuous. Then the collection of chirplets $(U_{I,n})_{I \in \mathcal{P}, n \geq 0}$ is an orthobasis of $L_2(\mathbb{R})$.*

The proof of this theorem is an adaptation of the argument presented in [8] and is omitted. Suppose that the chirping parameter b_I is discretized as before. Then an equivalent realization of the dictionary of orthonormal chirplets is as follows: for each $j_0 \leq j \leq j_1$, we mark out 2^j equally spaced vertical lines in the notional time-frequency square $[0,1]^2$; we put tick marks along the vertical lines at spacing δ_j; we then create a dictionary of 'chirplet lines' connecting tick marks on adjacent vertical lines and such that in absolute value, the slope of each line is less than B. For reference, we will denote this library by $\mathcal{L}^{\mathbb{O}}_{\text{CHIRPLETS}}$ where the symbol $\mathbb{O}$ is meant for "orthogonal."

Because exact orthogonality is not critical in this paper, and because this orthogonal system is more delicate to handle, we shall not make an extensive use of this construction and merely mention this possibility for the sake of completeness. Indeed, the continuity constraint would impose a larger dictionary size and more complex search algorithms, see section 4. In addition, enriching this collection to efficiently deal with real-valued chirps (2.11) would actually defeat the orthogonality property.

2.4 Discrete Chirplet Analysis

The ideas presented above are readily applicable to the analysis of discrete signals of finite length as one can, of course, define discrete analogs of chirplet dictionaries and associated libraries of discrete tight-frames. Suppose $N = 2^J$ is dyadic and fix a scale $0 < j < J$, so that the time interval $I = [k2^{-j}, (k+1)2^{-j})$ contains $N \cdot 2^{-j}$ sample points. Viewing a discrete signal (f_t), $t = 0, \ldots, N-1$, as equispaced samples of the form $f(t/N)$, we would introduce the family of discrete chirplets,

$$V^D_{b,I,n}[t] = V_{b,I,n}(t/N), \quad -N2^{-j} \leq n < N2^{-j}.$$

The only possible issue might concern the boundary windows; that is those dyadic intervals which abut the endpoints 0 and 1. For those intervals, one would need to consider special boundary-adapted windows using ideas such as symmetrization and folding. It is not the scope of this paper to discuss these issues. There are well-known techniques for constructing boundary adapted windows, see for example [29] and references therein.

In practical applications, one would need to specify a cut-off function ρ which generates the family of dyadic windows w_I. There is a vast literature [30] about the choice of such cut-offs and a frequently discussed approach is to take $\rho(t)$ as

$$\rho(t) = \sin\left(\frac{\pi}{4}(1 + \sin(\pi t))\right), \quad |t| < 1/2,$$

and $\rho(t) = 0$ for $t \le -1/2$ and $\rho(t) = 1$ for $t \ge 1/2$.

We now discuss the size of the dictionary of multiscale chirplets for signals of size N. At each scale 2^{-j}, the number M_j of distinct chirplets in $\mathcal{D}_{\text{CHIRPLETS}}$ would not exceed

$$M_j \le 4 \cdot 2^j \cdot 2N/2^j \cdot \# \text{ slopes};$$

the first term in the product is the number of different window genders, the second is the number of intervals at scale 2^{-j} and the third is the number of sines for each interval I. Take

$$\delta_j = \pi\, 2^j/N, \quad b_I = \pi \cdot \ell \cdot 2^{2j}/N \tag{2.13}$$

in (2.8) so that the number of slopes at scale 2^{-j} is about $\max(1, N/2^{2j})$, we would have

$$M_j \le 16 \cdot N^2 \cdot 2^{-2j}.$$

Unless specified otherwise, we will always assume the special discretization (2.13). Using the Fast Fourier transform, it is clear that one can compute all chirplet coefficient at a given scale in $O(M_j \log N)$ for a signal of size N.

In the remainder of the paper and for theoretical purposes, we shall primarily be interested in the range of scales $J_0 \le j \le J_1$, where

$$2^{-J_0} \sim N^{-1/3}, \qquad 2^{-J_1} \sim N^{-1/2}.$$

Assume that the dictionary is restricted to this range of scales. Then

- The number M_N of distinct elements in the dictionary obeys $M_N \le 32 \cdot N^{4/3}$.
- The *chirplet analysis* of a signal f is the collection of inner products $\langle f, g\rangle$ for all g in the dictionary $\mathcal{D}_{\text{CHIRPLETS}}$. It is possible to compute the chirplet analysis of a signal of size N in $O(N^{4/3} \log N)$. Moreover, given the chirplet coefficient of an object in a tight-frame $\Phi \in \mathcal{L}_{\text{CHIRPLETS}}$, we can synthesize the corresponding objects in $O(N \log N)$ operations.
- The number of distinct tight frames in $\mathcal{L}_{\text{CHIRPLETS}}$ is, however, exponential in N.

In short, just as cosine packets [9, 10] are libraries of rapidly constructible orthonormal bases for signal expansions, multiscale chirplets build libraries of tight-frames which are also rapidly constructible. Moreover, chirplet libraries share a powerful feature with cosine packets libraries [10], namely, they are rapidly searchable for a “best” tight-frame.

3 Main Result

This section introduces estimation strategies for recovering chirps from noisy data and states the main results of this paper. Before we develop our ideas, we would like to emphasize that we are interested in practical and flexible methods which do not heavily rely on modeling assumptions and can be deployed in a variety of different settings.

3.1 Complexity Penalized Estimation

Suppose we are given a library $\mathcal{L}$ of dictionaries $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L\}$. Here, a dictionary $\mathcal{D}$ is an arbitrary collection of waveforms $(g_\lambda)_{\lambda \in \Lambda}$; e.g., in the discrete setup (1.1), a dictionary would simply be a collection of finite-dimensional vectors. For a sequence θ, we let $\mathcal{N}_0(\theta)$ be the number of nonzero entries in θ, $\mathcal{N}_0(\theta) = \#\{|\theta_i| \neq 0\}$ and define the complexity $\mathcal{N}_{\mathcal{D}}(f)$ of an object f (with respect to the dictionary $\mathcal{D}$) as the minimum number of terms in $\mathcal{D}$ needed to represent f

$$\mathcal{N}_{\mathcal{D}}(f) = \inf_\theta \mathcal{N}_0(\theta), \quad f = \sum_\lambda \theta_\lambda g_\lambda. \tag{3.1}$$

Likewise, we let $\mathcal{N}_{\mathcal{L}}(f)$ be the minimum number of terms needed to represent f in any dictionary $\mathcal{D} \in \mathcal{L}$

$$\mathcal{N}_{\mathcal{L}}(f) = \inf_{\mathcal{D} \in \mathcal{L}} N_{\mathcal{D}}(f), \tag{3.2}$$

as in e.g. [15]. Equipped with these concepts, we now introduce the *complexity functional* K

$$K_\Lambda(y, \tilde{f}) = \|y - \tilde{f}\|_2^2 + \Lambda \cdot \mathcal{N}_{\mathcal{L}}(\tilde{f}), \tag{3.3}$$

and for data (1.1), define the estimator $\hat{f}$ as that object which minimizes the empirical complexity

$$\hat{f} = \operatorname*{argmin}_{\tilde{f}} K_\Lambda(y, \tilde{f}). \tag{3.4}$$

In other words, we seek an estimator which achieves the best trade-off between goodness of fit and complexity—a central principle in the modern literature on statistical estimation which is commonly referred to as *complexity penalized estimation.*

In truth, this estimation procedure is highly unrealistic. Indeed, for each dictionary in the library, the minimization (3.4) requires solving a combinatorial problem. For a signal of size N, one would need to compare at least 2^N models, assuming each dictionary of size greater or equal to N. This is not practical. To put it bluntly: we can certainly talk about $\hat{f}$ and its properties but this is about it!

3.2 Thresholding in the Best Empirical Tight Frame

Suppose the library $\mathcal{L}$ is a collection of tight frames $\{\Phi_1, \Phi_2, \dots, \Phi_L\}$. In a discrete setting, a tight frame Φ (with frame bounds equal to one) is a collection of vectors g_λ obeying

$$\|\Phi^* f\| = \|f\|, \quad \forall f \in \mathbb{C}^N,$$

where Φ is the matrix whose columns are the vectors g_λ (and Φ^* is the adjoint matrix). Note that the above tight-frame property implies $\Phi\,\Phi^* = I$. Define the *new* complexity functional

$$\Lambda(y; \tilde{\Phi}, \tilde{\theta}) = \|\tilde{\Phi}^* y - \tilde{\theta}\|^2 + \Lambda \cdot \mathcal{N}_0(\tilde{\theta}). \tag{3.5}$$

For data (1.1), we then define the estimator $\hat{f}$ by

$$\hat{f} = \hat{\Phi}\hat{\theta}, \qquad (\hat{\Phi}, \hat{\theta}) = \underset{\tilde{\Phi}\in\mathcal{L},\ \tilde{\theta}}{\operatorname{argmin}} J_\Lambda(y; \tilde{\Phi}, \tilde{\theta}). \tag{3.6}$$

Observe that if $\mathcal{L}$ is a library of orthogonal bases, then both estimation procedures (3.4)–(3.6) yield the same estimator. Indeed, when Φ is an orthonormal transform, there is a one-to-one correspondence between $\tilde{\theta}$ and $\tilde{f}$, namely, $\tilde{\theta} = \tilde{\Phi}^* \tilde{f}$ and thus

$$\|y - \tilde{f}\|_2^2 + \Lambda \cdot N_\Phi(\tilde{f}) = \|\tilde{\Phi}^* y - \tilde{\theta}\|^2 + \Lambda \cdot \mathcal{N}_0(\tilde{\theta}),$$

which shows that the two complexity functionals are actually identical. In general setups, however, the two proposals (3.4) and (3.6) yield different estimators. There are two aspects of special interest here: first, the complexity penalized estimator (3.6) provably attains optimal bounds over our classes of chirps and second, this estimator has a nice interpretation which makes it computationally attractive.

For convenience, we let $\theta_i[\Phi]$ denote the coordinates of the frame coefficients $\Phi^* f$ and define the entropy of an object f in the frame Φ as

$$\mathcal{E}_\Lambda(f, \Phi) = \sum_i \min(|\theta_i[\Phi]|^2, \Lambda). \tag{3.7}$$

Given a vector of observations, we let $\hat{\Phi}$ be the basis which minimizes the entropy:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}}\, \mathcal{E}_\Lambda(y, \Phi).$$

Then the estimator $\hat{f}$ (3.6) is obtained by hard-thresholding in the "best" empirical tight frame $\hat{\Phi}$: with $\eta_\Lambda(t)$ the scalar nonlinearity $\eta_\delta(t) = t \cdot 1_{\{|t|>\delta\}}$, we have

$$\hat{\theta}_i = \eta_{\sqrt{\Lambda}}(y_i[\hat{\Phi}]), \tag{3.8}$$

and of course $\hat{f} = \hat{\Phi}\hat{\theta}$. To see why this holds, simply observe that for a fixed Φ, that vector $\theta^\star$ which minimizes $J(y;\, \Phi, \tilde{\theta})$ is given by $\theta_i^\star = \eta_{\sqrt{\Lambda}}(y_i[\Phi])$ and, therefore, $J(y;\, \Phi, \theta^\star) = \mathcal{E}_\Lambda(y, \Phi)$.

The point here is that the hierarchical dyadic structure of $\mathcal{L}_{\text{CHIRPLETS}}$ or $\mathcal{L}^+_{\text{CHIRPLETS}}$ and the additivity of the entropy functional makes possible to search these libraries for the best empirical tight-frame very rapidly. In fact, in section 4 we will see that the computational cost of the search is in some sense negligible compared to the computational cost of the chirplet analysis. Hence, at least from a computational viewpoint, (3.5) is a very practical estimator.

3.3 Main Results

In what follows, we consider the following two estimation problems: first,

$$y_i = f_i + z_i, \quad i = 0, 1, \ldots, N-1, \tag{3.9}$$

where (z_i) is a Gaussian White Noise, z_i i.i.d. $N(0, \sigma^2)$; and second,

$$Y_i = F_i + Z_i, \quad i = 0, 1, \ldots, N-1, \tag{3.10}$$

where Y is a two-dimensional vector of observations $(y_{1,i}, y_{2,i})$, $F = (F_{1,i}, F_{2,i})$ is the object we wish to recover, and Z is a Gaussian error $Z = (z_{1,i}, z_{2,i})$, with z_1, z_2 independent and each i.i.d. $N(0, \sigma^2/2)$ so that the noise level is the same as in (3.9), $E|Z_i|^2 = \sigma^2$ (this will simplify the exposition). In this setting, the MSE of an estimator $\hat{F} = (\hat{F}_1, \hat{F}_2)$ is given by $MSE(F, \hat{F}) = E\|F_1 - \hat{F}_1\|^2 + E\|F_2 - \hat{F}_2\|^2$. Note that (3.10) is equivalent to recovering $f = F_1 + \mathbf{i}F_2$ from the noisy data

$$y = f + z, \tag{3.11}$$

where $y = y_1 + \mathbf{i}y_2$, and $z = z_1 + \mathbf{i}z_2$. We will refer to these two problems as the *real* and *complex/bivariate* problems.

We are interested in quantifying the best performance attainable over the classes of chirps and seek estimators which achieve or nearly achieve the minimax risk (1.8) for $\mathcal{F} = \text{Chirp}(s;\, \lambda, R)$, for some $s \in [2, 3]$. Recall from the introduction section that $f \in \text{Chirp}(s;\, \lambda, R)$ if $f(t) = A(t)\cos(\lambda\varphi(t))$ with the amplitude A and φ obeying (1.5), or$f(t) = A(t)\exp(\mathbf{i}\lambda\varphi(t))$ with the same assumptions on A and φ.

Theorem 3.1 *Set $\lambda = N$ in (1.5). The minimax risk for both the real and complex problems obeys*

$$M^*(N, \mathcal{F}) \geq c_R \cdot N^{-\frac{2(s-1)}{2s+1}}, \tag{3.12}$$

for $\mathcal{F} = \text{Chirp}(s;\, N, R)$ and $s \in [2, 3]$. (The constant c_R may be chosen of the form $c_R = c \cdot R^{2/(2s+1)}$, for some universal $c > 0$.)

Theorem 3.2 *Let $\mathcal{L}$ be either $\mathcal{L}_{\text{Chirplets}}$ or $\mathcal{L}^{+}_{\text{Chirplets}}$ (or $\mathcal{L}^{\mathbb{O}}_{\text{Chirplets}}$), i.e. $\mathcal{L}$ is one of the chirplet libraries defined in section 2 (with the discretization (2.8)). We let M_N be the number of distinct vectors occurring in $\mathcal{L}$ and put $\sqrt{\Lambda_N} = t \cdot \sigma \cdot (1 + \sqrt{2\log(M_N)})$ with $t > 4$ in (3.3) and (3.5).*

- *Put $\mathcal{L} = \mathcal{L}_{\text{Chirplets}}$ or $\mathcal{L}^{\mathbb{O}}_{\text{Chirplets}}$ for the complex problem and $\mathcal{L} = \mathcal{L}^{+}_{\text{Chirplets}}$ for the real problem. Then in both problems, the estimator (3.4) nearly achieves the minimax risk, simultaneously over the classes of objects $\mathcal{F} = \text{Chirp}(s;\, N, R)$, $2 \leq s \leq 3$:*

  $$\sup_{\mathcal{F}} MSE(f, \hat{f}) \leq A_R \cdot \log N \cdot N^{-\frac{2(s-1)}{2s+1}}. \tag{3.13}$$

  The constant A_R may be chosen of the form $A_R = A \cdot (R+1)^{2/(2s+1)}$, for some universal $A > 0$.

- *The estimator (3.6) with $\mathcal{L} = \mathcal{L}_{\text{Chirplets}}$ or $\mathcal{L}^{\mathbb{O}}_{\text{Chirplets}}$ obeys (3.13) for the complex problem and, therefore, nearly achieves the minimax risk for the classes $\mathcal{F} = \text{Chirp}(s;\, N, R)$, $2 \leq s \leq 3$.*

In short, our estimators come within log-factors of the minimax risk

$$\sup_{\mathcal{F}} MSE(f, \hat{f}) \le O(\log N) \cdot M^*(N, \mathcal{F}).$$

In the above theorems, we fixed $\lambda = N$ as to make the results more concrete and because this choice allows for signals oscillating at the sampling rate, and has more evocative power. There are, of course, equivalent formulations of Theorems 3.1 and 3.2 for arbitrary values of λ.

Theorem 3.3 *Let $\mathcal{F} = \textsc{Chirp}(s;\, \lambda, R)$ for some $s \in [2,3]$. Then the minimax risk for both the real and complex problems obeys*

$$M^*(N, \mathcal{F}) \ge c \cdot \lambda^{\frac{2}{2s+1}} \cdot N^{-\frac{2s}{2s+1}}. \tag{3.14}$$

Theorem 3.4 *With the same setup as in Theorem 3.2:*

- *In the real and complex problems, the estimator (3.4) nearly achieves the minimax risk, simultaneously over the classes of objects $\mathcal{F} = \textsc{Chirp}(s;\, \lambda, R)$, $2 \le s \le 3$:*

$$\sup_{\mathcal{F}} MSE(f, \hat{f}) \le A \cdot \log N \cdot \lambda^{\frac{2}{2s+1}} \cdot N^{-\frac{2s}{2s+1}}. \tag{3.15}$$

- *In the real problem, the estimator (3.6) obeys (3.15) for the complex problem and, therefore, nearly achieves the minimax risk for the classes $\mathcal{F} = \textsc{Chirp}(s;\, \lambda, R)$, $2 \le s \le 3$.*

Note that for $\lambda = N$, the statements of those two theorems are actually identical to those of Theorems 3.1 and 3.2 while not surprisingly, for $\lambda = 1$, the convergence rate is that of objects with bounded Hölder regularity s, i.e. $\mathcal{F} = \textsc{Hölder}^s(C)$. In fact, the proof of these last two results are minor modifications of those of Theorems 3.1 and 3.2 and will be omitted.

We witness one more time a higher degree of adaptivity of our estimators in the sense that one does not need to know the value of the parameter λ; our estimators are nearly minimax for any value of the degree of regularity of the chirp, $2 \le s \le 3$ and any value of the base oscillation frequency λ, $1 \le \lambda \le N$.

3.4 Real Signals

To study the recovery of real signals with best empirical tight-frame type of ideas, note that we can transform any arbitrary real-valued signal $s(t)$, $t \in [0,1]$, into a complex signal by considering the analytic part $S(t)$ of $s(t)$ whose Fourier coefficients

$$c_n(S) = \int S(t) e^{-\mathbf{i}2\pi n t}\, dt$$

are defined by

$$c_n(S) = \begin{cases} 2c_n(s) & n > 0 \\ c_0(s) & n = 0 \\ 0 & n < 0. \end{cases} \quad . \tag{3.16}$$

(There is a discrete analogous for discrete signals which uses the discrete Fourier transform. Note that one could have alternatively define the analytic part of S by means of the continuous Fourier transform: $\hat{S}(\xi) = 2\hat{s}(\xi)1_{\{\xi\geq 0\}}$, $\hat{S}(\xi) = \int S(t)e^{-\mathbf{i}\xi t}\,dt$. However, the definition (3.16) is better adapted to objects which live in the interval $[0, 1]$.) It follows from the definition that $s(t)$ is the real part of the analytic signal $S(t)$

$$s(t) = \Re(S(t)).$$

Of interest is the fact that this mapping transforms a real-valued and oscillatory chirp into a corresponding complex-valued and oscillatory chirp. Formally, take $s(t)$ to be a chirp of the form $s(t) = A(t)\cos(\lambda\varphi(t))$. Then

$$S(t) = A(t)e^{\mathbf{i}\lambda\varphi(t)} + r(t) \tag{3.17}$$

where $r(t)$ is a remainder term whose size in some sense decreases as the minimum instantaneous frequency $\inf \varphi'(t) \geq 0$, say, increases.

This transformation provides a mechanism to turn estimation procedures adapted to complex-valued objects into procedures aimed at recovering real-valued signals. Consider the following estimation strategy for the real problem (3.9):

1. Apply the transformation (3.16) to the noisy data y (3.9) and obtain Y.

2. Apply the thresholding rule in the best empirical tight-frame $\Phi \in \mathcal{L}_{\text{CHIRPLETS}}$ yielding an estimate $\hat{F}$.

3. Set $\hat{f}$ to be the real part of $\hat{F}$.

We prove that this estimation procedure is asymptotically nearly optimal as well.

Theorem 3.5 *Assume that $A(0) = A(1) = 0$ or that A and φ are periodic. Suppose $|\varphi'(t)| \in [\omega_N, \pi)$ with $\omega_N = \pi \cdot N^{-1/2+\beta}$, $\beta > \frac{1}{2(2s+1)}$. With the same setup as that of Theorem 3.2, we set $\mathcal{L} = \mathcal{L}_{\text{CHIRPLETS}}$ and put $\sqrt{\Lambda_N} = t \cdot \sigma \cdot (1 + \sqrt{2\log(M_N)})$ with $t > 4$ in (3.5). Then in the real problem, the above estimator obeys*

$$\sup_{\mathcal{F}} MSE(f, \hat{f}) \leq A \cdot \Lambda_N \cdot N^{-\frac{2(s-1)}{2s+1}}.$$

Here, the periodicity assumption or the vanishing condition $A(0) = A(1) = 0$ prevent the discussion of rather nonessential issues associated with the boundaries of our interval. Our condition $|\varphi'(t)| \in [\omega_N, \pi)$ says that the chirp is genuinely oscillatory and does not have large low-frequency components. Following up on the above discussion, this condition guarantees that the remainder term in (3.17) is appropriately small.

3.5 Discussion

A first contribution of this paper is the identification of optimal rates of convergence for recovering classes of chirps from noisy data, and the construction of estimators which nearly attain these rates. (We would like to remark that the study of these optimal rates was inspired by some results in [16] although our results, our models and our methodologies are very different.) These classes encompass a wide range of phenomena of interest, such as arbitrary base frequencies λ, roughness s, and regularity R.

Further, our statistical optimality results are stated for discrete models and discrete algorithms (see sections 2 and 4) and a second contribution is the design of computationally effective methods which also provably attain these optimal rates of convergence. In fact, the next section will show that the computational complexity of the estimator (3.6) for recovering objects from $\text{Chirp}(s;\, N, R)$ is—up to logarithmic factors—of the order of $N^{4/3}$ operations. Before turning to these issues, we would like briefly to comment on the limits of this strategy.

Both estimation strategies synthesize a reconstruction by extracting a finite linear combination of a few selected chirplets. The estimator (3.6) extracts such a linear combination by applying a thresholding rule (in an empirically selected tight-frame). This is expected to perform well when thresholding provides good partial reconstructions of the underlying unknown object f, as for any complex-valued chirps, or real-valued chirps with a minimum degree of oscillation. For slowly varying chirps, however, this is not the case. Optimally sparse chirplet decompositions of such chirps exist—and this is essentially why (3.4) yields optimal rates of convergence—but they cannot be synthesized using naive ideas such as thresholding. This is the reason why our thresholding estimate would not perform well for slowly varying real chirps. One can of course design other strategies which would adapt to this situation more effectively. This is, however, not the scope of this paper.

Finally, we would like to discuss possible extensions. Because our libraries of multiscale chirplets exhibit a dyadic structure, our estimators adapt to other phenomena such as isolated singularities. For example, it would be possible to extend the results of this paper as to handle discontinuities in the amplitude A, or in the phase φ, or in the instantaneous φ'. We leave the treatment of such natural extensions to a possible future publication.

4 Computational Complexity of the Search

The algorithm for searching the best empirical tight-frame is based on dynamic programming ideas and is similar to the best-basis algorithm for cosine packets [10] and for adapted bases of local cosines [35].

It is classical to associate an RDP with a binary tree and borrowing a terminology from [13], we may want to decorate binary trees by associating to each leaf (or equivalently a dyadic interval) a number b_I (our chirping parameter) so that each *balanced* and decorated tree is associated with a chirplet tight-frame Φ. With this terminology, we describe an algorithm which searches the space of balanced decorated trees and operates by bottom-up inspection of the complete tree.

To avoid discussing issues related to the endpoints of the interval $[0, 1)$, we will assume periodic boundary conditions, i.e. that data are given on the circle. There are variants for other kinds of boundary conditions—which only make the description more technical. To describe the algorithm, we need to introduce some notations. For each dyadic interval I, we let $\mathcal{L}_I^{\ell,r}$ be the library of tight-frames Φ which correspond to a BRDP of I with a cut-off radius equal to $2^{-\ell}$ (resp. 2^{-r}) on the left (resp. on the right) of the interval I; $j \le \ell, r \le J_1$, where 2^{-J_1} is some fixed minimum window-size. Note that because of the restriction on balanced RDPs, the sets $\mathcal{L}_I^{\ell,r}$ are empty if $\ell = j$ and $r > j+1$ or vice-versa, if $r = j$ and $\ell > j+1$. There are balanced RDPs corresponding to other pairs (ℓ, r) with $\ell, r \le J_1$. For each $\Phi \in \mathcal{L}_I^{\ell,r}$, we then define the localized entropy

$$\mathcal{E}_\Lambda(I, \ell, r;\, \Phi) := \sum_i \min(|\theta_i[\Phi]|^2, \Lambda),$$

where the sum is of course restricted to those coordinates corresponding to dyadic intervals which are subsets of I. For each $(\ell, r) \in \{(j, j), (j, j+1), (j+1, j), (j+1, j+1)\}$, define $a_I^{\ell,r}$ to be the minimum value of the localized entropy corresponding to a tight frame whose corresponding partition is I and whose cut-offs radii are $2^{-\ell}$ and 2^{-r}. With the notations of section 2, such a tight frame is of the form

$$\Phi = (V_{b,I,n})_n,$$

where we recall that the gender of the window w_I^η, $\eta = (\eta^-, \eta^+)$ and $\eta^\pm \in \{0, 1\}$, is related to the cut-offs via $\eta^- = j - \ell, \eta^+ = j - r$. Hence,

$$a_I^{\ell,r} = \inf_b \mathcal{E}_\Lambda(I, \ell, r;\, \Phi), \tag{4.1}$$

where Φ is as above .

Define now $C_I^{\ell,r}$ to be the optimal value of the localized entropy among all $\Phi \in \mathcal{L}_I^{\ell,r}$:

$$C_I^{\ell,r} = \min_{\Phi \in \mathcal{L}_I^{\ell,r}} \mathcal{E}_\Lambda(I, \ell, r;\, \Phi).$$

For simplicity, we set $C_I^{\ell,r} = \infty$ if $\mathcal{L}_I^{\ell,r} = \emptyset$, e.g. $C_I^{j,j+2} = \infty$. Consider a dyadic interval I and its two children I_L and I_R, the left and right halves of I. We let Φ be that tight-frame with minimal cost in $\mathcal{L}_I^{\ell,r}$. The corresponding partition is either I, or is composed by two balanced dyadic partitions on I_L and I_R. With

$$d_I^{\ell,r} = \min_{s \in \{j+1,\ldots,J_1\}} \left(C_{I_L}^{\ell,s} + C_{I_R}^{s,r} \right),$$

it follows from the additivity property of the entropy that

$$C_I^{\ell,r} = \begin{cases} a_I^{\ell,r} & (\ell, r) \in \{(j, j), (j+1, j), (j, j+1)\} \\ \min(a_I^{\ell,r}, d_I^{\ell,r}), & (\ell, r) = (j+1, j+1) \\ d_I^{\ell,r} & \textit{otherwise}. \end{cases} \tag{4.2}$$

This formula (4.2) shows that one can compute all the costs $C_I^{\ell,r}$ using a hierarchically organized algorithm which is described below.

Algorithm:

- Step 1: we initialize the full binary tree of depth J_1 by computing for each node I, $|I| \geq 2^{-J_1}$, and each pair (ℓ, r), the quantities $a_I^{\ell,r}$ (4.1) and $b_I^{\ell,r}$—the optimal value b^* of the chirping parameter.

- Step2: starting from the finer level, we move up the tree by computing all the costs. Loop: for each I of sidelength 2^{-j},
  - Compute $C_I^{\ell,r}$ using the formula (4.2).
  - At each node I, store the following array (indexed by (ℓ, r)) of information:
    - Record the value $C_I^{\ell,r}$
    - If the minimum is achieved by $a_I^{\ell,r}$, mark the node I (for this value of (ℓ, r)) "Terminal" and record the value of $b_I^{\ell,r}$.
    - If the minimum is achieved by $d_I^{\ell,r}$, mark the node I (for this value of (ℓ, r)) "Interior" and record the value s of the optimal cut-off at the middle-point.
  - Set $j = j - 1$.
  - If $j > 0$ goto Loop.

- Step 3: When the loop terminates, all the global costs $C_{[0,1]}^{\ell,r}$ (or $C_{[0,1]}^{\ell,\ell}$ assuming periodic boundary conditions) are computed and one can then find the global minimum cost $C_{[0,1]}^{\ell^*,r^*}$ (or $C_{[0,1]}^{\ell^*,\ell^*}$). To find the optimal BRDP and optimal tight-frame, we march down the tree, following the optimal splitting rules (which are stored) and stop whenever we find a node marked "Terminal." At each terminal node, we have available the optimal value of the chirping parameter $b_I^{\ell,r}$.

Hence, when the algorithm terminates, we hold a tight-frame Φ^* from our library $\mathcal{L}$ which minimizes the entropy $\mathcal{E}$ among all tight-frames. The optimality follows from the relation 4.2); that is, from of the additivity of the entropy functional.

The most expensive part of the algorithm is the initialization step (Step 1), i.e. the computation of the localized entropies. Note that the cost of this step is at most of the order of M_N additions/multiplications and comparisons. As far as the the complexity of the search algorithm (Step 2) is concerned, observe that for each interval of length 2^{-j}, we have to compute $O((J_1 - j)^2)$ costs and each calculation requires at most $O(J_1 - j)$ additions and comparisons. There are 2^j dyadic intervals of length 2^{-j} and, therefore, the cost of each loop is of the order of $O((J_1 - j)^3) \cdot 2^j$ so that the total number of operations is of the order of

$$\sum_{j=J_0}^{J_1} O((J_1 - j)^3) 2^j = O(2^{J_1}).$$

Hence, the cost of the search in a complete tree of depth $[J_0, J_1] = [\log_2 N, 0]$ is at most of the order of N operations. In section 2, we argued that it would not make much sense to consider intervals of size smaller than $N^{-1/2}$, i.e. $N^{-1/2} \sim 2^{-J_1}$. Hence, the complexity of the search restricted to this range would be much less, namely, of the order of $\sqrt{N}$ operations.

Therefore, in a full chirplet algorithm, *the workload associated with the search is negligible compared to the computational cost of the chirplet analysis.*

5 Oracle Inequalities

In the next three sections, we will assume that the noise level obeys $\sigma = 1$. A simple rescaling argument extends our discussion to arbitrary values of σ.

5.1 Oracle Inequalities

A key ingredient for proving the upper-bound are the so-called *oracle inequalities.* [15, 19, 3]. Suppose we are given a library of tight frames $\mathcal{L} = \{\Phi_1, \ldots, \Phi_L\}$, and let $(\hat{f}_\Phi)_{\Phi \in \mathcal{L}}$ be the family of estimators obtained by applying hard-thresholding rules in the frame $\Phi \in \mathcal{L}$ as in (3.8). Then the best performance this family may achieve is the so-called *ideal MSE* [19]

$$MSE(f, \mathcal{L}) = \inf_{\Phi} MSE(f, \hat{f}_\Phi). \tag{5.1}$$

This is called ideal because, of course, we would not know which estimator $\hat{f}_\Phi$ is best; that is, to achieve the ideal MSE, one would need an oracle which would tell us which frame to choose. In this setup, an oracle inequality would be an inequality of the form

$$MSE(f, \hat{f}) \le O(\log N) \cdot (1/N + MSE(f, \mathcal{L})), \quad \forall f. \tag{5.2}$$

In other words, an estimator $\hat{f}$ obeying an oracle inequality would attain a MSE which, ignoring log-like factors, would be close to the ideal MSE.

5.2 Thresholding with Noisy Data

Consider a model (complex or real) of the form

$$y_t = f_t + z_t, \quad t = 0, \ldots, n-1, \tag{5.3}$$

where $z = (z_t)$ is Gaussian White Noise, $E|z_t|^2$, and f is the vector of unknown sampled values. Let Φ be a tight-frame. Note that this estimation problem is equivalent to

$$\Phi^* y = \Phi^* f + \Phi^* z$$

which we rewrite as

$$y_i[\Phi] = \theta_i[\Phi] + z_i[\Phi], \quad i = 1, \ldots, p. \tag{5.4}$$

Here $y_i[\Phi]$ (resp. $\theta_i[\Phi]$) are the coordinates of the vector y (resp. f) in the frame Φ and the stochastic term $z[\Phi]$ is a Gaussian vector with mean zero and covariance matrix $\Phi^*\Phi$. In particular, it follows from the tight-frame property that $\mathrm{Var}(z_i[\Phi]) \le 1$, $\forall i$. Models like (5.4) have been extensively studied [19, 27, 12, 5] where it is suggested that one applies a thresholding rule, say

$$\hat{\theta}_i[\Phi] = \eta_{\sqrt{\Lambda}}(y_i[\Phi]),$$

with $\Lambda \sim 2 \log p$. Assuming that $p = O(n)$, it is now well-known that this rule yields an estimator obeying [19, 5]

$$\begin{aligned} E\|f - \hat{f}\|_{\ell^2}^2 &\le E\|\theta[\Phi] - \hat{\theta}[\Phi]\|_{\ell^2}^2 \\ &\le O(\log n) \cdot \left(1 + \sum_i \min(|\theta_i[\Phi]|^2, 1)\right), \end{aligned}$$

with $\|\cdot\|_{\ell_2}$ the usual Euclidean norm. The first inequality follows from the fact that for any tight-frame Φ and object $f = \Phi\theta$, we have

$$\|f\| \le \|\theta\|,$$

which we apply to $f - \hat{f} = \Phi(\theta[\Phi] - \hat{\theta}[\Phi])$. The second inequality is an oracle inequality of the type developed in [19] which provides a bound about the performance of thresholding rules.

Ignoring for the moment the logarithmic factor and the constant term immediately inside brackets, we focus attention on the expression

$$\mathcal{R}(f, \Phi) = \sum_i \min(|\theta_i[\Phi]|^2, 1). \tag{5.5}$$

This acts as proxy for the mean-squared error of estimation of a threshold estimator; in studies [19, 11] it has been shown that its behavior mimics, to within logarithmic factors, the true mean-squared error of estimation.

The risk proxy (5.5) enjoys a natural interpretation in terms of the classical bias-variance trade-off. We let $|\theta[\Phi]|_{(i)}$ denote the decreasing rearrangement of the coefficients of f in the frame Φ and define $C(\theta[\Phi], m)$ be the compression number

$$C(\theta[\Phi], m) = \sum_{k>m} |\theta[\Phi]|_{(k)}^2 \tag{5.6}$$

measuring the error of reconstruction of the signal f from its m-largest terms in an expansion in the frame Φ. Letting m^* be the number of coefficients exceeding in absolute value the threshold $t = 1$, we may rewrite (5.5) as

$$\mathcal{R}(f, \Phi) = C(\theta[\Phi], m^*) + m^* = \inf_m C(\theta[\Phi], m) + m. \tag{5.7}$$

In other words, the risk proxy is the sum of the number of the terms in the expansion (which we may think of as a variance term) and the squared approximation error from the linear combination of its m-largest terms in an expansion in Φ (which we may think of as a squared-bias term).

5.3 Thresholding in a Library of Tight-Frames

Suppose we have a library of tight frames $\mathcal{L}$ and let M_n be the number of distinct vectors occurring among all tight frames in the library. We now develop an oracle inequality.

Theorem 5.1 *Pick $\lambda > 4$ and set $\sqrt{\Lambda_n} = \lambda \cdot (1 + \sqrt{2 \log M_n})$. Then the MSE of the estimator obtained by hard-thresholding in the best empirical tight-frame with threshold $t = \sqrt{\Lambda_n}$ obeys*

$$E\|\hat{f} - f\|_2^2 \leq (1 - 4/\lambda)^{-1} \cdot \Lambda_n \cdot \mathcal{R}^*(f, \mathcal{L}). \tag{5.8}$$

Although this inequality bears some resemblance with that stated in [15], it is actually quite different. First and foremost, our inequality applies to a library of tight frames whereas [15] is concerned with orthonormal bases. This is significant because it led us to construct an estimator which is different than that studied in that reference and those which are commonly studied, see section 3. Second, (5.8) develops a bound on the risk of $\hat{f}$ while instead, [15] gives a probability inequality between the loss $\|\hat{f} - f\|_2^2$ (which is random) and the ideal MSE which we recall is equal to $\inf_\Phi MSE(f, \hat{f}_\Phi)$. Therefore, (5.8) is new and of independent interest, especially in light of the fact that it may be more applicable since tight-frames are usually much easier to construct than orthobases.

Proof of Theorem. With the notations of section 3, we recall the definition of the complexity

$$J(f;\, \tilde{\Phi}, \tilde{\theta}) = \|\tilde{\Phi}^* f - \tilde{\theta}\|_2^2 + \Lambda_n \cdot \mathcal{N}(\hat{\theta})$$

and consider a pair (Φ_0, f_0) which achieves the minimum theoretical complexity

$$\Phi_0 = \operatorname{argmin} J(f; \Phi_0, \theta_0), \qquad f_0 = \Phi_0 \theta_0.$$

Since the pair $(\hat{\Phi}, \hat{f})$ has minimum empirical complexity, $(\hat{\Phi}, \hat{f})$ obeys

$$J_{\Lambda_n}(y;\, \hat{\Phi}, \hat{\theta}) \leq J_{\Lambda_n}(y;\, \Phi_0, \theta_0). \tag{5.9}$$

For convenience, put $\hat{J} = J_{\Lambda_n}(f;\, \hat{\Phi}, \hat{\theta})$ and $J_0 = J_{\Lambda_n}(f;\, \Phi_0, \theta_0)$. It follows from the decomposition $y = f + z$ that

$$\begin{aligned} J_{\Lambda_n}(y;\, \hat{\Phi}, \hat{\theta}) &= \|\hat{\Phi}^* f - \hat{\theta}\|_2^2 + 2\Re(\langle \hat{\Phi}^* z, \hat{\Phi}^* f - \hat{\theta}\rangle) + \|\hat{\Phi}^* z\|_2^2 + \Lambda_n \cdot \mathcal{N}(\hat{\theta}) \\ &= \|\hat{\Phi}^* f - \hat{\theta}\|_2^2 + 2\Re(\langle z, f - \hat{f}\rangle) + \|z\|_2^2 + \Lambda_n \cdot \mathcal{N}(\hat{\theta}) \\ &= \hat{J} + 2\Re(\langle z, f - \hat{f}\rangle) + \|z\|_2^2. \end{aligned}$$

Here we used the tight frame-property which says that $\Phi\Phi^* = I$ or equivalently that for any pair (f, g), $\langle \Phi^* f, \Phi^* g\rangle = \langle f, g\rangle$. We may develop a similar expression for J_0 and plugging these equalities on both sides of (5.9) gives

$$\hat{J} \leq J_0 + 2\Re(\langle z, \hat{f} - f_0\rangle). \tag{5.10}$$

We now let X be the random variable defined by $X = \Re(\langle z, \hat{f} - f_0\rangle)$. The following lemma gives a bound on the expectation of X.

Lemma 5.2

$$E(X) \leq 2 \cdot \lambda^{-1} \cdot E(\hat{J}). \tag{5.11}$$

Taking expectation on both sides of (5.10) and applying (5.11) gives

$$E(\hat{J}) \leq (1 - 4/\lambda)^{-1} \cdot J_0. \tag{5.12}$$

The oracle inequality (5.8) now easily follows from (5.12) and

$$\|f - \hat{f}\|^2 \leq \|\theta - \hat{\theta}\|^2 \leq \hat{J},$$

together with

$$\begin{aligned} J_0 &= \min_{\Phi} \sum_i (|\theta_i[\Phi]|^2, \Lambda_n) \\ &\leq \Lambda_n \cdot \min_{\Phi} \sum_i (|\theta_i[\Phi]|^2, 1) = \Lambda_n \cdot \mathcal{R}^*(f, \mathcal{L}). \end{aligned}$$

The theorem is established provided we verify Lemma 5.2. ∎

Proof of Lemma 5.2. Observe that $J_0 \leq \hat{J}$ and consider the event $A = \{\hat{J} \leq t\}$. On this event, we have $\|\hat{f} - f\|^2 \leq \hat{J} \leq t$ and $\|f_0 - f\|^2 \leq J_0 \leq \hat{J} \leq t$. Next, set $m = \lfloor t/\Lambda_n \rfloor$. Note that each element $\hat{f}, f_0$ is a linear combination of at most m nonzero vectors and therefore the difference $\hat{f} - f_0$ is a linear combination of at most $2m$ distinct vectors from our dictionary; we let π be the linear space of dimension at most $2m$ spanned by those vectors and denote by P_π the orthogonal projection onto π. On the event A, the Cauchy-Schwartz inequality gives

$$|\langle z, \hat{f} - f_0 \rangle| \leq \|P_\pi z\| \cdot \|\hat{f} - f_0\| \leq 2\sqrt{t} \cdot \|P_\pi z\|.$$

Obviously,

$$E\|P_\pi z\| \leq \sqrt{t/\Lambda_n} \cdot E \sup_g |\langle g, z \rangle|,$$

where g ranges over all possible vectors from our dictionary $\mathcal{D}$. Since

$$E \sup_g |\langle g, z \rangle| \leq (1 + \sqrt{2 \log M_n}),$$

the projection $P_\pi z$ obeys

$$E\|P_\pi z\| \leq \sqrt{t}/\lambda.$$

Hence, we proved that

$$E(X|\hat{J}) \leq 2 \cdot \lambda^{-1} \cdot \hat{J}.$$

Taking expectations gives the lemma. ∎

It is possible to develop oracle inequalities for other complexity functionals. Consider for example the complexity functional (3.3)

$$K(y, \tilde{f}) = \|y - \tilde{f}\|^2 + \Lambda_n \cdot \mathcal{N}_{\mathcal{L}}(\tilde{f})$$

and let $\hat{f}$ denote that object which minimizes the empirical complexity (3.4). Then, the same argument would also give

Corollary 5.3 *Pick* $\lambda > 4$ *and set* $\sqrt{\Lambda_n} = \lambda \cdot (1 + \sqrt{2 \log M_n})$. *Then the MSE of the estimator (3.4) obeys*

$$E\|\hat{f} - f\|_2^2 \le (1 - 4/\lambda)^{-1} \cdot \Lambda_n \cdot \inf_{\tilde{f}} K(f, \tilde{f}). \tag{5.13}$$

The proof would merely follows that of Theorem 5.1 and is omitted.

6 Upper Bounds

6.1 Proof of Theorem 3.2.

We will see below that for a complex chirp in $f \in \text{Chirp}(s;\, N, R)$, there is a tight frame $\Phi \in \mathcal{L}$ where, $\mathcal{L}$ is either $\mathcal{L}_{\text{Chirplets}}$ or $\mathcal{L}^{\mathbb{O}}_{\text{Chirplets}}$, in which the approximation error (5.6) obtained from the m-largest term in that frame obeys

$$N^{-1} \cdot C(\theta[\Phi], m) \le C \cdot (1 + R)^2 \cdot \min(N^2 m^{-2s}, 1), \tag{6.1}$$

for each $m > 0$, and for some universal constant C which does not depend upon R, N, and m. Hence

$$N^{-1} \cdot \mathcal{R}(f, \Phi) \le C \cdot (1 + R)^2 \cdot \inf_m \left(\min(N^2 m^{-2s}, 1) + N^{-1} m \right). \tag{6.2}$$

Pick $m^* = \lfloor (1 + R)^{2/(2s+1)} \cdot N^{3/(2s+1)} \rfloor$. Evaluating the right-hand side of (6.2) shows that for this frame, the risk proxy (5.5)–(5.7) obeys

$$N^{-1}\mathcal{R}(f, \mathcal{L}) \le N^{-1}\mathcal{R}(f, \Phi) \le C \cdot (1 + R)^{2/(2s+1)} \cdot N^{-\frac{2(s-1)}{2s+1}}.$$

Consider $\mathcal{L}_{\text{Chirplets}}$. Then $M_N = O(N^{4/3})$ so that $\Lambda_N = 8\lambda^2/3 \cdot \log(N)(1 + o(1))$, and the oracle inequality (5.8) proves the second part of the Theorem. (For $\mathcal{L}^{\mathbb{O}}_{\text{Chirplets}}$, $M_N = O(N^2)$, and the conclusion is identical.)

The first part of the Theorem is essentially a consequence of (5.13) together with (6.1). Indeed, observe that (6.1) says that for $f(t) = A(t)e^{\mathbf{i}N\varphi(t)}$, $f \in \text{Chirp}(s;\, N, R)$, there is $\Phi \in \mathcal{L}^{+}_{\text{Chirplets}}$ such that for each m, there is an m-term linear combination $f_m[\Phi]$ of elements of Φ with the property

$$N^{-1} \cdot \|f - f_m[\Phi]\|^2 \le C \cdot (1 + R)^2 \cdot N^2 m^{-2s}$$

and hence

$$N^{-1} \cdot \inf_{\tilde{f}} K(f, \tilde{f}) \le C \cdot \inf_m \left((1 + R)^2 \cdot N^2 m^{-2s} + N^{-1} m \right) \le C \cdot (1 + R)^{2/(2s+1)} \cdot N^{-\frac{2(s-1)}{2s+1}}.$$

Note that for $f(t) = A(t) \cos(N\varphi(t))$, the same basic estimates would apply since we could take $(f_m[\Phi] + \overline{f[\Phi]}_m)/2$ (with f_m as before) as a linear combination of at most $2m$ elements taken from Φ. The first part of the Theorem then follows from (5.13). The theorem is proved —provided of course that one verifies the claim (6.1). ∎

6.2 Ideal Risk Calculations

Recall our assumptions which say that $f \in \textsc{Chirp}(s; N, R)$, if f is of the form $f(t) = A(t)\cos(N\varphi(t))$ where the phase φ and amplitude A belong to $\textsc{Hölder}^s(R)$. In the remainder of this paper, we will denote by $\|\cdot\|_s$ the homogeneous Hölder norm

$$\|g\|_s = \sup_{t,t'} \frac{|g^{(m)}(t) - g^{(m)}(t')|}{|t-t'|^{s-m}}, \quad m < s \le m+1.$$

To approximate $f \in \textsc{Chirp}(s; N, R)$, we select j such $2^j \le N^{1/s} < 2^{j+1}$, i.e. $2^{-j} \sim N^{-1/s}$, and consider the uniform dyadic partition $\mathcal{P}_j$ at that scale; that is the collection of intervals of the form $I = [k2^{-j}, (k+1)2^{-j})$, $k = 0, 1, \ldots, 2^j - 1$. Associated with $\mathcal{P}_j$ is a collection of windows we shall simply denote by w_I. In the sequel, $\tilde{I}$ will stand for the support of w_I. Note that $|\tilde{I}| \le 2^{-j+1}$.

Lemma 6.1 *Assume that b_I is discretized as in section 2, namely, $b_I = \pi \cdot \ell \cdot 2^{2j}/N$, $\ell \in \mathbb{Z}$, $|b_I| \le B$, where $B \le R$, say. Then, there is a continuous broken line $(a_I + b_I t)_{I \in \mathcal{P}_j}$ obeying the following two properties:*

(i) for each dyadic interval I,

$$\sup_{\tilde{I}} |\varphi''(t) - b_I| \le (2R + \pi) \cdot 2^{-j(s-2)}; \tag{6.3}$$

(ii) and for each dyadic interval I,

$$\sup_{\tilde{I}} |\varphi'(t) - (a_I + b_I t)| \le (R + \pi) \cdot 2^{-j(s-1)}. \tag{6.4}$$

In the above, a_I may be taken of the form $\pi \cdot \ell \cdot 2^j/N$ for some interval dependent integer ℓ.

Proof of Lemma. Both (6.3) and (6.4) follow from the assumptions about the phase and the fact that the collection of "phaselets" $(a_I + b_I t)_I$ is appropriately discretized. We let t_I, t'_I, be the dyadic points $k/2^{-j}, (k+1)2^{-j}$. For each $k = 0, \ldots 2^j - 1$, we define (t_I, φ'_I) such that φ'_I is the point on the lattice $\ell \cdot \pi 2^j/N$, $\ell \in \mathbb{Z}$, closest to $\varphi'(t_I)$; the broken line $a_I + b_I t$ is the piecewise linear and continuous function which goes through the points (t_I, φ'_I). This approximation obeys (6.3) and (6.4).

To see why this is true, let g be a C^α function, i.e. such that $\|g\|_\alpha$ is finite for some $1 \le \alpha \le 2$, and such that $g(0) = g(1) = 0$. Then

$$\sup_{[-1/2,3/2]} |g'(t)| \le 2\|g\|_\alpha, \quad \text{and} \quad \sup_{[-1/2,3/2]} |g(t)| \le \|g\|_\alpha. \tag{6.5}$$

Now for each I, let

$$\delta_I(t) = \varphi'(t_I) + 2^j \left(\varphi'(t'_I) - \varphi'(t_I)\right)(t - t_I).$$

Following (6.5), a simple rescaling argument gives

$$\sup_{\tilde{I}} |\varphi''(t) - \delta'_I(t)| \le 2^{-j(s-2)} \cdot 2\|\varphi\|_s, \quad \text{and} \quad \sup_{\tilde{I}} |\varphi'(t) - \delta_I(t)| \le 2^{-j(s-1)} \cdot \|\varphi\|_s.$$

Put $\omega_I(t) = a_I + b_I t = \varphi'_I + b_I(t - t_I)$. At scale 2^{-j}, the spacing Δ between two consecutive intercepts obeys $\Delta \le \pi \cdot 2^j/N$ and hence,

$$|\varphi'(t_I) - \omega_I(t_I)| \le \pi \cdot 2^j/2N, \quad |\varphi'(t'_I) - \omega_I(t'_I)| \le \pi \cdot 2^j/2N.$$

It then follows that

$$|\delta_I(t) - \omega_I(t)| \le \pi \cdot 2^j/N, \quad \forall t \in \tilde{I},$$

and the triangle inequality then gives

$$|\varphi'(t) - \omega_I(t)| \le 2^{-j(s-1)} \cdot \|\varphi\|_s + \pi \cdot 2^j/N.$$

Likewise, $|\delta'_I(t) - \omega'_I(t)|$ is less or equal to $\pi \cdot 2^{2j}/N$ and, therefore,

$$|\varphi''(t) - \omega'_I(t)| \le 2^{-j(s-2)} \cdot 2\|\varphi\|_s + \pi \cdot 2^{2j}/N.$$

The last two displays prove (6.3) and (6.3), since by assumption $N \ge 2^{js}$, and $\|\varphi\|_s \le R$. ∎

In the remainder of this section, we will let $\Phi \in \mathcal{L}_{\text{CHIRPLETS}}$ be the tight-frame implied by Lemma 6.1; that is, that frame with chirping rate b_I over the interval I. Put h_I to be the phase difference

$$h_I(t) = \varphi(t) - a_I t - \frac{1}{2} b_I t^2.$$

It follows from Lemma 6.1 that $\|h_I\|_s \le R$, and

$$\sup_{\tilde{I}} |h'_I(t)| \le (R + \pi) \cdot 2^{-j(s-1)}, \qquad \sup_I |h''_I(t)| \le (2R + \pi) \cdot 2^{-j(s-2)}. \tag{6.6}$$

Put $n_I = N \cdot 2^{-j} \cdot a_I/\pi$ and note that our discretization implies $n_I \in \mathbb{Z}$. With these notations and for each dyadic interval, the coefficients of f in Φ are given by

$$\theta_{I,n} = 2^{(j-1)/2} \int A(t) e^{\mathbf{i} N h_I(t)} w_I(t)\, e^{-\mathbf{i}\pi(n-n_I)2^j t}\, dt. \tag{6.7}$$

Lemma 6.2 *Let I be a fixed interval. The sequence $(\theta_{I,n})_n$ obeys*

$$\sum_{n:\, |n-n_I| > M} |\theta_{I,n}|^2 \le C \cdot (R+1)^2 \cdot 2^{-j} \cdot M^{-2s}, \tag{6.8}$$

for some constant $C > 0$.

Remark. The previous calculations assumed we were working with a tight-frame $\Phi \in \mathcal{L}_{\text{CHIRPLETS}}$. Suppose instead that Φ is an orthobasis in $\mathcal{L}^{\mathbb{O}}_{\text{CHIRPLETS}}$, then

$$\theta_{I,n} = 2^{(j+1)/2} \int A(t) e^{\mathbf{i} N h_I(t)} w_I(t) \sin\left[\pi(n+1/2)(2^j t - k)\right] dt, \quad n = 0, 1, \ldots, \tag{6.9}$$

and the statement (6.8) of Lemma would need to be replaced with

$$\sum_{n > M} |\theta_{I,n}|^2 \le C \cdot (R+1)^2 \cdot 2^{-j} \cdot M^{-2s},$$

whose proof consists in a minor modification of that of Lemma 6.2. We postpone the proof of this lemma to a later section and state a consequence of (6.8).

Theorem 6.3 *Suppose that $f \in \mathcal{F}_s$ and let Φ be the tight-frame (or basis) implied in Lemma 6.1. Then the partial reconstruction obtained by keeping the m largest terms in the expansion of f obeys*

$$\|f - f_m[\Phi]\|_{L_2}^2 \leq C \cdot (R+1)^2 \cdot \min(N^2 m^{-2s}, 1). \tag{6.10}$$

Proof of Theorem. Lemma 6.2 gives

$$\sum_{I \in \mathcal{P}_j} \sum_{n:|n-n_I|>M} |\theta_{I,n}|^2 \leq C \cdot (R+1)^2 \cdot M^{-2s}.$$

Let $|\theta|_{(k)}$ be the kth largest entry in the sequence $(|\theta_{I,n}|)_{I,n}$ and set $m = 2^j \cdot M$. We then just established that

$$\sum_{k>m} |\theta|_{(k)}^2 \leq \sum_{I \in \mathcal{P}_j} \sum_{|n|>M} |\theta_{I,n}|^2 \leq C \cdot (R+1)^2 \cdot 2^{2js} \cdot m^{-2s},$$

which proves (6.10) since $2^j \leq N^{1/s} < 2^{j+1}$. ∎

6.3 Analysis of Demodulated Chirps

In this section, we will give a proof of Lemma 6.2. Suppose we have a chirp $A(t)e^{\mathbf{i}N\varphi(t)}$. We demodulate and window the chirp, multiplying by $e^{-iN(a_I+\frac{1}{2}b_I t)t} 2^{j/2} w(2^j t - k)$, where w is smooth and compactly supported. We then rescale the domain so the resulting object is supported in a unit-scale interval. An equivalent realization of this object is of the form,

$$\tilde{f}(t) = w(t)\tilde{A}(t)e^{\mathbf{i}N\delta(t)}, \tag{6.11}$$

so that the coefficient (6.9) is given by

$$\theta_{I,n} = 2^{-(j+1)/2} \int w(t)\tilde{A}(t)e^{\mathbf{i}N\delta(t)}\, e^{-\mathbf{i}\pi(n-n_I)t}\, dt.$$

For such an object to correspond to a rescaling of the type discussed above, we must impose constraints on the function $\tilde{A}$ and on the phase δ. As for A we impose

$$\|A\|_\infty \leq R, \quad \|A\|_s \leq R \cdot 2^{-js}. \tag{6.12}$$

As for δ we impose

$$2^{js} \cdot \|\delta'\|_{L_\infty} \leq (R+\pi), \quad 2^{js} \cdot \|\delta''\|_{L_\infty} \leq (R+\pi), \quad 2^{js} \cdot \|\delta\|_s \leq R. \tag{6.13}$$

Indeed this follows from (6.4) and (6.6) by a simple rescaling. In this sense, the phase δ is nearly non-oscillatory.

Set $g(t) = w(t)\tilde{A}(t)$ so that $f(t) = g(t)e^{\mathbf{i}N\delta(t)}$. One the one hand, g clearly belongs to Hölder$^s(\gamma R)$, and on the other

$$\|e^{\mathbf{i}N\delta}\|_s \leq \gamma \cdot (R+1),$$

for some positive $\gamma > 0$. The latter statement is a simple consequence of the condition (6.13). Indeed, we calculate the second derivative

$$D^2 e^{\mathbf{i}N\delta(t)} = ([\mathbf{i}N\delta'(t)]^2 + \mathbf{i}N\delta''(t))e^{\mathbf{i}N\delta(t)} = T_0(t) + T_1(t).$$

(Recall that $2^j \leq N^{1/s} < 2^{j+1}$). The first term is differentiable and obeys $|(DT_0)(t)| \leq \gamma \cdot (R+1)$ and, therefore, $|T_0(t) - T_0(t')| \leq \gamma \cdot (R+1) \cdot |t - t'|$. For the second, note that (6.13) gives $N|\delta''(t) - \delta''(t')| \leq (2R+\pi) \cdot |t-t'|^{s-2}$ and obviously, T_1 obeys $|T_1(t) - T_1(t')| \leq \gamma \cdot (R+1) \cdot |t - t'|^{s-2}$. We then conclude that $\tilde{f}$ obeys the smoothness estimate

$$\|\tilde{f}\|_s \leq \gamma \cdot (R+1). \tag{6.14}$$

For a fixed I, the coefficients $\theta_{I,n}$ are the shifted Fourier coefficient of $2^{-j/2}\tilde{f}$ where $\tilde{f} \in \text{HÖLDER}^s(\gamma \cdot (R+1))$, supported on the interval $[-1/2, 3/2]$ and obeying (6.14). With $c_n(\tilde{f})$ the Fourier coefficients of $\tilde{f}$,

$$c_n(\tilde{f}) = \frac{1}{\sqrt{2}} \int \tilde{f}(t) \exp(-\mathbf{i}\pi n t)\, dt,$$

it is well-known that (6.14) implies the decay

$$\sum_{|n|>m} |c_n(\tilde{f})|^2 \leq C \cdot (R+1)^2 \cdot m^{-2s} \cdot \|\tilde{f}\|_s^2.$$

The claim (6.8) then follows from $\theta_{I,n} = 2^{-j/2} c_{n-n_I}(\tilde{f})$. Lemma 6.2 is established. ∎

6.4 Accuracy of Trapezoidal Quadrature Rules

The reader may have been surprised by the argument underlying the proof of Lemma 6.2. Indeed, the expression for the chirplet coefficients were given by integrals and thus, assumed signals on the continuum. Because our model is discrete, this is a distortion of reality. Before deriving a version of Theorem 6.3 for sampled signals, we would like to point out that we are guilty of a common "crime" as many papers in the literature of statistics claim results for discrete models while the proofs operate at the level of the continuum, see the discussion in [26] and references therein.

With the notations of the previous section, we considered

$$\theta_{I,n} = \frac{2^{-j/2}}{\sqrt{2}} \cdot \int \tilde{f}(t) e^{-\mathbf{i}\pi n t}\, dt,$$

while we should have considered, instead,

$$\theta^D_{I,n} = \frac{2^{-j/2}}{\sqrt{2}} \cdot \frac{1}{N_j} \sum_t \tilde{f}(t/N_j) e^{-\mathbf{i}\pi n t/N_j}.$$

The point here is that the discrete coefficients are very accurate approximations of the underlying continuous coefficients. This is a well-known phenomenon in numerical analysis which is briefly explained below.

Consider for example a periodic function g on the interval $[0,1]$. We let I denote the integral $I = \int_0^1 g(t)\,dt$ and let I_N be the discrete approximation given by the trapezoidal rule

$$I_N = \frac{1}{N}\sum_{t=0}^{N-1} g(t/N).$$

Then a classical result [2] states that the accuracy of the trapezoidal rule obeys

$$|I - I_N| \le C \cdot \|g\|_s \cdot N^{-s}, \tag{6.15}$$

for some uniform constant C. Now, let $(\hat{g}_n)$, $-N/2 \le n < N/2$, be the discrete Fourier transform of the sampled values of g

$$\hat{g}_n = \frac{1}{N}\sum_{t=0}^{N-1} g(t/N) e^{\mathbf{i}2\pi nt/N}.$$

Then the accuracy estimate (6.15) generalizes to Fourier coefficients (observe that for $n = 0$, $\hat{g}_0 = I_N$) and there is a constant C such that for all $-N/2 \le n < N/2$

$$|\hat{g}_n - \hat{g}(2\pi n)| \le C \cdot N^{-s}. \tag{6.16}$$

In our setup, note that $\tilde{f}$ is zero at the endpoint of the interval of interest (and hence periodic). We then apply (6.16) and obtain

$$\theta^D_{I,n} = \theta_{I,n} + \epsilon_{I,n}, \qquad |\epsilon_{I,n}| \le C(R) \cdot 2^{-j/2} \cdot N_j^{-2s},$$

where here and below, the constant $C(R)$ may be taken to be of the form $C \cdot (R+1)$ for some C whose value may change from line to line. It then follows from

$$\sum_I \sum_{-N_j/2 \le n < N_j/2,\ |n|>M} |\theta^D_{I,n} - \theta_{I,n}|^2 \le \sum_I \sum_{|n|>M} |\epsilon_{I,n}|^2 \le C(R)^2 \cdot N_j^{-(2s-1)}$$

together with (6.8) that

$$\sum_I \sum_{-N_j/2 \le n < N_j/2,\ |n|>M} |\theta^D_{I,n}|^2 \le C(R)^2 \cdot (M^{-2s} + N_j^{-(2s-1)}).$$

Recall that N_j is the number of sampled values in an interval of length $2^{-j} \sim N^{-1/s}$, i.e. $N_j \sim N^{(s-1)/s}$. Hence, following the proof of Theorem 6.3, one obtains that the tail of θ^D obeys

$$\sum_{k>m} |\theta^D|^2_{(k)} \le C(R)^2 \cdot \left(2^{2js} \cdot m^{-2s} + N^{-\alpha}\right), \quad \alpha = (2s-1)(s-1)/s.$$

We then conclude by observing that for our discrete model, the previous analysis gives

$$\frac{1}{N} \cdot \mathcal{R}(f, \Phi) \le C(R)^2 \cdot \inf_m \left(\min(N^2 m^{-2s}, 1) + N^{-1}m\right) + C(R)^2 \cdot N^{-\alpha}. \tag{6.17}$$

Since on the one hand we calculated earlier that $\inf_m \min(N^2 m^{-2s}, 1) + N^{-1}m \sim N^{-\frac{2(s-1)}{2s+1}}$, and on the other $\alpha > \frac{2(s-1)}{2s+1}$ for $2 \le s \le 3$, we see that the discretization error is negligible as compared to the minimax risk (1.8). Hence, our discrete algorithm provably attains the optimal estimation bounds.

6.5 Proof of Theorem 3.5

This theorem relies on the following lemma.

Lemma 6.4 *Let $f(t) = A(t)\cos(N\varphi(t))$ be a chirp in* $\textsc{Chirp}(s; N, R)$ *with $A(0) = A(1) = 0$ or with A and φ periodic. Put F to be the analytic part of f as in (3.16). Then under the assumptions of Theorem 3.5*

$$F(t) = A(t)e^{\mathbf{i}N\varphi(t)} + r(t),$$

where

$$\|r\|^2 \leq B \cdot N^{-\frac{2(s-1)}{2s+1}}. \tag{6.18}$$

The proof of (6.18) involves a delicate estimate whose proof in the Appendix. We first explain, however, how Theorem 3.5 follows from this lemma.

Put $F_{\text{IDEAL}}(t) = A(t)e^{\mathbf{i}\lambda\varphi(t)}$. Since $f(t) = \Re(F(t))$, our estimator obeys

$$\|\hat{f} - f\|^2 \leq \|\hat{F} - F\|^2.$$

We then use the risk proxy (5.5) to bound the MSE $E\|\hat{F} - F\|^2$. (Because the noise is now lower-dimensional, we could actually derive an oracle inequality with better bounds than (5.8).) Letting $\theta(F)$ be the coefficients of F in the tight-frame Φ, the proxy obeys

$$\begin{aligned}
\sum_i \min\left(|\theta_i(F_{\text{IDEAL}}) + \theta_i(r)|^2, 1\right) &\leq \sum_i \min(2|\theta_i(F_{\text{IDEAL}})|^2, 1) + \min(2|\theta_i(r)|^2, 1) \\
&\leq \sum_i \min(2|\theta_i(F_{\text{IDEAL}})|^2, 1) + 2\sum_i |\theta_i(r)|^2 \\
&\leq 2\sum_i \min(|\theta_i(F_{\text{IDEAL}})|^2, 1) + 2\|r\|^2.
\end{aligned}$$

To conclude, we simply observe that an earlier section proved that

$$\sum_i \min(|\theta_i(F_{\text{IDEAL}})|^2, 1) \leq C \cdot (R^2 + 1) \cdot N^{-\frac{2(s-1)}{2s+1}},$$

in some frame Φ, while Lemma 6.4 gives a bound on the second term which is of the same order, namely $\|r\|^2 = O\left(N^{-\frac{2(s-1)}{2s+1}}\right)$. The oracle inequality allows to conclude that

$$E\|\hat{f} - f\|^2 \leq E\|\hat{F} - F\|^2 \leq O(\log N) \cdot N^{-\frac{2(s-1)}{2s+1}},$$

which establishes the theorem. ∎

We now turn our attention to Lemma 6.4. Let $c_n(g)$ be the Fourier coefficients of an object $g \in L_2[0,1]$

$$c_n(g) = \int g(t)\, e^{-\mathbf{i}\pi nt}\, dt.$$

The coefficients of r are then simply given by

$$c_n(r) = \begin{cases} \overline{c_{-n}(F_{\mathrm{IDEAL}})} & n > 0 \\ -i\Im\left(c_0(F_{\mathrm{IDEAL}})\right) & n = 0 \\ -c_n(F_{\mathrm{IDEAL}})) & n < 0 \end{cases}.$$

This follows from the definition. Indeed, for $n \geq 0$, $2f = F_{\mathrm{IDEAL}} + \overline{F_{\mathrm{IDEAL}}}$ which gives $2c_n(f) = c_n(F_{\mathrm{IDEAL}}) + \overline{c_n(F_{\mathrm{IDEAL}})}$ and, therefore, $c_n(r) = 2c_n(f) - c_n(F_{\mathrm{IDEAL}}) = \overline{c_{-n}(F_{\mathrm{IDEAL}})}$. The claim for $n = 0$ and $n < 0$ are similar. Hence,

$$\|r\|^2 = \sum_{n \leq 0} |c_n(F_{\mathrm{IDEAL}})|^2,$$

and we then need to verify that the right-hand side is appropriately bounded.

Suppose now that A and φ are members of the class $\mathrm{H\ddot{O}LDER}^s(R)$, and obey the assumptions of Lemma 6.4. Assume without loss of generality that $\varphi' \geq 0$. Set Ω such that $\Omega \leq N\varphi'(t)$ and suppose $\Omega \geq N^{1/2}$. Then for any $\delta > 0$, there is a constant C_δ such that

$$\sum_{n<0} |c_n(F_{\mathrm{IDEAL}})|^2 \leq C_\delta \cdot \Omega^\delta \cdot \left(\frac{N}{\Omega^2}\right)^{2(s-1)}. \tag{6.19}$$

This inequality is the difficult part of the argument and the proof may be found in the Appendix. We now specialize this inequality by selecting $\Omega = N^{1/2+\beta/2}$ with $\beta > 1/(2s+1)$ as in Theorem 3.5. This gives

$$\begin{aligned} \sum_{n<0} |c_n(F_{\mathrm{IDEAL}})|^2 &\leq C_\delta \cdot N^{\delta(1+\beta)/2} \cdot N^{-2(s-1)\beta} \\ &= C_\delta \cdot N^{-2(s-1)(\beta-\delta^*)}, \quad \delta^* = \delta(1+\beta)/(4(s-1)). \end{aligned}$$

Picking δ small enough so that $\delta^* < \beta - 1/(2s+1)$ gives (6.18).

7 Lower Bounds

In this section, we provide a proof of Theorem 3.1 and establish lower bounds of estimation for $\mathcal{F} = \mathrm{CHIRP}(s, N, R)$, with $s \geq 2$. We assume that $\lambda = N$ as the proofs for arbitrary values of λ are identical. We follow a standard strategy for establishing lower bounds and construct a family of hypercubes $\mathcal{H}_m$ which are embedded in $\mathcal{F}$, see [16, 14, 5] and perhaps [28]. We then study the subproblem of estimation when the object comes from one of the vertices of the hypercube. Suitable choices of hypercubes then allow to derive sharp lower bounds on the minimax mean-squared error of this subproblem. We begin by considering the case of real signals, and later explain why the complex problem is not easier.

7.1 Hypercubes

We let φ be a smooth function supported in the interval $[0, 1]$ and obeying $\|\varphi\|_s \leq R$. We then define 'phaselets' by the rule

$$\varphi_{i,m}(t) = m^{-s}\varphi(m(t - t_i)), \quad t_i = i/m,\ i = 0, 1, \ldots, m-1. \tag{7.1}$$

Of course, each phaselet also obeys $\|\varphi_{i,m}\|_s \leq R$ and for $t \in [0,1]$, we then consider superpositions of such phaselets

$$\Phi_\xi(t) = \sum_{i=0}^{m-1} \xi_i \varphi_{i,m}(t), \quad \xi \in \{0,1\}.$$

We will use these linear combinations to synthesize a collection of chirps as follows:

$$\mathcal{H}_m = \{f(t),\ f(t) = \sin\left(N\Phi_\xi(t)\right), t \in [0,1]\}. \tag{7.2}$$

Note that the embedding $\mathcal{H}_m \subset \mathcal{F}$ is obviously a consequence of the membership of Φ_ξ to $\text{HÖLDER}^s(R)$. Moreover, the set $\mathcal{H}_m$ is indeed a hypercube. Indeed, observe that for any t, there is at most one term in the sum $\Phi_\xi(t) = \sum_i \xi_i \varphi_{i,m}(t)$ which is nonzero and that we can thus rewrite $f \in \mathcal{H}_m$ as

$$f(t) = \sum_{i=0}^{m-1} \xi_i \, \sin(N\varphi_{i,m}(t)) := \sum_{i=0}^{m-1} \xi_i \, a_{i,m}(t)$$

with

$$a_{i,m}(t) = \sin(N\varphi_{i,m}(t)).$$

The vertices $a_{i,m}$ of the set $\mathcal{H}_m$ are 'little chirps' which are orthogonal because of their disjoint support. Hence, we are entitled to think of $\mathcal{H}_m$ as a functional hypercube in L_2.

7.2 Lower Bound Calculations

When restricted to the set $\mathcal{H}_m$, the estimation problem (1.1) becomes simply the problem of determining which of the many vertices of the hypercube may have generated the observed data. First, letting P be the orthogonal projection onto the span of the $a_{i,m}$'s, we have

$$\|P\hat{f} - f\|_{L_2}^2 = \|P\hat{f} - Pf\|_{L_2}^2 \leq \|\hat{f} - f\|_{L_2}^2,$$

and, therefore, we may restrict attention to estimators which are linear combinations of $a_{i,m}$'s, i.e. of the form

$$\hat{f} = \sum_i \hat{\xi}_i a_{i,m}.$$

Second, owing to the orthogonality of the chirps $a_{i,m}$, we have that for these estimators

$$\|\hat{f} - f\|_{L^2}^2 = \|\hat{\xi} - \xi\|_{\ell^2}^2 \cdot \|a_{i,m}\|_{L_2}^2, \tag{7.3}$$

and so the problem reduces to one of estimating ξ.

Next, define the statistic $X = (X_1, \ldots, X_m)$ by $X_i = \langle Y, a_{i,m}\rangle / \|a_{i,m}\|_{L_2}^2$; that is, Y_i is the inner product between the data Y and the renormalized blobs $a_{i,m}$. In vector notations, $X \sim N(\xi, \sigma_m^2 \cdot I)$, where

$$1/\sigma_m^2 = N \cdot \|a_{i,m}\|_{L_2}^2.$$

This construction is valid for any m and we now select m such that the 'noise' level σ_m is approximately equal to the size of the coordinates to be estimated, i.e. $\sigma_m \sim 1$. Formally, we choose a noise-dependent $m(N)$ such that

$$m(N) = \inf \{m, \ \sigma_m \geq 1\}. \tag{7.4}$$

To specify the value of $m(N)$, we now develop a result about the size of the elements $a_{i,m}$.

Lemma 7.1 *The norm of $a_{i,m}$ obeys the following size estimate:*

$$\|a_{i,m}\|^2 \geq A \cdot m^{-1} \cdot \min(N^2 m^{-2s}, 1) \tag{7.5}$$

for some fixed positive constant A.

Here the norm is either the continuous L_2-norm $\|f\|^2 = \int_0^1 |f(t)|^2 \, dt$ or the analogous discrete Euclidean norm $\|f\|^2 = N^{-1} \sum_{t=0}^{N-1} |f(t/N)|^2$. The proof of this lemma consists of a simple calculation and is omitted.

Hence, selecting $m(N) \geq N^{1/s}$ as

$$m(N) \sim N^{3/(2s+1)}$$

gives $\sigma_m \geq 1$ as in (7.4). Now for $\sigma_m \geq 1$, it is intuitively obvious that the minimax mean-squared error in estimating the vector ξ from the observations $X \sim N(\xi, \sigma_m^2 \cdot I)$ is bounded below by by a constant times the number of coordinates m. Formally,

$$\inf_{\hat{\xi}} \sup_{\xi} E\|\hat{\xi} - \xi\|_{\ell^2}^2 \geq B \cdot m. \tag{7.6}$$

This last inequality is indeed classical. In a nutshell, note that our problem is harder than that of estimating the mean vector ξ from the data $X = N(\xi, I)$. For this latter problem, consider the prior π where the ξ_i's are i.i.d. with $P(\xi_i = 1) = 1/2$. Then the minimax mean-squared error is bounded below by the Bayes risk which for this specific prior is of the form $B \cdot m$—with B the Bayes risk of estimating any coordinate $\xi_i \in \{0, 1\}$ from $X_i \sim N(\xi, 1)$ and the one-dimensional prior $P(\xi_i = 1) = 1/2$. To see why this is true, observe that the Bayes estimate is of the form

$$\hat{\xi}_i = E(\xi_i | X) = E(\xi_i | X_i)$$

and, therefore, the Bayes estimate obeys

$$E\|\hat{\xi} - \xi\|^2 = \sum_i E(\hat{\xi}_i - \xi_i) = m \cdot E(\hat{\xi}_i - \xi_i) = B \cdot m.$$

Specializing (7.6) to the choice $m(N)$ of (7.4) and invoking the isometry (7.3) gives Theorem 3.1.

7.3 Complex Signals

Suppose now that we wish to recover a complex signal $f(t) = A(t)\exp(\mathbf{i}N\varphi(t))$ from noisy data as in (3.10) or (3.11). It turns out that this problem is about just as hard as the corresponding real problem. With the same phase function as before, we consider the embedded hypercube

$$\mathcal{H}_m = \{f(t),\ f(t) = \exp\left(\mathbf{i}N\Phi_\xi(t)\right), t \in [0,1]\}; \tag{7.7}$$

$\mathcal{H}_m$ is a hypercube since we can rewrite a generic element $f \in \mathcal{H}_m$ as

$$f(t) = 1 + \sum_{i=0}^{m-1} \xi_i\,(e^{\mathbf{i}N\varphi_{i,m}(t)} - 1) := 1 + \sum_{i=0}^{m-1} \xi_i\, u_{i,m}(t)$$

with

$$u_{i,m}(t) = \left(e^{\mathbf{i}N\varphi_{i,m}(t)} - 1\right) 1_{[i/m,(i+1)/m)}(t).$$

The vertices $u_{i,m}$ of the set $\mathcal{H}_m$ are orthogonal since they have disjoint supports. For $t \in [i/m, (i+1)/m)$, rewrite $u_{i,m}$ as

$$u_{i,m}(t) = (\cos(N\varphi_{i,m}(t)) - 1) + \mathbf{i}\sin(N\varphi_{i,m}(t)) = b_{i,m}(t) + \mathbf{i}a_{i,m}(t)$$

($a_{i,m}(t)$ is as before). Our problem then consists in recovering the bivariate signal $(\sum_i \xi_i a_{i,m}, \sum_i \xi_i b_{i,m})$ from the noisy data

$$\begin{aligned} Y_1[t] &= \sum_i \xi_i a_{i,m}[t] + z_1[t] \\ Y_2[t] &= \sum_i \xi_i b_{i,m}[t] + z_2[t]. \end{aligned}$$

Following our earlier argument, consider the random vector $X_{i,j} = \langle Y_j, a_{i,m}\rangle/\|a_{i,m}\|^2$, $j = 1, 2$,

$$\begin{aligned} X_{i,1} &= \xi_i + \sigma_m\, z_{1,i} \\ X_{i,2} &= \epsilon_m \cdot \xi_i + \sigma_m\, z_{2,i} \end{aligned}. \tag{7.8}$$

Here $\epsilon_m = \langle b_{i,m}, a_{i,m}\rangle/\|a_{i,m}\|^2$. Because $\|b_{i,m}\|$ is much smaller than $\|a_{i,m}\|$ over the range of m's of interest, ϵ_m is very small; selecting m as before gives

$$|\epsilon_m| \le \|b_{i,m}\|/\|a_{i,m}\| \le C \cdot N^{-(s-1)/(2s+1)}.$$

Note that our problem is obviously harder than just recovering the first component of $\sum_i \xi_i a_{i,m}$ (or equivalently recovering the vector ξ) from the above data which is nearly the problem we studied in the previous section; the only difference is that we have additional data about ξ_i in the form of the second component of (7.8). However, since we argued that ϵ_m is negligible, these additional data do not make the problem any easier; in other words, the vector X_2 contains practically no information about ξ.

Formally, we proceed as before and observe that our our problem is harder than that of estimating the mean vector (ξ_i) from the independent observations $X_i = N((\xi_i, \epsilon_m \xi_i), I)$. For this latter problem, consider the prior π where the ξ_i's are i.i.d. with $P(\xi_i = 1) = 1/2$. Then the Bayes estimate $\hat{\xi}_i$ is again of the form $\hat{\xi}_i = E(\xi_i | X_i)$ and obeys $E(\hat{\xi}_i - \xi_i) = B'$ which gives

$$E\|\hat{\xi} - \xi\|^2 = B' \cdot m;$$

in fact because $\epsilon_m \to 0$ as $N \to \infty$, $B' = B(1 + o(1))$ where the second term goes to zero as N tends to infinity. The rest of the argument is as before.

8 Discussion

The main contribution of this paper is merely to establish a bridge between between concepts in time-frequency analysis and statistical estimation by showing how well one can hope to recover a class of rapidly oscillating signals. Next, building upon ideas Computational Harmonic Analysis (CHA), we have been able to derive reasonable, flexible, and low-complexity algorithms which come close to the best theoretical performances. Encouraged by the excellent track-record of CHA for applications, we look forward to testing the potential of these ideas for practical significance. In this direction, the author has recently engaged in the development of computational tools and hope to report on numerical experiments shortly.

8.1 Further Issues

The methodology presented in this paper relies on the assumption that the signal we wish to recover is a single chirp, i.e. a signal of the form $f(t) = A(t)\cos(\lambda\varphi(t))$, and would be ineffective in situations where the unknown signal is actually a superposition of a finite number of chirps, with distinct amplitudes and phases, e.g. $f(t) = \sum_{i=1}^m A_i(t)\cos(\lambda\varphi_i(t))$. It would certainly be of importance to extend and develop other ideas to adapt to such signals effectively.

It might very well be possible to find algorithms with lower computational complexity than those we discussed in this paper. In particular, we may not need to compute all chirplet coefficients as one might imagine quickly scanning the time-frequency plane to identify regions in the plane or subsets of the chirplet parameter space in order to narrow the search. A relevant problem in this line of research would concern the design of algorithms with minimum computational complexity, and which would still arguably produce estimators with nearly optimal asymptotic properties.

Suppose we have noisy data $Y(t) = \alpha f(t) + z(t)$, $t = 0, 1, \ldots, N$ where α is unknown and f is chirping. Then a problem of interest is to detect whether $\alpha = 0$. Work in progress also develops a statistical theory for problems of this kind. Assume for example that the chirp has unit-energy $\|f\| = 1$ and Gaussian white noise with noise level $E\|z\|^2 = 1$. Then preliminary calculations suggest the existence of constants (α_N^0, α_N^1) such that no procedure can detect if $\alpha < \alpha_N^0$ and that one can detect with asymptotically full power if $\alpha \geq \alpha_N^1$ with again low complexity algorithms.

8.2 Challenge

In some sense, there is a higher-level question that we have not addressed here: Is there an optimal representation for chirps? That is, is there a fixed basis or tight-frame in which thresholding would be asymptotically nearly optimal over our classes of chirps. To our knowledge, no pre-existing basis or tight-frames would yield convergence rates that could come even close to the optimal convergence rates we identified in this paper. In fact, the best published nonadaptive result is given by the thresholding of Gabor type of expansions which gives markedly suboptimal convergence rates, as we have seen. Hence, we conclude with this important challenge: *Is there a nonadaptive representation which provides optimally sparse decompositions of chirps?* In other words, is there something beyond Gabor?

9 Appendix

In the appendix, we study properties of analytic chirps and prove (6.19).

Theorem 9.1 *Let* $f(t) = A(t)\exp(\mathbf{i}2\pi N\varphi(t))$ *be a chirp such that* $A, \varphi \in \text{HÖLDER}^s(R)$ *with* $A(0) = A(1) = 0$ *or with* A *and* φ *periodic. Suppose that* $2\pi \cdot N\varphi'(t) > \pi \cdot \Omega \geq \pi \cdot N^{1/2}$. *Then, for any* $\delta > 0$, *the Fourier coefficients* $c_n(f)$ *obey*

$$\sum_{n \leq 0} |c_n(f)|^2 \leq C_\delta \cdot \Omega^\delta \cdot \left(\frac{N}{\Omega^2}\right)^{2(s-1)}. \tag{9.1}$$

We prove the result for $s = 2$ (the extension to arbitrary values of s is treated similarly). An integration by parts gives

$$\begin{aligned} c_n(f) &= \int A(t) e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt}\, dt \\ &= \int e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt} \frac{d}{dt}\left(\frac{A(t)}{2\pi\mathbf{i}(N\varphi'(t) - n)}\right). \end{aligned}$$

That is,

$$c_n(f) = \frac{1}{2\pi\mathbf{i}}(a_n + b_n)$$

where

$$\begin{aligned} a_n &= \int e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt} \frac{A'(t)}{(N\varphi'(t) - n)}\, dt \\ b_n &= \int e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt} \frac{NA(t)\varphi''(t)}{(N\varphi'(t) - n)^2}\, dt. \end{aligned}$$

It then suffices to prove that both sequences obey (9.1). We begin with the difficult estimate, namely, b_n. From now on, g will denote the function

$$g(t) = e^{\mathbf{i}2\pi N\varphi(t)}\, NA(t)\varphi''(t)$$

so that $b_n = \int e^{-\mathbf{i}2\pi nt}\, g(t)\,(N\varphi'(t) - n)^{-2}\, dt$.

We introduce block of indices and say that $n \in B_\ell$ if

$$-n \in [\ell \cdot \Omega, (\ell + 1) \cdot \Omega), \quad \ell = 0, 1, \ldots.$$

We may think of this partitioning as a kind of Littlewood-Paley decomposition with *a further decomposition within each dyadic block*. For $n \in B_\ell$, set

$$n = n_\ell + n', \quad n_\ell = -\ell \cdot \Omega.$$

Now recall that for $x < 1$

$$\frac{1}{(1+x)^2} = \sum_{k=0}^{K-1} (k+1)x^k + O(x^K).$$

Suppose that $n \in B_\ell$, then

$$(N\varphi'(t) - n)^{-2} = (N\varphi'(t) - n_\ell)^{-2} \left(\sum_{k=0}^{K-1} \frac{(k+1)(n')^k}{(N\varphi'(t) - n_\ell)^k} + h_K(t) \right), \quad h_K(t) = O(\Omega^{-\alpha k}).$$

The last assertion comes from the fact that for $n \in B_\ell$, $N\varphi'(t) - n_\ell \geq \Omega \cdot (1 + \ell \cdot \Omega^{-\alpha})$ together with $|n'| \leq \cdot \Omega^{(1-\alpha)}$.

With these notations

$$b_n = \sum_{k=0}^{K-1} b_{k,n} + b_{K,n}$$

where for $k = 0, 1, \ldots, K-1$,

$$b_{k,n} = (k+1)(n')^k \int e^{-\mathbf{i}2\pi nt}\, g(t)\, (N\varphi'(t) - n_\ell)^{-(2+k)}\, dt.$$

and

$$b_{K,n} = \int e^{-\mathbf{i}2\pi nt}\, g(t)\, h_K(t)\, (N\varphi'(t) - n_\ell)^{-2}\, dt.$$

We set $K(\alpha)$ to be the smallest integer such that $\alpha \cdot K > 1/2$. We show that each sequence $b_{k,n}$ obeys

$$\sum_{n \leq 0} |b_{k,n}|^2 \leq C_\alpha \cdot \Omega^\alpha \cdot N^2 \cdot \Omega^{-4}. \tag{9.2}$$

Since K is finite, this gives

$$\sum_{n \leq 0} |b_n|^2 \leq C'_\alpha \cdot \Omega^\alpha \cdot N^2 \cdot \Omega^{-4},$$

which is what we are seeking to establish.

Begin with $b_{K,n}$. The function h_K obeys $h_K(t) \le C_\alpha \cdot \Omega^{-\alpha K} \le C_\alpha \cdot \Omega^{-1/2}$ and, therefore,

$$|b_{K,n}| \le C_\alpha \cdot N \cdot \Omega^{-2} \cdot (1 + \ell \cdot \Omega^{-\alpha})^{-2} \cdot \Omega^{-1/2}.$$

Because the number of elements in B_ℓ is less or equal to $\Omega^{1-\alpha}$, this gives

$$\sum_{n \in B_\ell} |b_{K,n}|^2 \le C_\alpha \cdot \Omega^{1-\alpha} \cdot (1 + \ell \cdot \Omega^{-\alpha})^{-4} \cdot \Omega^{-1} \cdot (N^2/\Omega^4).$$

Further, from the summation

$$\sum_{\ell=0}^{\infty} (1 + \ell \cdot \Omega^{-\alpha})^{-4} \le B \cdot \Omega^{\alpha},$$

we conclude that

$$\sum_{n \le 0} |b_{K,n}|^2 \le C_\alpha \cdot N^2/\Omega^4,$$

which is actually better than (9.2).

Next, for $k = 0, 1, \ldots, K-1$, we have

$$\int \left| g(t) \, (N\varphi'(t) - n_\ell)^{-(2+k)} \right|^2 \, dt \le C_k \cdot \Omega^{-2k} (1 + \ell \cdot \Omega^{-\alpha})^{-2(2+k)} \cdot N^2/\Omega^4.$$

Therefore, the Plancherel formula gives

$$\sum_{n \in B_\ell} (1 + |n'|)^{-2k} |b_{k,n}|^2 \le C_k \cdot \Omega^{-2k} (1 + \ell \cdot \Omega^{-\alpha})^{-2(2+k)} \cdot N^2/\Omega^4.$$

Since $|n'| \le \Omega^{1-\alpha}$, this gives

$$\sum_{n \in B_\ell} |b_{k,n}|^2 \le C_k \cdot \Omega^{-2\alpha k} (1 + \ell \cdot \Omega^{-\alpha})^{-2(2+k)} \cdot N^2/\Omega^4.$$

We conclude as before and obtain

$$\sum_{n} |b_{k,n}|^2 \le C_k \cdot \Omega^{\alpha(1-2k)} \cdot N^2/\Omega^4 \le C_k \cdot \Omega^{\alpha} \cdot N^2/\Omega^4.$$

as claimed.

We need turn attention to a_n. Note that we can integrate this term by parts and obtain

$$a_n = \frac{1}{2\pi \mathbf{i}} \left(a_n^0 + a_n^1 \right)$$

where

$$\begin{aligned} a_n^0 &= \int e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt} \, \frac{A''(t)}{(N\varphi'(t) - n)^2} \, dt \\ a_n^1 &= \int e^{\mathbf{i}2\pi N\varphi(t)} e^{-\mathbf{i}2\pi nt} \, \frac{NA'(t)\varphi''(t)}{(N\varphi'(t) - n)^3} \, dt. \end{aligned}$$

The first term obeys

$$|a_n^0| \leq \int \frac{|A''(t)|}{(N\varphi'(t) - n)^2}\, dt \leq C \cdot (\Omega - n)^{-2}.$$

Therefore,

$$\sum_{n<0} |a_n^0|^2 \leq C \cdot \Omega^{-3};$$

since $\Omega \leq N$, this bound is better than N^2/Ω^4. The second term obeys

$$|a_n^1| \leq \int \frac{|NA'(t)\varphi''(t)|}{(N\varphi'(t) - n)^3}\, dt \leq C \cdot N \cdot (\Omega - n)^{-3}.$$

Therefore,

$$\sum_{n<0} |a_n^1|^2 \leq C \cdot N^2 \cdot \Omega^{-5},$$

which again is of course better than N^2/Ω^4. This finishes the proof of the Theorem. ∎

References

[1] W. G. Anderson and R. Balasubramanian. Time-frequency detection of gravitational waves. *Phys. Rev. D.*, 60, 1999.

[2] K. E. Atkinson. *An Introduction to Numerical Analysis.* Wiley, 2nd edition, 1989.

[3] L. Birgé and P. Massart. Gaussian model selection. *J. Eur. Math. Soc.*, 3:203–268, 2001.

[4] A. Bultan. A four-parameter atomic decomposition of chirplets. *IEEE Trans. Signal Proc.*, 47:731–745, 1999.

[5] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Ann. Statist.*, 30:784 –842, 2002.

[6] R. Carmona, W. Hwang, and B. Torresani. *Practical Time-Frequency Analysis.* Academic Press, San Diego, 1998.

[7] E. Chassande-Mottin and P. Flandrin. On the time-frequency detection of chirps. *Appl. Comput. Harmon. Anal.*, 6:252–281, 1999.

[8] R. R. Coifman, G. Matviyenko, and Y. Meyer. Modulated Malvar-Wilson bases. *Appl. Comput. Harmon. Anal.*, 4:58–61, 1997.

[9] R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad. Sci. Paris Sér. I Math.*, 312:259–261, 1991.

[10] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38:713–718, 1992.

[11] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[12] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.

[13] D. L. Donoho. Wedgelets: nearly-minimax estimation of edges. *Ann. Statist.*, 27:859–897, 1999.

[14] D. L. Donoho. Sparse components of images and optimal atomic decomposition. *Constr. Approx.*, 17:353–382, 2001.

[15] D. L. Donoho and I. M. Johnstone. Ideal denoising in an orthonormal basis chosen from a library of bases. *C. R. Acad. Sci. Paris Sér. I Math.*, 319:1317–1322, 1994.

[16] D. L. Donoho and I. M. Johnstone. *Empirical atomic decomposition.* Manuscript, 1995.

[17] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26:879–921, 1998.

[18] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57:301–369, 1995.

[19] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[20] S.Y. Efroimovich and M.S. Pinsker. Estimation of square-integrable [spectral] density based on a sequence of observations. *Problems of Information Transmission*, pages 182–196, 1982.

[21] P. Flandrin. *Time-Frequency/Time-Scale Analysis*, volume 10 of *Wavelet Analysis and its Applications*. Academic Press Inc., San Diego, CA, 1999.

[22] D. Gabor. Theory of communication. *J. IEE*, 93:429–457, 1946.

[23] S. Jaffard, Y. Meyer, and R. D. Ryan. *Wavelets: Tools for Science & Technology.* Society for Industrial and Applied Mathematics, Philadelphia, PA, revised edition, 2001.

[24] E.T. Jaynes. *Maximum Entropy and Bayesian Spectrum Analysis and Estimation Problems*, chapter Bayesian Spectrum and Chirp Analysis. Riedel, 1987.

[25] F. A. Jenet and T. A. Prince. Detection of variable frequency signals using a fast chirp transform. *Physical Review D*, 62:122001–+, 2000.

[26] I. M. Johnstone. Wavelets and the theory of nonparametric function estimation. Available at: `http://www-stat.stanford.edu/~imj`, 1999.

[27] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc. B.*, 59:319–351, 1997.

[28] A. P. Korostelev and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.

[29] B. Larson and N. Saito. The continuous boundary local Fourier transform. In A. F. Laine, M. A. Unser, and A. Aldroubi, editors, *Wavelets IX*, volume Proc. SPIE 4478, pages 415–426, 2001.

[30] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[31] S. Mann and S. Haykin. Adaptive 'chirplet' transform: an adaptive generalization of the wavelet transform. *Optical Engineering*, 31:1243–1256, 1992.

[32] B. J. Owen and B. S. Sathyaprakash. Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement. *Physical Review D*, 60:22002–+, 1999.

[33] M.S. Pinsker. Optimal filtering of square integrable signals in Gaussian white noise. *Problems of Information Transmission*, 16:120–133, 1980.

[34] C. J. Stone. Consistent nonparametric regression, with discussion. *Ann. Statist.*, 5:595–645, 1977.

[35] L. F. Villemoes. Adapted bases of time-frequency local cosines. *Appl. Comput. Harmon. Anal.*, 10:139–162, 2001.

[36] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA, 1990.

[37] K. G. Wilson. Renormalization group and critical phenomena II: Phase-space cell analysis of critical behavior. *Physical Review B*, 4:3184–3205, 1971.