Locally Stationary Covariance and Signal Estimation With Macrotiles

David L. Donoho, Stéphane Mallat, Rainer von Sachs, and Yann Samuelides

Abstract—A macrotile estimation algorithm is introduced to estimate the covariance of locally stationary processes. A macrotile algorithm uses a penalized method to optimize the partition of the space in orthogonal subspaces, and the estimation is computed with a projection operator. It is implemented by searching for a best basis among a dictionary of orthogonal bases and by constructing an adaptive segmentation of this basis to estimate the covariance coefficients. The macrotile algorithm provides a consistent estimation of the covariance of locally stationary processes, using a dictionary of local cosine bases. This estimation is computed with a fast algorithm.

Macrotile algorithms apply to other estimation problems such as the removal of additive noise in signals. This simpler problem is used as an intuitive guide to better understand the case of covariance estimation. Examples of removal of white noise from sounds illustrate the results.

Index Terms—Best basis, covariance estimation, cosine basis, local stationarity, noise removal, spectrum estimation.

I. INTRODUCTION

ESTIMATING the covariance of nonstationary processes from few realizations is a challenging and mostly open statistical problem. This problem often appears in signal processing where only one realization is available. For a realization of size N, the N^2 expected values of the covariance matrix must be estimated from only N data values, which is extremely difficult. In a nonparametric framework, in order to reduce the number of coefficients to be estimated, one approach is to find a sparse representation of the covariance matrix. This type of strategy has been developed in the context of noise removal, where signals are estimated by thresholding their decomposition coefficients in an orthonormal basis [1]. This produces a small risk if the original signal is well approximated by few nonzero coefficients in this basis. When one does not know which orthonormal basis produces a small estimation risk, best bases algorithms [2] search automatically for an optimized basis, given the noisy data.

The Karhunen–Loève basis, which diagonalizes the covariance matrix, provides a very sparse representation of the covariance since all nondiagonal coefficients are equal to zero.

For stationary processes, we know that the Karhunen–Loève basis is the Fourier basis, and we thus only need to estimate the diagonal coefficients that correspond to the power spectrum. For nonstationary processes, the Karhunen–Loève basis is not known in advance, and it is therefore necessary to search for a best basis which approximates this Karhunen–Loève basis. This paper introduces a *macrotile* estimation algorithm inspired from techniques recently studied in [3] and [4] to estimate the Karhunen–Loève basis among a dictionary of orthogonal bases.

A macrotile model can be interpreted as a projection subspace procedure [5], where the subspace is adaptively selected using a penalization technique on the data. Section II introduces the general principles of the macrotile algorithm. A macrotile estimation procedure can be applied to the removal of noise in signals, where it can be interpreted as a particular best basis selection procedure. It is first studied in this simpler context, and Section III gives examples for the removal of noise from sounds by searching for a best basis in a dictionary of local cosine bases.

The main result of the paper concerns the estimation of locally stationary covariance matrices with the macrotile algorithm. It is more difficult than the removal of additive white noise because the diagonalization of matrices has much more sophisticated properties than the decomposition of signals in orthogonal bases. Locally stationary processes appear in physical systems where random fluctuations are produced by a mechanism which varies slowly in time or which has few brutal transitions. A locally stationary process can thus be qualitatively defined as a process that is approximately stationary over sufficiently short time intervals of unknown size [6]. For such processes, it is shown that the Karhunen–Loève basis can be approximated by an appropriate local cosine basis [3], which is estimated by the macrotile algorithm. The paper concentrates on algorithmic and numerical aspects of macrotile estimation as opposed to the mathematical proofs of consistency, which are long and technical, and can be found in [7] and [8]. For signals or processes of size N, fast algorithms on trees are described to compute macrotile estimators with $O(N(\log N)^2)$ operations.

II. PENALIZED MACROTILE ESTIMATION

A. Macrotile Projection Estimators

To better understand the estimation of covariance matrices, this problem is related to the estimation of a signal contaminated by an additive noise. The risk of a projection estimator is computed for covariance estimation and noise removal. We write $\widehat{z}[n]$ as a random signal as opposed to a deterministic signal, which we write as $z[n]$.

Let $x[n]$ be a signal of size N, contaminated by a zero-mean white Gaussian noise $\widehat{\epsilon}[n]$:

$$\widehat{x} = x + \widehat{\epsilon} \tag{1}$$

with $\mathbf{E}\{\widehat{\epsilon}[n]\widehat{\epsilon}[m]\} = \delta[n-m]\,\sigma^2$. The signal x is unknown but considered as deterministic so that $x[n] = \mathbf{E}\{\widehat{x}[n]\}$. Given the empirical observations $\widehat{x}$, noise removal consists of finding an estimator $\tilde{x}$ of x, which yields a small risk. As often in signal estimation [18], we quantify this risk with a mean-square Euclidean norm:

$$\mathbf{E}\{\|\tilde{x} - x\|^2\} = \sum_{n=1}^{N} \mathbf{E}\{|\tilde{x}[n] - x[n]|^2\}.$$

The estimation of the covariance operator Γ of a zero-mean Gaussian random vector $\widehat{y}[n]$ of size N can be casted as a similar estimation problem. This operator is characterized by a matrix of N^2 covariance coefficients

$$\gamma[n,m] = \mathbf{E}\{\widehat{y}[n]\,\widehat{y}^*[m]\} \quad \text{for } 1 \le n, m \le N. \tag{2}$$

Let $\widehat{\Gamma}$ be the operator whose matrix is composed of sample-mean estimators computed from R independent realizations $\{\widehat{y}_r\}_{1 \le r \le R}$ of $\widehat{y}$:

$$\widehat{\gamma}[n,m] = \frac{1}{R} \sum_{r=1}^{R} \widehat{y}_r[n]\,\widehat{y}_r^*[m]\,. \tag{3}$$

It is clearly unbiased: $\mathbf{E}\{\widehat{\Gamma}\} = \Gamma$ but yields a large risk.

The risk is calculated with the Hilbert–Schmidt norm of $\Gamma - \widehat{\Gamma}$, which can be rewritten as a Euclidean norm in $\mathbb{C}^{N^2}$

$$\|\Gamma - \widehat{\Gamma}\|^2 = \sum_{n=1}^{N} \sum_{m=1}^{N} |\gamma[n,m] - \widehat{\gamma}[n,m]|^2.$$

In signal processing, the number of realizations R is much smaller than N, and we often have $R = 1$. The number of data points NR is thus much smaller than the N^2 covariance coefficients that we intend to estimate, and as a result, $\|\Gamma - \widehat{\Gamma}\|^2$ is typically much larger than $\|\Gamma\|^2$. This means that estimating Γ as 0 instead of $\widehat{\Gamma}$ would reduce the estimation risk. Improving the empirical estimator $\widehat{\Gamma}$ can be viewed as a denoising problem similar to (1)

$$\widehat{\Gamma} = \Gamma + \widehat{\epsilon} \tag{4}$$

where the "noise" $\widehat{\epsilon} = \widehat{\Gamma} - \Gamma$ is the estimation error. A first major difficulty is that $\widehat{\epsilon}$ is not white. Its coefficients have non-Gaussian distributions that are highly correlated, and they depend on Γ. Moreover, Γ is characterized by a matrix of N^2 coefficients that can be assimilated to an element of $\mathbb{C}^{N^2}$ as opposed to the much smaller signal space $\mathbb{C}^N$ in the denoising problem. The covariance estimation problem is therefore much more complicated than the removal of white noise. Yet, we will develop a similar strategy to perform the estimation in both cases, and the noise removal problem provides a simpler context to better understand the properties of the resulting algorithm.

The denoising and covariance estimation problems are incorporated in the same framework where we are computing an estimator $\tilde{z}$ of a vector $z \in \mathbb{C}^P$ given $\widehat{z} = z + \widehat{\epsilon}$. For denoising,

$P = N$ and $z = x$, whereas for covariance estimation, $P = N^2$ and $z = \Gamma$. Projection estimators compute $\tilde{z}$ by projecting $\widehat{z}$ in a subspace $\mathcal{M}$ of dimension $\dim(\mathcal{M})$ much smaller than P:

$$\tilde{z} = P_{\mathcal{M}}(\widehat{z})\,.$$

The risk can be decomposed according to the Pythagorean theorem:

$$\|z - P_{\mathcal{M}}(\widehat{z})\|^2 = \|z - P_{\mathcal{M}}(z)\|^2 + \|P_{\mathcal{M}}(\widehat{\epsilon})\|^2. \tag{5}$$

Since $\mathbf{E}\{\widehat{z}\} = z$, the first term $\|z - P_{\mathcal{M}}(z)\|^2$ is the estimation bias, whereas $\|P_{\mathcal{M}}(\widehat{\epsilon})\|^2$ carries the remaining noise. An analysis of both terms shows that a projection estimator is an efficient estimator only if z can be well approximated by an element in $\mathcal{M}$ and if the space $\mathcal{M}$ has a relatively small dimension.

The space $\mathcal{M}$ plays the role of an estimation model, and one difficulty is to find an appropriate space that produces a small risk. Instead of choosing the subspace $\mathcal{M}$ *a priori*, one can try to optimize it, depending on $\widehat{z}$. We study a procedure where $\mathcal{M}$ is selected among a predefined family $\mathcal{F}$. The family $\mathcal{F}$ of subspaces must be constructed from prior information on the properties of z to guarantee that within $\mathcal{F}$, there exists at least one space $\mathcal{M}$ that leads to a small estimation risk. The issue will then be to estimate a "best" $\mathcal{M}$ within $\mathcal{F}$, which will be done with a penalization procedure.

The macrotile approach constructs the spaces $\mathcal{M}$ in $\mathcal{F}$ from a relatively small number $\mathcal{V}^{\#}$ of spaces $\{\mathbf{W}_k\}_{1 \le k \le \mathcal{V}^{\#}}$ of dimension 1 that are called macrotiles. Any *macrotile model* space $\mathcal{M}$ in $\mathcal{F}$ is constructed as a sum of $K = \dim(\mathcal{M})$ orthogonal macrotile spaces, whose indexes are in a set I of size K:

$$\mathcal{M} = \oplus_{k \in I} \mathbf{W}_k.$$

The number of such combination of orthogonal macrotiles is very large, and the total number of macrotile models $\mathcal{M}$ in $\mathcal{F}$ is typically an exponential function of the number $\mathcal{V}^{\#}$ of macrotiles. However, the fact that these models are obtained by a rearrangement of a limited number of macrotile spaces of dimension 1 will allow us to implement the search for a best model with a fast algorithm.

In the following, we will assume that the macrotile family $\mathcal{F}$ satisfies a progressive refinement property. This means that for each subspace macrotile model $\mathcal{M} \neq \mathbb{C}^P \in \mathcal{F}$, there exists a richer model $\mathcal{M}' \in \mathcal{F}$ and, hence, a larger macrotile subspace $\mathcal{M} \subset \mathcal{M}'$, whose dimension is at most twice as large:

$$\dim(\mathcal{M}) < \dim(\mathcal{M}') \le 2\dim(\mathcal{M})\,. \tag{6}$$

With this property, one can test models by refining them progressively.

B. Penalization

A general framework for penalization estimation is presented in [9] and [10] for the estimation of probability densities. These mathematical results do not apply to a penalized estimation of covariance matrices, but the underlying ideas of the macrotile approach are similar.

Given a family of macrotile models $\mathcal{F}$, we want to find a model $\widehat{\mathcal{M}} \in \mathcal{F}$, which reduces as much as possible the estima-

tion risk (5). For this purpose, we first estimate an upper bound of $\|P_{\mathcal{M}}(\widehat{\epsilon})\|^2$. This upper bound is then used by a penalization procedure that finds $\widehat{\mathcal{M}}$.

In the denoising context, $\widehat{\epsilon}$ is a Gaussian white noise of variance σ^2, and as a consequence, the analysis of the penalization procedure is much simpler than for covariance estimation. For noise removal, one can derive the penalization formula from general results in [9] and [10]. Since $P_{\mathcal{M}}$ is an orthogonal projection onto a space of dimension $\dim(\mathcal{M})$

$$\mathbf{E}\{\|P_{\mathcal{M}}(\widehat{\epsilon})\|^2\} = \sigma^2 \dim(\mathcal{M}).$$

Moreover, using a standard inequality on Gaussian distributions, since all models $\mathcal{M}$ in $\mathcal{F}$ are constructed from a family of $\mathcal{V}^{\#}$ macrotile spaces, one can prove [7] that if $\mathcal{V}^{\#} \geq \exp(1/3\sigma^2)$, the probability that

$$\|P_{\mathcal{M}}(\widehat{\epsilon})\| \leq \sigma \sqrt{3 \log \mathcal{V}^{\#} \dim(\mathcal{M})} \qquad (7)$$

is larger than $1 - 0.5 * [\mathcal{V}^{\#}]^{-1/2}$. Therefore, when $\mathcal{V}^{\#}$ tends to $+\infty$, this probability tends to 1. Note that in our framework, the total number of macrotiles $\mathcal{V}^{\#}$ is larger than the signal size N, which is itself large.

For covariance estimation, $\widehat{\epsilon} = \widehat{\Gamma} - \Gamma$ is a matrix of empirical estimation error, whose entries have non-Gaussian distributions. The variances of these errors are inversely proportional to the number R of realizations used to compute the sample means, and proportional to the covariance coefficients of Γ, which are bounded by the largest eigenvalue λ of Γ. In the Appendix, we derive that

$$\mathbf{E}\{\|P_{\mathcal{M}}(\widehat{\epsilon})\|^2\} \leq 2\lambda^2 \frac{\dim(\mathcal{M})}{R}. \qquad (8)$$

Like in the denoising case, the variance of the projected noise is bounded by a constant proportional to the dimension of the model, and the largest eigenvalue λ of Γ plays the same role as the variance of the noise in the denoising problem. Moreover, in [8], it is shown that similarly to (7), if Γ is the covariance of a Gaussian process, then $\widehat{\epsilon} = \Gamma - \widehat{\Gamma}$ satisfies

$$\|P_{\mathcal{M}}(\widehat{\epsilon})\| \leq 3\lambda \log \mathcal{V}^{\#} \sqrt{\frac{\dim(\mathcal{M})}{R}} \qquad (9)$$

with a probability that is larger than $1 - 6 * [\mathcal{V}^{\#}]^{-1/2}$ when $\mathcal{V}^{\#} \geq \exp(1/3\sigma^2)$. Again, this probability tends to 1 when $\mathcal{V}^{\#}$ tends to $+\infty$.

For noise removal and covariance estimation, inserting (7) or (9) into (5) give similar upper bounds for the estimation risk. With probability tending to 1 when $\mathcal{V}^{\#}$ tends to $+\infty$

$$\|z - P_{\mathcal{M}}(\widehat{z})\|^2 \leq \|z - P_{\mathcal{M}}(z)\|^2 + C \dim(\mathcal{M}) \qquad (10)$$

with $C = 3\sigma^2 \log \mathcal{V}^{\#}$ in the denoising case and $C = 9\lambda^2(\log \mathcal{V}^{\#})^2/R$ for covariance estimation. The penalization approach tries to find a model which nearly minimizes this upper bound and therefore produces a risk close to

$$\text{Risk}_{\min} = \inf_{\mathcal{M} \in \mathcal{F}} \{\|z - P_{\mathcal{M}}(z)\|^2 + C \dim(\mathcal{M})\}. \qquad (11)$$

For estimating a model $\mathcal{M}$ that can nearly reach the minimum $\text{Risk}_{\min}$, observe that

$$\|z - P_{\mathcal{M}}(z)\|^2 = \|z\|^2 - \|P_{\mathcal{M}}(z)\|^2.$$

Therefore, $\|z - P_{\mathcal{M}}(z)\|^2$ can be replaced by $-\|P_{\mathcal{M}}(z)\|^2$ when minimizing the upper bound (10). Since $\|P_{\mathcal{M}}(z)\|^2$ is unknown, the penalization approach replaces it by $\|P_{\mathcal{M}}(\widehat{z})\|^2$, and we compute

$$\widehat{\mathcal{M}} = \arg \min_{\mathcal{M} \in \mathcal{F}} \{-\|P_{\mathcal{M}}(\widehat{z})\|^2 + C' \dim(\mathcal{M})\}. \qquad (12)$$

Of course, the main difficulty is to prove that the risk $\|z - P_{\widehat{\mathcal{M}}}(\widehat{z})\|^2$ obtained with the penalized model $\widehat{\mathcal{M}}$ is of the same order as $\text{Risk}_{\min}$, with probability close to 1, despite the fact that we have approximated $\|P_{\mathcal{M}}(z)\|^2$ by $\|P_{\mathcal{M}}(\widehat{z})\|^2$. This requires to choose a penalization constant C' at least larger than C by a fixed factor.

In the context of noise removal, where $z = x$ and $\widehat{z} = \widehat{x}$ are signals of size N, since $C = 3\sigma^2 \log \mathcal{V}^{\#}$, we write

$$C' = C_p \sigma^2 \log \mathcal{V}^{\#}.$$

Using techniques developed by Barron *et al.* [9] based on Talagrand inequalities, if $C_p \geq 32\sigma^2$, one can prove [7] that the optimal penalized model

$$\widehat{\mathcal{M}} = \arg \min_{\mathcal{M} \in \mathcal{F}} \{-\|P_{\mathcal{M}}(\widehat{z})\|^2 + C_p \sigma^2 \log \mathcal{V}^{\#} \dim(\mathcal{M})\} \qquad (13)$$

has a nearly optimal risk

$$\|x - P_{\widehat{\mathcal{M}}}(\widehat{x})\|^2 \leq 4 C_p \text{Risk}_{\min} \qquad (14)$$

with a probability that tends to 1 when $\mathcal{V}^{\#}$ tends to infinity. Note that the multiplicative constant does not depend on the signal size N. This result is also similar to the one obtained for removing Gaussian white noise by thresholding coefficients in a best basis selected among a dictionary [2].

For covariance estimation, no mathematical result guarantees that in general, the penalized model yields a risk that is close to the minimum risk. However, Section IV-B explains that when estimating locally stationary processes in a dictionary of local cosine bases, this near-optimality result is valid.

III. Denoising With Local Cosine Macrotiles

This section studies the application of the macrotile penalization algorithm to noise removal, with a dictionary of macrotile models constructed with local cosine basis. In this context, finding the best macrotile model amounts to searching for a best basis and estimating the coefficients of the signal with an adaptive averaging in this basis. In a dictionary of local cosine bases, a fast algorithm computes the signal estimator with $O(N (\log_2 N)^2)$ operations for a signal of size N. This fast algorithm as well as many ideas introduced in this section will be used for covariance estimation.

A. From Bases to Macrotile Models

For noise removal, we saw in (14) that a penalized macrotile search yields a risk $\|x - P_{\widehat{\mathcal{M}}}(\widehat{x})\|^2$ that is of the same order as the minimum risk

$$\text{Risk}_{\min} = \inf_{\mathcal{M} \in \mathcal{F}} \left\{ \|x - P_{\mathcal{M}}(x)\|^2 + C_p \sigma^2 \log_2 \mathcal{V}^\# \dim(\mathcal{M}) \right\}. \tag{15}$$

The main issue is now to construct a family of macrotiles models which can guarantee that $\text{Risk}_{\min}$ is small. We explain how to construct such a family from a dictionary of bases and segmentations of these bases.

Consider a dictionary $\mathcal{D} = \{\mathcal{B}_b\}_{b \in \mathcal{I}}$, where each $\mathcal{B}_b = \{g_m^b\}_{1 \le m \le N}$ is an orthonormal basis of $\mathbb{C}^N$. A segmentation $S = \{I_k\}_{1 \le k \le K}$ is a partition of the integers in $[1, N]$ into K disjoint integer sets I_k, each of which has a cardinal written $I_k^\#$. The segmentation S applied to the basis $\mathcal{B}_b$ associates with each I_k a one-dimensional (1-D) macrotile space $\mathbf{W}_k$ generated by the sum vector $\sum_{m \in I_k} g_m^b$. The resulting *macrotile model* space $\mathcal{M} = \oplus_{k=1}^K \mathbf{W}_k$ contains all vectors of $\mathbb{C}^N$, whose coefficients in the basis $\mathcal{B}_b$ are constant over each macrotile $\{g_m^b\}_{m \in I_k}$. One can easily verify that the orthogonal projection $P_{\mathcal{M}}(\widehat{x})$ has decomposition coefficients in $\mathcal{B}_b$ that are equal to the average of the coefficients of $\widehat{x}$ on each macrotile:

$$P_{\mathcal{M}}(\widehat{x}) = \sum_{k=1}^K \frac{1}{I_k^\#} \left(\sum_{m \in I_k} \langle \widehat{x}, g_m^b \rangle \right) \sum_{m \in I_k} g_m^b \tag{16}$$

and hence

$$\|P_{\mathcal{M}}(\widehat{x})\|^2 = \sum_{k=1}^K \frac{1}{I_k^\#} \left| \sum_{m \in I_k} \langle \widehat{x}, g_m^b \rangle \right|^2. \tag{17}$$

Each model $\mathcal{M}$ is characterized by a basis $\mathcal{B}_b$ and by the segmentation S of this basis to estimate the coefficients with an averaging of the noisy ones. The penalization algorithm of Section II-B that computes a best empirical model $\widehat{\mathcal{M}}$ from $\widehat{x}$ thus amounts to finding a best basis in the dictionary $\mathcal{D}$ and a best segmentation of this basis. A macrotile algorithm can therefore be interpreted as a "best basis" denoising algorithm. However, as opposed to other "best basis" estimation algorithms [2], instead of estimating x by thresholding the coefficients of $\widehat{x}$ in a best basis, its coefficients are calculated with (16) using an adaptive averaging obtained by finding the best segmentation of the selected basis.

The family of macrotile models $\mathcal{M}$ in $\mathcal{F}$ corresponds to a choice of possible segmentations S applied to the bases $\mathcal{B}_b$ of the dictionary $\mathcal{D}$. To guarantee that $\text{Risk}_{\min}$ in (15) is small, there must exist a model $\mathcal{M}$ for which $\|x - P_{\mathcal{M}}(x)\|$ and $\dim(\mathcal{M})$ are simultaneously small. This means that one can obtain an approximation of x from its projection in a space $\mathcal{M}$ in $\mathcal{F}$ of small dimension $\dim(\mathcal{M})$. For a family $\mathcal{F}$ constructed from a dictionary of orthonormal bases $\mathcal{D}$, $P_{\mathcal{M}}(\widehat{x})$ is a signal whose decomposition coefficients in the basis $\mathcal{B}_b$ take only $\dim(\mathcal{M})$ different values. An important particular case is when x has a large number of decomposition coefficients in $\mathcal{B}_b$, which are close to zero and is thus well approximated by a partial sum of only K vectors in $\mathcal{B}_b$. Such an approximation has decompo-

sition coefficients in $\mathcal{B}_b$, which takes at most $K + 1$ different values (including 0), and hence, x can be well approximated by a projection $P_{\mathcal{M}}(x)$ with an appropriate macrotile subspace model that satisfies $\dim(\mathcal{M}) \le K + 1$.

B. Local Cosine Macrotile Models

For sound signals, the performance of the audio compression standard MPEG-2-AAC shows that most sounds can be well reconstructed from relatively few nonzero coefficients in a local cosine basis. This means that one can construct local cosine macrotile models, of small dimension, that efficiently approximate sound signals. This section constructs a family of macrotile models from a dictionary of local cosine bases, which will be used for noise removal.

A local cosine basis divides an interval $[0, N]$ with windows of varying sizes, which are multiplied by cosine vectors of varying frequencies. Let $1/2 = a_1 < a_2 < \cdots < a_{L+1} = N + 1/2$ be half-integers that define a partition of $[0, N]$ in L intervals. Let $l_p = a_{p+1} - a_p$, and $\eta \le \min_p l_p/2$. One can construct for each $p \in [1, L]$ a regular window $g_p[n]$ with support $[a_p - \eta, a_{p+1} + \eta]$ such that the local cosine family

$$\left\{ g_{p,q}[n] = g_p[n] \sqrt{\frac{2}{l_p}} \cos\left[\pi \left(q + \frac{1}{2} \right) \frac{n - a_p}{l_p} \right] \right\}_{1 \le p \le L, \, 0 \le q \le l_p} \tag{18}$$

is an orthonormal basis of $\mathbb{C}^N$[11]–[13]. Each local cosine vector $g_{p,q}$ has a support nearly localized in the interval $[a_p, a_{p+1}]$, and its Fourier transform has its energy mostly localized in the interval $[q\pi/l_p, (q + 1)\pi/l_p]$. It can thus be represented by the Heisenberg rectangle $[a_p, a_{p+1}] \times [q\pi/l_p, (q + 1)\pi/l_p]$ in a time-frequency plane. This Heisenberg box is called a *tile* because the union of the Heisenberg rectangles associated with a local cosine basis defines an exact tiling of the time–frequency plane. This is illustrated by Fig. 1, which displays the local cosine coefficients of an audio recording $x[n]$. Darker tiles correspond to high amplitude coefficients $|\langle x, g_{p,q} \rangle|$, whereas white tiles coefficients $|\langle x, g_{p,q} \rangle|$ that are nearly zero.

A difficult issue is to find a time partition that defines a local cosine basis that is well adapted to an estimation problem. Coifman and Meyer [11], [13] have thus constructed a dyadic dictionary of local cosine bases, which includes orthonormal bases defined over partitions whose intervals have sizes which are powers of 2. The dictionary is organized as a tree. For $a_p = pN2^{-j} + 1/2$ and $a_{p+1} = (p + 1)N2^{-j} + 1/2$, we denote by g_p^j the window whose support is $[a_p - \eta, a_{p+1} + \eta]$ and by $\mathcal{B}_p^j$ the corresponding family of $l_p = N 2^{-j}$ local cosine vectors

$$\mathcal{B}_p^j = \left\{ g_{p,q,j}[n] = g_p^j[n] \sqrt{\frac{2}{2^{-j}N}} \right.$$
$$\left. \cos\left[\pi \left(q + \frac{1}{2} \right) \frac{n - p2^{-j}N - \frac{1}{2}}{2^{-j}N} \right] \right\}_{0 \le q < N2^{-j}}.$$

The family of cosine vectors $\mathcal{B}_p^j$ is stored at the depth j and position p (from left to right) of a binary tree. The 2^j local cosine families $\mathcal{B}_p^j$ at the depth j divide the interval $[0, N]$ in 2^j

(a)

(b)

(c)
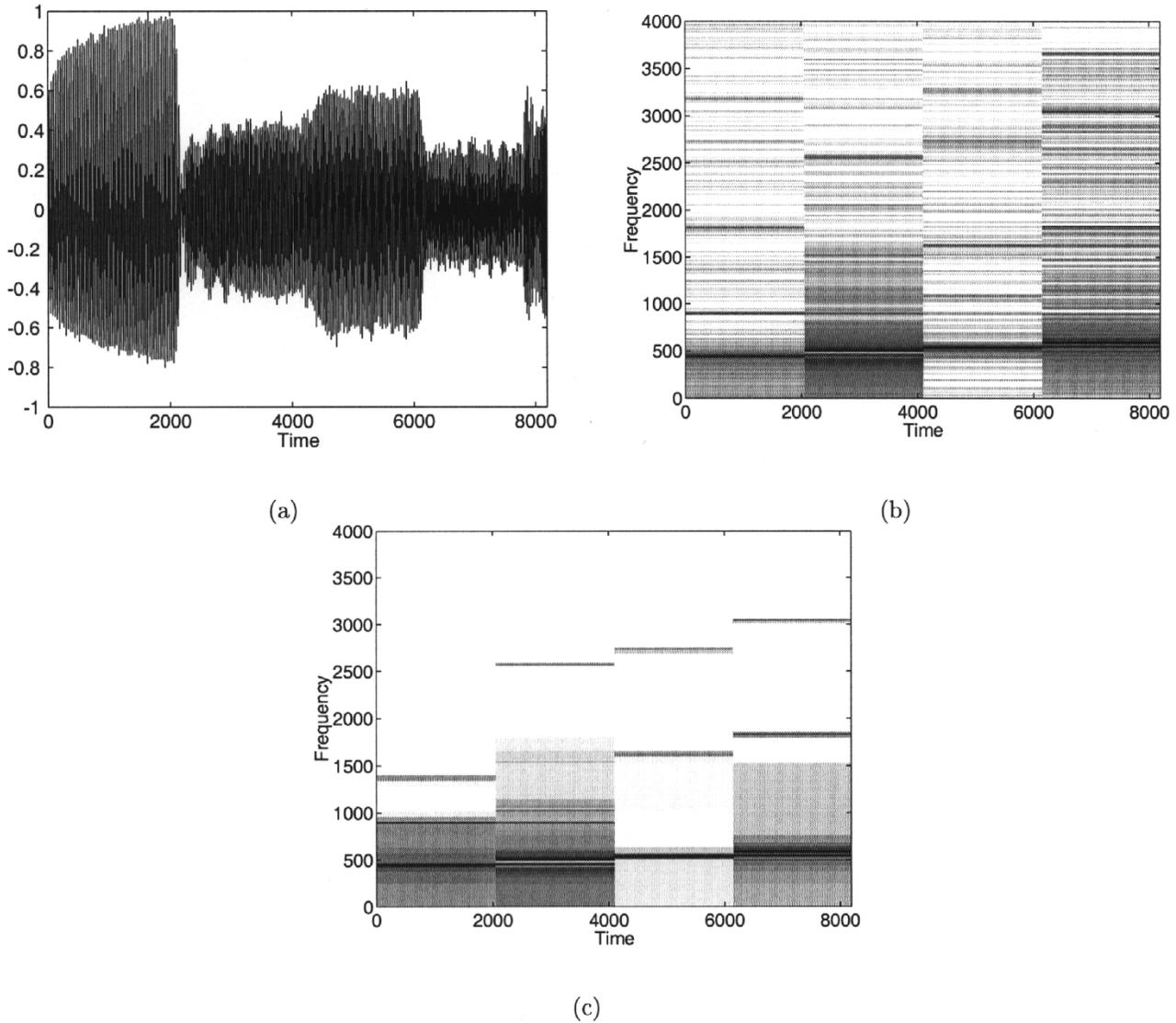
Fig. 1.   (a) Recording f of four notes played on a clarinet. (b) Tiling corresponding to the local cosine coefficients $|\langle f, g_{p,q}\rangle|$. The darker the grey level of the tile, the larger the amplitude. (c) Macrotiles obtained by averaging the tiles over groups of varying sizes for each time interval.

overlapping intervals. The local cosine family $\mathcal{B}_0^0$ at the root of the local cosine tree has N local cosine vectors that cover the whole signal support. The maximum depth of this tree is $J = \log_2(N/(2\eta))$ because $\eta = \min_p l_p/2 = 2^{-J-1}N$. There are thus $JN \le N \log_2 N$ local cosine vectors stored in the tree.

A local cosine basis in this dictionary is constructed with an admissible subtree, whose nodes have either 0 or 2 children. Let $b = \{p_l, j_l\}_{1 \le l \le L}$ be the position and depth of the L leaves of an admissible subtree. One can verify that the family

$$\mathcal{B}_b = \{\mathcal{B}_{p_l}^{j_l}\}_{1 \le l \le L} \qquad (19)$$

is a local cosine basis of $\mathbb{C}^N$ [13]. It defines a particular segmentation of the interval $[0, N]$ into windows of various sizes.

Remember that a family of macrotile models $\mathcal{F}$ is defined with a dictionary of bases $\mathcal{D}$ and a set of admissible segmentations. For a local cosine basis $\mathcal{B}_b = \{\mathcal{B}_{p_l}^{j_l}\}_{1 \le l \le L}$, we will consider segmentations that are performed independently within each family $\mathcal{B}_p^j$ and that satisfy the refinement property (6).

Sound signals can include frequency tones that are localized over a time interval and have narrow frequency bandwidth. Let $\mathcal{B}_j^p = \{g_{p,q,j}\}_{0 \le q < N\,2^{-j}}$ be a local cosine family whose window covers a time interval $[p\,2^{-j}\,N, (p+1)\,2^{-j}\,N]$, where the signal has several narrow frequency tones. To efficiently approximate such a signal with a macrotile model of small dimension, it is necessary to define a frequency segmentation that uses narrow macrotiles in the neighborhood of the frequency tones to approximate them efficiently. Constructing a local cosine macrotile model is similar to finding a piecewise constant approximation of the local signal coefficients $\langle f, g_{p,q,j}\rangle$ when the frequency index q varies for p and j fixed. For each p, j, we thus consider any possible grouping, which corresponds to a piecewise constant sequence of the $2^{-j}N$ local cosine coefficients. We will, however, impose that each macrotile has a size which is a power of two. For a given p, j, a segmentation in $K_{j,p}$ macrotile frequency intervals is thus defined by a sequence of length $S = \{2^{s_k}\}_{1 \le k \le K_{j,p}}$. Imposing that each group has a number of elements that is a power of 2
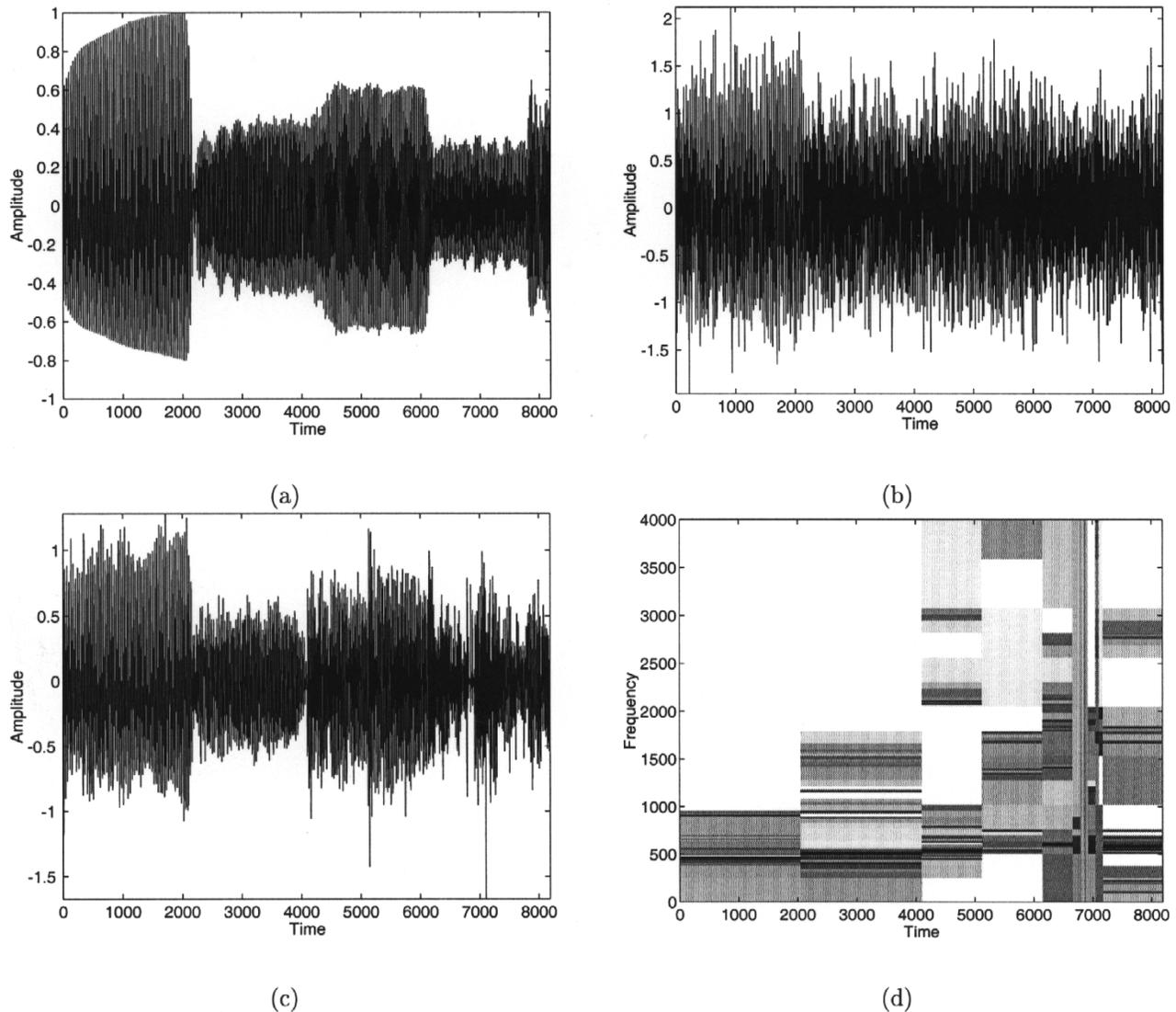
Fig. 2.   (a) Music signal f. (b) Noisy signal $\widehat{x} = x + \widehat{\epsilon}$. (c) Penalized macrotile estimation $P_{\widehat{\mathcal{M}}}(\widehat{x})$ calculated with (21). (d) Representation of the time-frequency macrotiles of $P_{\widehat{\mathcal{M}}}(\widehat{x})$. The darkest macrotiles correspond to high energy coefficients.

has no asymptotic effect on the performance of the macrotile averaging and allows one to optimize this segmentation with a fast CART algorithm [14]. A macrotile model $\mathcal{M}$ is specified by the indices of the local cosine basis $\mathcal{B}_b = \{\mathcal{B}_{pl}^{jl}\}_{1 \leq l \leq L}$ and the adaptive segmentation S_l for each $\mathcal{B}_{pl}^{jl}$

$$\mathcal{M} = \{j_l, p_l, S_l\}_{1 \leq l \leq L}. \tag{20}$$

Fig. 1 displays the representation of $P_{\mathcal{M}}(x)$ for one such macrotile model. This macrotile model is just an example and is not optimal in any sense with respect to this signal. The gray level of each macrotile indicates the amplitude of the average inner product corresponding to a particular grouping of local cosine tiles. For a local cosine tree of depth $J \leq \log_2 N$, one can verify that the total number of different macrotile models $\mathcal{M}$ is larger than $2^{N/2}$, but all these models are constructed from only a number $\mathcal{V}^\#$ of macrotiles, which satisfies

$$(2J - 1)N \leq \mathcal{V}^\# \leq 2JN.$$

C. Denoising With Fast Model Search

We now study the application of local cosine macrotiles to noise removal with a fast implementation using the tree structure of local cosine dictionaries. Given a noisy signal $\widehat{x} = x + \widehat{\epsilon}$, a best local cosine model is selected according to (13). Since $\log_2 \mathcal{V}^\# \approx \log_2 N$, this amounts to computing

$$\widehat{\mathcal{M}} = \arg \min_{\mathcal{M} \in \mathcal{F}} \left\{ -\|P_{\mathcal{M}}(\widehat{x})\|^2 + C_p \sigma^2 \log N \dim(\mathcal{M}) \right\}. \tag{21}$$

Fig. 2 illustrates this macrotile denoising for the music signal with four notes, which is shown in Fig. 1(a). Fig. 2(b) is the signal contaminated by a white noise for an SNR of 0 db. Fig. 2(c) gives the macrotile estimation $P_{\widehat{\mathcal{M}}}(\widehat{x})$ calculated with the macrotile denoising algorithm. The resulting SNR is 7 db, which corresponds to a gain of 7 db. Fig. 2(d) displays the macrotiles in the time–frequency plane. The harmonics appear as darker and more narrow macrotiles; the time–frequency structures of the signal are clearly preserved. Two other experiments were performed with the same original signal and other

white noises corresponding to SNR of 6.5 db and -6 db; the gains were, respectively, 6 and 8 db. The audio quality of the signal is considerably improved by this macrotile denoising, yet this example is only an illustration, and further refinements in the models are needed to optimize the noise removal for music signals. It is shown in [7] that as for the asymptotic performance, the macrotile denoising is equivalent to thresholding in the best basis. However, the macrotile denoising offers more flexibility since it averages the coefficients instead of replacing the small ones with 0.

When the family of models is constructed from dictionaries of orthonormal bases having a tree structure, like a local cosine dictionary, the best penalized model is computed with a fast dynamic programming algorithm similar to the one of Coifman and Wickerhauser [15], which takes advantage of the additivity property of the penalization cost for partial models. A partial model $\mathcal{M}$ is characterized by an orthonormal basis $\mathcal{B}$ of a subspace $\mathbf{V}$ of $\mathbb{C}^N$, as opposed to a basis of the whole space, and a segmentation S of $\mathcal{B}$. The cost of this model is defined by

$$\text{Cost}(\mathcal{M}) = -\|P_{\mathcal{M}}(\widehat{x})\|^2 + C_p\,\sigma^2\,\log N\,\dim(\mathcal{M})$$

where $\|P_{\mathcal{M}}(\widehat{x})\|^2$ is calculated with (17) by summing only the coefficients restricted to the basis $\mathcal{B}$ of $\mathbf{V}$. Let $\mathcal{M}_1 = (\mathcal{B}_1\,,\,S_1)$ and $\mathcal{M}_2 = (\mathcal{B}_2\,,\,S_2)$ be two partial models where $\mathcal{B}_1$ and $\mathcal{B}_2$ are two orthogonal families of vectors. One can define a new model $\mathcal{M} = (\mathcal{B}_1 \cup \mathcal{B}_2\,,\,S_1 \cup S_2)$, where $S_1 \cup S_2$ is the segmentation of $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ obtained by segmenting $\mathcal{B}_1$ with S_1 and $\mathcal{B}_2$ with S_2. Clearly

$$\text{Cost}(\mathcal{M}) = \text{Cost}(\mathcal{M}_1) + \text{Cost}(\mathcal{M}_2)\,. \tag{22}$$

Using this additive property in a tree dictionary, one can now identify the model that minimizes the cost with the fast bottom up strategy of Coifman and Wickerhauser [15].

To any node of the local cosine tree is associated a family of local cosine vectors $\mathcal{B}_p^j$, which generates a vector space $\mathbf{V}_p^j$. Let $\mathcal{M}_p^j$ be the optimal model of $\mathbf{V}_p^j$ that minimizes $\text{Cost}(\mathcal{M})$ among all models $\mathcal{M}$ defined by $(\mathcal{B}, S)$, where $\mathcal{B}$ is a local cosine basis of $\mathbf{V}_p^j$. One possible candidate is the model $\mathcal{M}_p^{j*} = (\mathcal{B}_p^j, S_p^j)$, where the segmentation S_p^j is adjusted to minimize the cost associated with the basis $\mathcal{B}_p^j$. However, other models can also be considered by building other bases of $\mathbf{V}_p^j$ from the local cosine vectors located in the descendant nodes. The space $\mathbf{V}_p^j$ can be decomposed into two orthogonal subspaces $\mathbf{V}_{2p}^{j+1}$ and $\mathbf{V}_{2p+1}^{j+1}$ generated by the two local cosine families $\mathcal{B}_{2p}^{j+1}$ and

$\mathcal{B}_{2p+1}^{j+1}$ located at the children nodes in the binary tree. Using the additivity property (22) we easily verify that we have (23), shown at the bottom of the page. At the leaves of a local cosine tree of depth J, the only possible choice is $\mathcal{M}_p^J = \mathcal{M}_p^{J*}$ since $\mathcal{B}_p^J$ is the only basis available for $\mathbf{V}_p^J$. Using the aggregation relation (23), the optimal model is computed by going up the tree along each branch, until the root where $\mathbf{V}_0^0 = \mathbb{C}^N$. The resulting optimal model is $\mathcal{M}_0^0 = \widehat{\mathcal{M}}$ and is the model that minimizes the cost over $\mathbb{C}^N$. The overall computational complexity of the algorithm is driven by the number of operations to compute the models $\mathcal{M}_p^{j*} = (\mathcal{B}_p^j, S_p^j)$ at all nodes of the tree.

For this purpose, we first compute the JN local cosine coefficients of the noisy signal $\widehat{\alpha}_p^j[q] = \langle \widehat{x}, g_{p,q,j} \rangle$ for all local cosine vectors $g_{j,q,p}$ in the tree. With a fast lapped cosine transform [12], this requires $O(JN \log_2 N)$ operations. Then, for each j and p, we compute the segmentation S_p^j, which yields a model $\mathcal{M}_p^{j*}$ of minimum cost for local cosine family $\mathcal{B}_p^j$. The cost is calculated from the $N2^{-j}$ coefficients $\{\widehat{\alpha}_p^j[q]\}_{1 \le q \le N2^{-j}}$ associated with $\mathcal{B}_p^j$.

The optimal segmentation S_p^j can be computed with $O(2^{-j}N)$ operations with a CART algorithm [14]. In the following, we write $\text{Cost}(\mathcal{S})$ as the cost of a partial macrotile model corresponding to a segmentation $\mathcal{S}$ of $\mathcal{B}_p^j$. By definition, S_p^j is the segmentation that minimizes this cost. To compute it, we use a segmentation tree, which subdecomposes any segmentation in partial segmentations whose cost are summed. At the level l of this segmentation tree, each node of index $0 \le k < 2^l$ corresponds to segmentations of the group of $N2^{-l-j}$ coefficients $\widehat{\alpha}_p^j[q]$ for $kN2^{-l-j} < q \le (k+1)N2^{-l-j}$. Let $\mathcal{S}_l^k$ be the optimal segmentation of this group, which yields a partial model of minimum cost, which we write $\text{Cost}(\mathcal{S}_l^k)$. One possible candidate is the segmentation $\mathcal{S}_l^{k*}$, where all coefficients of the group are averaged into one macrotile whose average value is

$$\bar{\alpha}_l[k] = \frac{1}{N2^{-l-j}} \sum_{q=kN2^{-l-j}+1}^{(k+1)N2^{-l-j}} \widehat{\alpha}_p^j[q].$$

The cost of the partial model associated with this segmentation is

$$\text{Cost}(\mathcal{S}_l^{k*}) = -N\,2^{-l-j}\,|\bar{\alpha}_l[k]^2| + C_p\,\log N.$$

As in (23), using the additivity property of the cost function, one can verify that we have (24), shown at the bottom of the page. At

$$\mathcal{M}_p^j = \begin{cases} \mathcal{M}_p^{j*}, & \text{if } \text{Cost}(\mathcal{M}_p^{j*}) \le \text{Cost}(\mathcal{M}_{2p}^{j+1}) + \text{Cost}(\mathcal{M}_{2p+1}^{j+1}) \\ \mathcal{M}_{2p}^{j+1} \cup \mathcal{M}_{2p+1}^{j+1}, & \text{if } \text{Cost}(\mathcal{M}_p^{j*}) > \text{Cost}(\mathcal{M}_{2p}^{j+1}) + \text{Cost}(\mathcal{M}_{2p+1}^{j+1}). \end{cases} \tag{23}$$

$$\mathcal{S}_k^l = \begin{cases} \mathcal{S}_k^{l*}, & \text{if } \text{Cost}(\mathcal{S}_k^{l*}) \le \text{Cost}(\mathcal{S}_{2k}^{l+1}) + \text{Cost}(\mathcal{S}_{2k+1}^{l+1}) \\ \mathcal{S}_{2k}^{l+1} \cup \mathcal{S}_{2k+1}^{l+1}, & \text{if } \text{Cost}(\mathcal{S}_k^{l*}) > \text{Cost}(\mathcal{S}_{2k}^{l+1}) + \text{Cost}(\mathcal{S}_{2k+1}^{l+1}). \end{cases} \tag{24}$$

the maximum depth $L = \log_2 N - j$ of this tree, the only possible segmentation is $\mathcal{S}_k^L = \mathcal{S}_k^{L*}$ because there is only one coefficient in each group. Using (24), a bottom-up strategy requires $O(2^{-j}N)$ operations to compute $\mathcal{S}_0^0$, which is the optimal segmentation associated to all coefficients $\{\widehat{\alpha}_p^j[q]\}_{1 \leq q \leq N2^{-j}}$, and hence, $S_p^j = \mathcal{S}_0^0$. This segmentation defines $\mathcal{M}_p^{j*}$, and by definition

$$\text{Cost}(\mathcal{M}_p^{j*}) = \text{Cost}(\mathcal{S}_0^0).$$

Since $\mathcal{M}_p^{j*}$ is computed with $O(2^{-j}N)$ operations, the optimal model $\widehat{\mathcal{M}}$ model is obtained with $O(JN)$ operations using (23). Since $J \leq \log_2 N$, the penalization algorithm optimizing the model over adaptive segmentations thus requires $O(N(\log_2 N)^2)$ operations. Note that the computational complexity of a thresholding estimator in a best local cosine basis is of the same order because it is driven by the number of operations to compute all the local cosine coefficients in the dictionary [13].

IV. ESTIMATION OF COVARIANCE OF LOCALLY STATIONARY PROCESSES

The study of the additive noise removal problem in Section III provides important ideas and algorithmic tools to apply the penalized macrotile procedure to covariance estimation. The next section shows that a dictionary of macrotile models can also be constructed with a segmentation of orthonormal bases, in which case, the search for a best macrotile model can be interpreted as a simultaneous estimation of an approximate Karhunen–Loève basis as well as the diagonal covariance coefficients in this basis. The consistency and numerical performance of this estimator is studied for locally stationary processes.

A. Macrotile Estimation of Covariance

To estimate the covariance Γ of a process $\widehat{y}[n]$ of size N, according to (12), we compute the best penalized model

$$\widehat{\mathcal{M}} = \arg \min_{\mathcal{M} \in \mathcal{F}} \left\{ -\|P_{\mathcal{M}}(\widehat{\Gamma})\|^2 + C_p \frac{\lambda^2}{R} \left(\log \mathcal{V}^\# \right)^2 \dim(\mathcal{M}) \right\}$$
(25)

where $P_{\mathcal{M}}$ is an orthogonal projection in a space $\mathcal{M}$ of operators, and λ is an upper bound of the largest eigenvalue of Γ that we suppose known *a priori*. We study the calculation and properties of such an estimator for a family of macrotile models $\mathcal{F}$ constructed from a dictionary of orthonormal bases $\mathcal{D} = \{\mathcal{B}_b\}_b$ that are segmented with a strategy similar to the one used in Section III-A for the noise-removal problem.

A segmentation $S = \{I_k\}_{1 \leq k \leq K}$ makes a partition of $[1, N]$ in K disjoint integer sets I_k, each of which has a cardinal written $I_k^\#$. The segmentation S applied to the basis $\mathcal{B}_b$ associates with each I_k a 1-D macrotile space $\mathbf{W}_k$ of operators. It is defined as the 1-D space of operators, which are diagonalized in $\mathcal{B}_b$ and whose eigenvalues are the same for all $\{g_m^b\}_{m \in I_k}$ and equal to zero for all other vectors of $\mathcal{B}_b$.

Let us denote by $\mathbf{V}_k$ the subspace of $\mathbb{C}^N$ generated by the vectors $\{g_m^b\}_{m \in I_k}$. A macrotile model $\mathcal{M} = \oplus_{k=1}^K \mathbf{W}_k$ is a space of operators of dimension K whose eigenspaces are the $\{\mathbf{V}_k\}_{1 \leq k \leq K}$. The basis $\mathcal{B}_b$ diagonalizes the operators in $\mathcal{M}$ and

their eigenvalues are constant over each macrotile $\{g_m^b\}_{m \in I_k}$. Let $P_{\mathbf{V}_k}$ be the orthogonal projection onto the vector space $\mathbf{V}_k$. It applies to signals in $\mathbb{C}^N$. The orthogonal projection $P_{\mathcal{M}}(\widehat{\Gamma})$ on the operator space $\mathcal{M}$ is an operator that can be decomposed as a linear combination of the projectors $P_{\mathbf{V}_k}$ on each eigenspace $\mathbf{V}_k$. One can verify that the multiplicative factors are averages of the diagonal coefficients of $\widehat{\Gamma}$ in the direction of vectors in $\mathbf{V}_k$:

$$P_{\mathcal{M}}(\widehat{\Gamma}) = \sum_{k=1}^K \frac{1}{I_k^\#} \left(\sum_{m \in I_k} \langle \widehat{\Gamma} g_m^b, g_m^b \rangle \right) P_{\mathbf{V}_k} \qquad (26)$$

and since $\{P_{\mathbf{V}_k}\}_{1 \leq k \leq K}$ is an orthonormal family of operators

$$\|P_{\mathcal{M}}(\widehat{\Gamma})\|^2 = \sum_{k=1}^K \frac{1}{I_k^\#} \left| \sum_{m \in I_k} \langle \widehat{\Gamma} g_m^b, g_m^b \rangle \right|^2. \qquad (27)$$

Using (3), we verify that each diagonal coefficient of the empirical covariance matrix is calculated from the realizations $\{\widehat{y}_r\}_{1 \leq r \leq R}$ of $\widehat{y}$ with

$$\langle \widehat{\Gamma} g_m^b, g_m^b \rangle = \frac{1}{R} \sum_{r=1}^R |\langle \widehat{y}_r, g_m^b \rangle|^2. \qquad (28)$$

The main difficulty is now to evaluate the resulting risk $\|\Gamma - P_{\mathcal{M}}(\widehat{\Gamma})\|$ depending on the properties of the process $\widehat{y}[n]$. If the vectors $\{g_m^b\}_{1 \leq m \leq N}$ of $\mathcal{B}_b$ are discrete Fourier vectors multiplied by square windows that localize them over subintervals of $[1, N]$, then (28) can be interpreted as a simple periodogram. Section IV-B studies the application of local cosine bases computed with smooth windows to estimate the spectrum of locally stationary processes.

Let $\mathcal{B}_{\widehat{b}}$ be the orthonormal basis corresponding to the selected penalized model $\widehat{\mathcal{M}}$. The estimator $P_{\widehat{\mathcal{M}}}(\widehat{\Gamma})$ of Γ is an operator that is diagonal in $\mathcal{B}_{\widehat{b}}$ and whose eigenvalues have been averaged in $\dim(\widehat{\mathcal{M}})$ groups. The best basis $\mathcal{B}_{\widehat{b}}$ can thus be interpreted as an approximate Karhunen–Loève basis of Γ, which is optimized to reduce the risk $\|\Gamma - P_{\widehat{\mathcal{M}}}(\widehat{\Gamma})\|$. Within this basis, the segmentation is adapted to estimate the diagonal coefficients of Γ with an appropriate averaging of the empirical diagonal coefficients of $\widehat{\Gamma}$. Several approaches have already been studied to estimate covariance matrices by searching for an approximate Karhunen–Loève basis [3], [4]. The macrotile algorithm improves over these approaches by simultaneously estimating the basis and the diagonal covariance coefficients in this basis. This provides better numerical estimations and leads to mathematical consistency results, which could not be obtained with the previous approaches. In the following, we evaluate both numerically and mathematically the efficiency of the macrotile penalized estimator in the context of locally stationary processes.

B. Locally Stationary Covariances

Intuitively, a locally stationary process can be locally approximated by stationary processes. Many different mathematical modeling of such processes have been proposed [3], [6], [16], [17], [19], but the derivation of consistent estimators of the resulting covariance is still an open issue. A zero-mean process $\widehat{y}[n]$ for $n \in \mathbb{Z}$ is wide-sense stationary if its covariance

$\gamma[n, m] = \mathbf{E}\{\widehat{y}[n]\,\widehat{y}[m]\}$ satisfies $\gamma[n, m] = c[n - m]$. The corresponding covariance operator Γ is a convolution operator, diagonalized in the Fourier basis, and whose eigenvalues (power spectrum) are given by

$$P(\omega) = \sum_\tau c[\tau]\, e^{-i\omega\tau}.$$

If the process is not strictly stationary, we can still write the covariance

$$c[n, \tau] = \gamma[n, n - \tau]..$$

Priestley [6] and Martin and Flandrin [19] have proposed several definitions of locally stationary processes by introducing a time-varying power spectrum defined by

$$P[n, \omega] = \sum_{\tau \in \mathbb{Z}} c[n, \tau]\, e^{-i\omega\tau}.$$

It can also be rewritten as the expected value of the Wigner–Ville distribution of the process $\widehat{y}$:

$$\widetilde{P}[n, \omega] = \mathbf{E}\left\{\sum_{\tau \in \mathbb{Z}} \widehat{y}[n]\,\widehat{y}[n + \tau]\, e^{-i\omega\tau}\right\}.$$

If the process is locally stationary, then one can expect that $P[n, \omega]$ has slow variations as a function of the position parameter n. If $\widehat{y}$ also has a fast decorrelation, which means that $c[n, \tau]$ decays quickly when τ increases, then $P[n, \omega]$ is a smooth function along the frequency variable ω. This time-varying spectrum has been further studied in [16] and [17], as well as in [3], where it is observed that if $P[n, \omega]$ is nearly constant over the time–frequency Heisenberg rectangle of a local cosine vector, then this vector is nearly an eigenvector of the covariance operator Γ. More precisely, if $P[n, \omega]$ is approximatively equal to a constant $P_{p,q}$ over the Heisenberg rectangle $[a_p, a_{p+1}] \times [q\pi/l_p, (q + 1)\pi/l_p]$ of a local cosine vector $g_{p,q}[n]$ defined in (18), then

$$\Gamma g_{p,q}[n] \approx P_{p,q}\, g_{p,q}[n].$$

In this case, the eigenvectors of the covariance can be approximated by local cosine vectors and the eigenvalues are given by the time-varying power spectrum. This shows that the Karhunen–Loève basis of a locally stationary process can be approximated by an appropriate local cosine basis, whose vectors have a time–frequency localization adapted to the variations of the time-varying power spectrum [3].

C. Local Cosine Approximate Karhunen–Loève

Since the Karhunen-Loève basis of a locally stationary process can be approximated by a local cosine basis, one can estimate its covariance with a penalized macrotile algorithm that searches for a best macrotile model in a family $\mathcal{F}$ constructed with a dictionary of local cosine bases. A fast algorithm is described to compute the resulting covariance estimator.

To specify a family of macrotiles $\mathcal{F}$, we still need to choose the set of all possible segmentations for each basis. This section concentrates on locally stationary processes having a fast decorrelation, which means that $c[n, \tau]$ has a fast decay as a function of τ, uniformly in n. In this case, the time-varying power

spectrum $P[n, \omega]$ is a uniformly smooth function of ω. An appropriate estimator of the power spectrum (eigenvalues) is thus obtained by averaging uniformly in frequency the sample-mean estimator of the diagonal covariance coefficients in a local cosine basis. This corresponds to a particular family of admissible segmentations of vectors in a local cosine basis.

We saw in (19) that a local cosine basis can be written $\mathcal{B}_b = \{\mathcal{B}_{p_l}^{j_l}\}_{1 \leq l \leq L}$, where each $\mathcal{B}_p^j$ regroups $2^{-j}N$ cosine vectors whose support is $[p\,2^{-j}N - \eta,\ (p + 1)\,2^{-j}N + \eta]$. Among these vectors, we choose segmentations that perform a uniform averaging in frequency. This means that local cosine coefficients in this interval are averaged along frequencies in consecutive groups of 2^s coefficients, where 2^s is a fixed length with $1 \leq 2^s \leq 2^{-j}N$. Observe that this family of segmentations is more restrictive than the segmentations used in the local cosine macrotile family defined in Section III-B, which were not necessarily uniform. The number of macrotile models is therefore smaller, but the total number $\mathcal{V}^\#$ of macrotiles combined to construct these models is the same in both cases, and $\log_2 \mathcal{V}^\#$ is nearly equal to $\log_2 N$.

Following (25), the best local cosine macrotile $\widehat{\mathcal{M}}$ is calculated by minimizing

$$\mathrm{Cost}(\mathcal{M}) = -\|P_{\mathcal{M}}(\widehat{\Gamma})\|^2 + C_p\, \frac{\lambda^2}{R}\left(\log N\right)^2 \dim(\mathcal{M}). \quad (29)$$

Each model $\mathcal{M}$ is specified by a local cosine basis $\mathcal{B}_b$ and its segmentation. The norm $\|P_{\mathcal{M}}(\widehat{\Gamma})\|^2$ is computed from the decomposition coefficients of each realization $\widehat{y}_r$ in the basis $\mathcal{B}_b$ with (27) and (28). The minimization of this cost in a local cosine dictionary is implemented with minor modifications of the fast algorithm described in Section III-C because it has the same additivity properties.

The best model $\widehat{\mathcal{M}}$ is constructed from partial optimal models $\mathcal{M}_p^j$ associated with the local cosine spaces $\mathbf{V}_p^j$ along the tree nodes, using the aggregation property (23). The only difficulty is to construct the model $\mathcal{M}_p^{j*}$ where the segmentation S_p^j is adjusted to minimize the cost associated with the basis $\mathcal{B}_p^j$.

For each realization $\widehat{y}_r$ of the process, we first compute the JN coefficients $\langle \widehat{y}_r, g_{p,q,j}\rangle$ corresponding to all local cosine vectors $g_{j,q,p}$ in the tree depth J. With a fast lapped cosine transform [12], for R realizations, we require $O(R\,J\,N \log_2 N)$ operations. Following (28), we then compute

$$\widehat{\alpha}_p^j[q] = \langle \widehat{\Gamma} g_{p,q,j},\, g_{p,q,j}\rangle = \frac{1}{R}\sum_{r=1}^{R} |\langle \widehat{y}_r, g_{p,q,j}\rangle|^2$$

with $O(R\,J\,N)$ operations.

To compute the segmentation S_p^j, which minimizes the cost associated with the basis $\mathcal{B}_p^j$, is a much simpler operation than in Section III-C. Indeed, we are restricted to uniform segmentations that average the $2^{-j}N$ coefficients $\widehat{\alpha}_p^j[q]$ by intervals of same size 2^s, for any $1 \leq 2^s \leq 2^{-j}N$. The resulting averaged coefficients are

$$\bar{\alpha}_s[k] = \frac{1}{2^s}\sum_{q=k2^s+1}^{(k+1)2^s} \alpha_p^j[q] \quad \text{for } 0 \leq k \leq (2^{-s-j}N - 1).$$
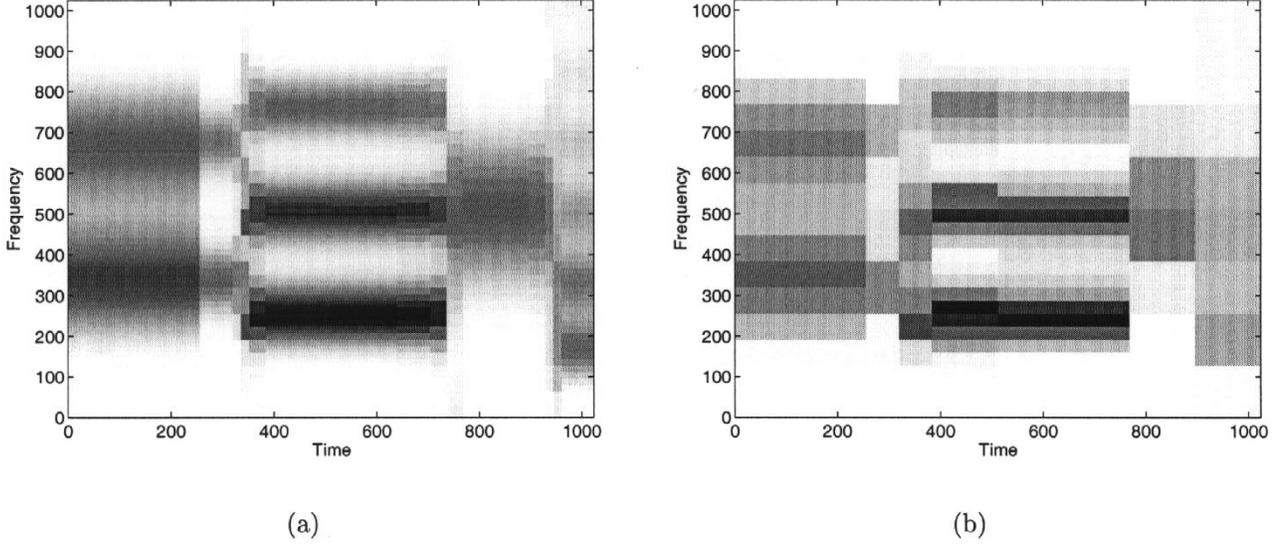
Fig. 3. (a) Diagonal coefficients of the covariance Γ in the ideal local cosine basis. (b) Macrotiles calculated by the penalized estimator from $R = 5$ realizations.

With a standard dyadic averaging cascade by groups of 2, this is calculated for all $0 \leq s \leq \log_2 N - j$ with $O(2^{-j} N)$ operations. The cost associated with the only model $\mathcal{M}_s$ of dimension $K = 2^{-s-j} N$ is

$$\text{Cost}(\mathcal{M}_s) = -2^s \sum_{k=1}^{K} |\bar{\alpha}_s[k]|^2 + C_p (\log N)^2 \frac{K}{R}.$$

Finding the model $\mathcal{M}_p^{j*}$ that minimizes this cost over s thus requires $O(2^{-j} N)$ operations. The total number of operations to compute all models $\mathcal{M}_p^{j*}$ in the tree is thus $O(J N)$. With the aggregation property (23), the optimal model $\widehat{\mathcal{M}}$ is then also computed with $O(J N)$ operations. The overall complexity of the penalization algorithm is therefore dominated by the number of operations to compute the local cosine coefficients of all realizations, which is $O(R N (\log_2 N)^2)$, since $J \leq \log_2 N$.

Fig. 3 illustrates this macrotiles covariance estimation with a synthetic example. A locally stationary process is constructed by aggregating independent stationary processes over the intervals $[0,240]$, $[240\,340]$, $[340\,740]$, $[740\,970]$, $[970,1024]$. Let $\Gamma_{\mathcal{B}}$ be the diagonal operator in a local cosine basis $\mathcal{B}$ whose diagonal coefficients are equal to the diagonal coefficients of Γ. The ideal cosine basis is the one where $\|\Gamma_{\mathcal{B}} - \Gamma\|$ is minimum. Fig. 3(a) shows the exact diagonal values of Γ in this ideal basis. No macrotile averaging of coefficients was performed. The resulting minimum error is $\|\Gamma - \Gamma_{\mathcal{B}}\|^2 / \|\Gamma\|^2 = 0.09$. Fig. 3(b) represents the penalized macrotile estimator $P_{\widehat{\mathcal{M}}}(\widehat{\Gamma})$ computed from $R = 5$ realizations of the process $\widehat{y}$. The risk is only twice above the ideal minimum risk: $\|\Gamma - \widehat{\Gamma}_{\widehat{\mathcal{M}}}\|_H^2 / \|\Gamma\|_H^2 = 0.19$.

Speech signals are typical realizations of locally stationary processes, and unvoiced speech signals have a fast decorrelation. Their covariance can thus be estimated with the penalized macrotile algorithm. The problem is particularly difficult since we have only $R = 1$ realization and do not use any parametric model. Before computing the macrotiles, a preprocessing normalizes the average signal energy over intervals of 500 samples. This improves the estimation in time domains where the signal

has a smaller energy. Fig. 4(a) shows an example of unvoiced speech recording including a succession of five sounds. When listening to this signal, the intervals of stationarity are approximatively $[0,800]$, $[800,2800]$, $[2800,5000]$, $[5000,7300]$, and $[7300,8192]$. Observe that the penalized macrotile estimation selects time windows that correspond approximatively to these intervals. In particular, the two [œ] sounds (phonetic writing) are represented by similar macrotiles. Transient sounds such as [p] are represented by macrotiles having a poor frequency localization. Despite the relatively good performance of such an estimation, the fast decorrelation model is too restrictive for voiced speech signals which can include narrow frequency tones. This issue is further addressed in Section IV-D.

The consistency of this macrotile estimation procedure has been proved using a formal definition of locally stationary processes that imposes conditions directly on the covariance $c[n,\tau]$ [4] as opposed to its time varying power-spectrum. The fact that $c[n,\tau]$ does not vary too much along the time variable n is imposed by supposing that $c[n,\tau]$ has a uniformly bounded variation when n varies. This condition is satisfied if $c[n,\tau]$ has regular variations as a function of n or few isolated brutal transitions. The fast decorrelation property of these processes imposes a minimum decay rate of $|c[n,\tau]|$ as a function of τ. If $\widehat{y}$ is a locally stationary Gaussian process that satisfies these conditions, then the minimum risk of a penalized macrotile estimator

$$\text{Risk}_{\min} = \inf_{\mathcal{M} \in \mathcal{F}} \left\{ \|x - P_{\mathcal{M}}(x)\|^2 + C_p \lambda^2 (\log_2 N)^2 \dim(\mathcal{M}) \right\}$$

can be bounded by [8]

$$\frac{\text{Risk}_{\min}}{N} \leq D_1 \frac{N^{-\gamma}}{R^\gamma} \qquad (30)$$

where D_1 is a constant, and the exponent $\gamma > 2/5$ depends only upon the decay rate of $c[n,\tau]$ as a function of τ. Moreover, if the penalization constant C_p is sufficiently large, independently of N, then there exists $D_2 > 0$ and $\beta < 3$ such that

$$\|\Gamma - P_{\widehat{\mathcal{M}}}(\widehat{\Gamma})\|^2 \leq D_2 (\log N)^\beta \text{Risk}_{\min}$$

(a)

(b)
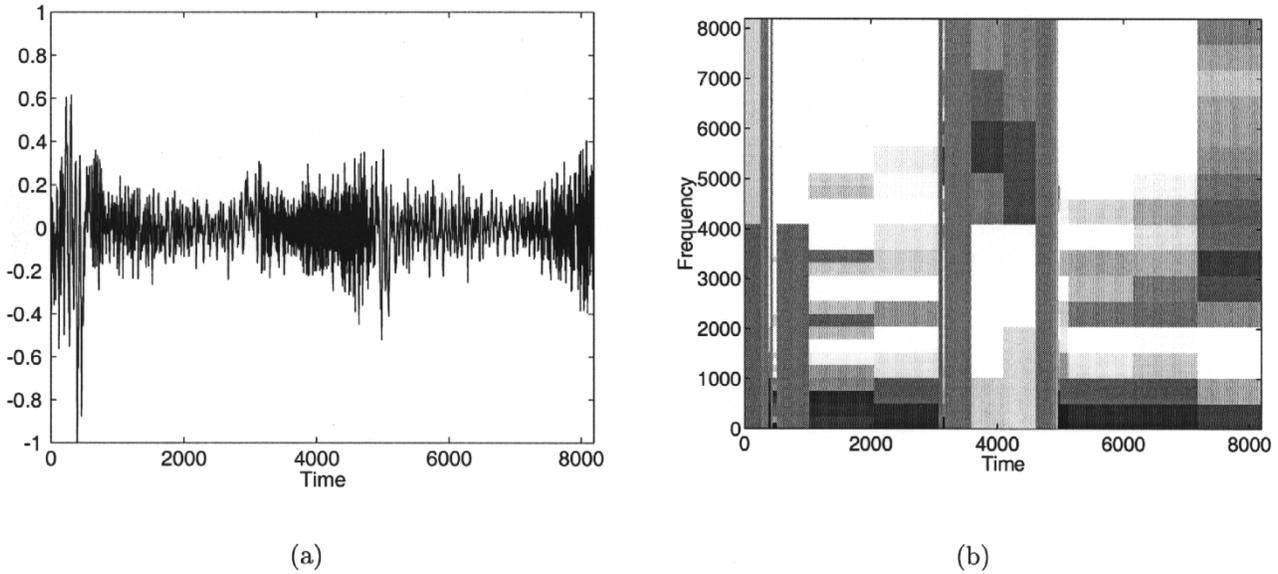
Fig. 4.   (a) Unvoiced speech signal composed of $[p]$, $[\infty]$, $[s]$, $[\infty]$, and $[ch]$ sampled at 11 kHz. (b) Penalized macrotile estimator of the signal above calculated with $C_p = 5\,10^{-4}$.

with a probability that tends to 1 when N goes to ∞. When N increases, the rate of convergence given by (30) is always better than what can be obtained with a linear estimator that does not adapt the basis to the specific process [8]. This linear strategy is most often implemented in a local cosine basis or with a window Fourier transform, which decomposes $[0, N]$ in intervals of same size, chosen independently from the realizations of the process [20].

D. Estimation of Long-Range Locally Stationary Covariances

The previous section has studied locally stationary processes that have a relatively fast decorrelation and whose time-varying spectrum is therefore uniformly smooth as a function of frequency. The macrotile models of Section IV-B perform a uniform averaging in frequency, which is consistent with this smoothness. Some locally stationary processes may not satisfy this fast decorrelation property. This means the macrotile segmentations of Section IV-B, which correspond to a uniform averaging in frequency, must be replaced by more flexible segmentations, which allow nonuniform frequency averaging within each time interval. This is exactly what is performed by the family of macrotile models of Section III-B, which are used for noise removal in sounds.

The macrotile covariance estimator is calculated by computing the best model

$$\widehat{\mathcal{M}} = \arg \min_{\mathcal{M} \in \mathcal{F}} \left\{ -\|P_{\mathcal{M}}(\widehat{\Gamma})\|^2 + C_p \frac{\lambda^2}{R} \left(\log N \right)^2 \dim(\mathcal{M}) \right\}.$$
(31)

In this case, the family of local cosine macrotile models $\mathcal{F}$ includes adaptive piecewise-constant segmentations in each time interval. Therefore, the fast algorithm of Section IV-B has to be modified when computing the optimal segmentation S_p^j of the partial models $\mathcal{M}_p^{j*}$ at each node of the local cosine tree. Since the dictionary is formally identical as the one used for the denoising problem, the same CART tree algorithm is used to find

the optimal adaptive segmentation. The resulting fast algorithm requires $O(N(\log_2 N)^2)$ operations to calculate $\widehat{\mathcal{M}}$.

Fig. 5 shows numerical results with a process constructed by aggregating independent stationary processes over the intervals $[0,80]$, $[80\,180]$, $[180\,430]$, $[430\,830]$, $[830\,930]$, and $[930,1024]$. On each interval, the process is the sum of two independent stationary processes: one whose spectrum includes a narrow frequency spike and one whose spectrum is uniformly regular. Fig. 5(a) shows one realization of this spliced process. Fig. 5(b) shows the diagonal coefficients of Γ in the ideal local cosine basis $\mathcal{B}$, where Γ is best approximated by a diagonal matrix $\Gamma_{\mathcal{B}}$: $\|\Gamma - \Gamma_{\mathcal{B}}\|^2 / \|\Gamma\|^2 = 0.10$. Fig. 5(d) represents the penalized macrotile estimator $P_{\widehat{\mathcal{M}}}(\widehat{\Gamma})$ computed with (31): $\|\Gamma - \widehat{\Gamma}_{\widehat{\mathcal{M}}}\|^2 / \|\Gamma\|^2 = 0.23$. Note that the frequency spikes observed in Fig. 5(b) are retrieved in Fig. 5(d); this is possible because the adaptive smoothing allows for macrotiles of very different sizes in the same time interval. At fixed time n, the macrotile model produces a piecewise-constant approximation of the true spectrum. Fig. 5(e) and (f) show that the frequency spikes and the smoother parts are well approximated, given the small number of realizations. To illustrate the importance of the penalization, Fig. 5(c) represents the macrotile estimator $P_{\widehat{\mathcal{M}_0}}(\widehat{\Gamma})$ computed from the same $R = 5$ realizations for a penalization constant $C_p = 0$, which means that the choice of model is not penalized by its dimension. As a result, a model of maximum dimension is chosen, namely many intervals of stationary and no averaging in frequency. The corresponding risk is much larger: $\|\Gamma - P_{\widehat{\mathcal{M}_0}}(\widehat{\Gamma})\|_H^2 / \|\Gamma\|_H^2 = 0.45$.

This penalized macrotile algorithm has also been tested on a voiced-speech recording shown in Fig. 6, which is composed of five vowels. The number of realizations is, of course, $R = 1$. The corresponding macrotile penalized estimator is shown in Fig. 6(b). The stationarity intervals can be distinguished and the harmonics are clearly retrieved. Small intervals at the borders of different sounds in Fig. 6(b) are due to the dyadic nature of admissible segmentations.
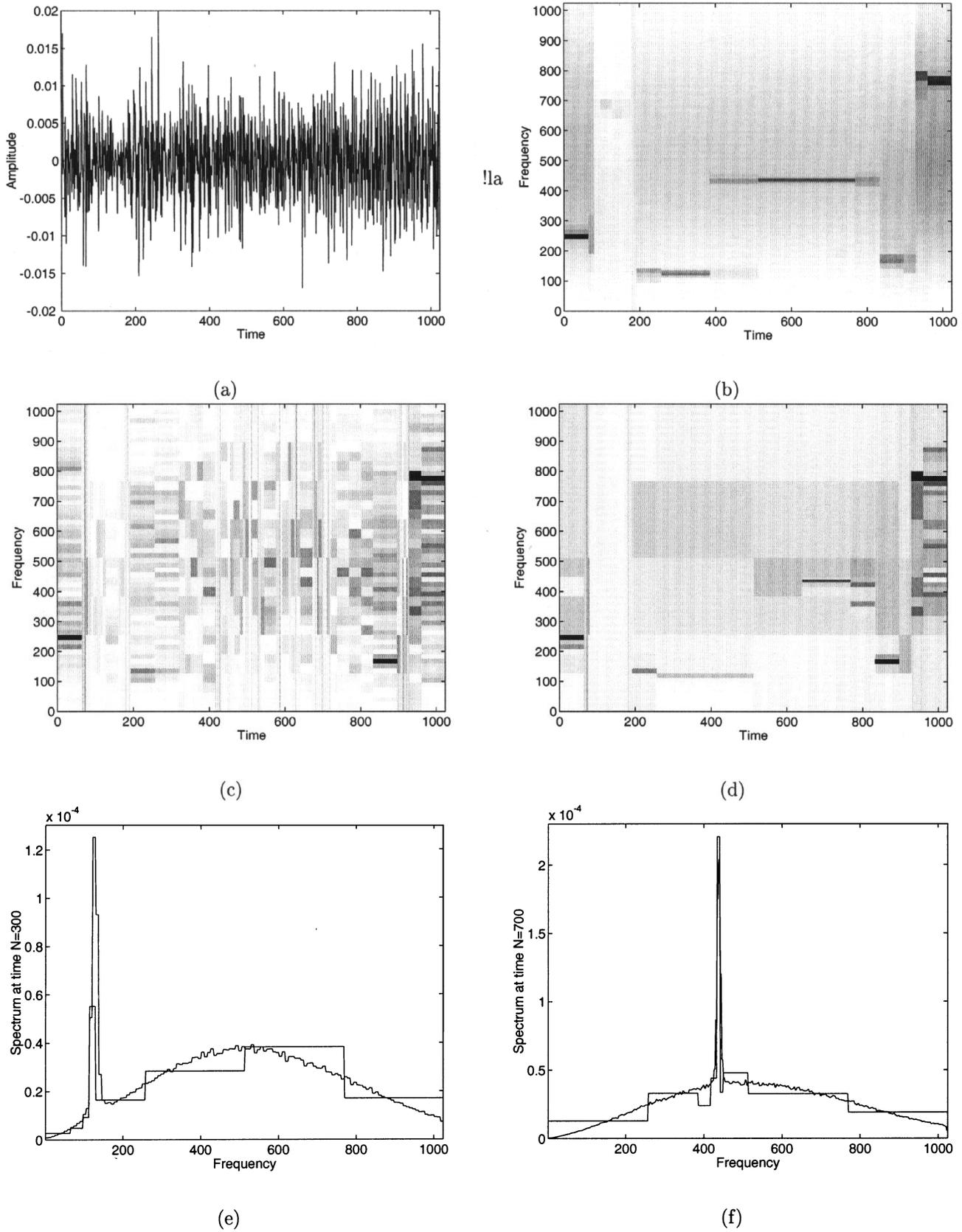
Fig. 5.   Piecewise stationary process of size $N = 1024$. (a) One realization of the process. (b) Ideal diagonal covariance. (c) Macrotile estimator $\widehat{\Gamma}_{\widehat{\mathcal{M}_0}}$ computed from $R = 5$ realizations with no penalization ($C_p = 0$). (d) Macrotile estimator $\widehat{\Gamma}_{\widehat{\mathcal{M}}}$ computed with a penalization. (e), (f) Spectrum of the spliced process at $n = 300$ and $n = 700$ are superimposed with the piecewise-constant estimations of the macrotile model in (d). The true spectrum is approximated by an estimator using a very large number of realizations.

(a)

(b)
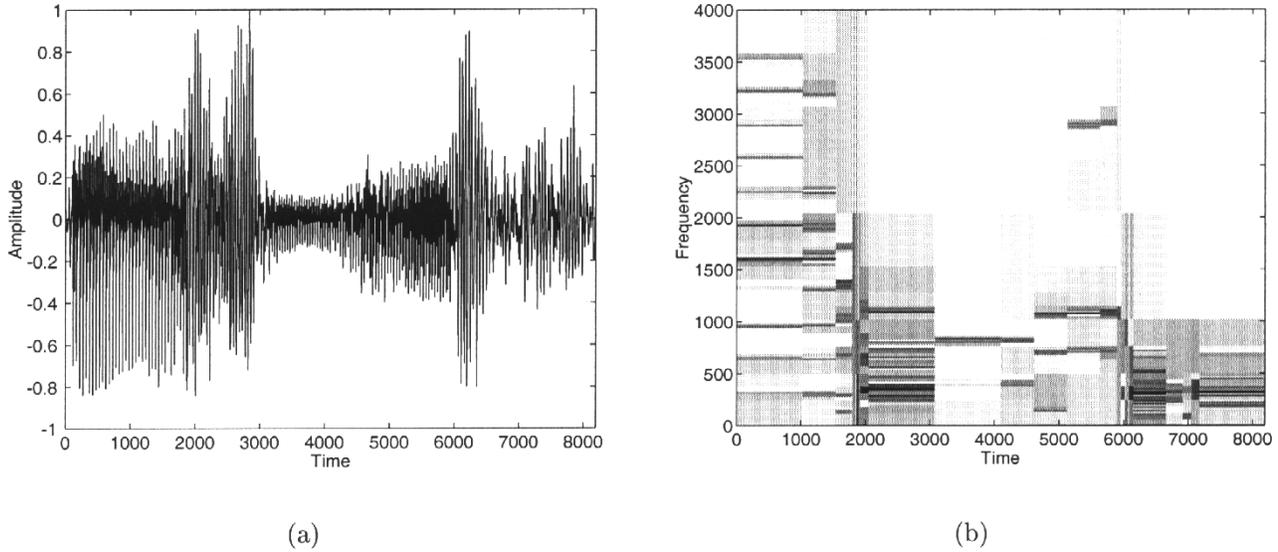
Fig. 6. (a) Voiced speech signal composed of $[a]$, $[\mathit{oe}]$, $[i]$, $[o]$, and $[y]$ (phonetic writing), sampled at 5.5 kHz. (b) Penalized macrotile estimator of the signal above calculated with $C_p = 10^{-5}$.

V. CONCLUSION

This paper introduces adaptive macrotile models to estimate covariances of nonstationary processes. These models are selected with a penalization algorithm. Macrotile estimations provide general estimation procedures that apply to other problems such as the removal of additive noise. For noise removal, macrotile estimations can be interpreted as a best basis denoising algorithm. This simpler problem is thus used as an introduction to better understand the properties of macrotile estimations. An application to noise removal in sounds is presented with macrotiles constructed with the local cosine dictionary. These macrotiles define a coarse pavement of the time–frequency plane, which is adapted to the signal time–frequency properties.

The main result of the paper concerns the covariance estimation of locally stationary processes. A fast algorithm is presented to compute the penalized macrotile estimator with $O(N(\log N)^2)$ operations. For processes with a fast decorrelation, the statistical consistency of this estimator has been proved [8], but not for complex processes whose spectrum include narrow frequency spikes. Although the numerical experiments are encouraging, the proof of statistical consistency remains an open problem in this case.

APPENDIX

A. Proof of (8)

We saw in (26) that

$$P_{\mathcal{M}}(\widehat{\Gamma}) = \sum_{k=1}^{K} \frac{1}{I_k^{\#}} \left(\sum_{m \in I_k} \langle \widehat{\Gamma} g_m^b, g_m^b \rangle \right) P_{\mathbf{V}_k}.$$

We can also derive from (28) that

$$\sum_{m \in I_k} \langle \widehat{\Gamma} g_m^b, g_m^b \rangle = \sum_{m \in I_k} \frac{1}{R} \sum_{r=1}^{R} |\langle \widehat{y}_r, g_m^b \rangle|^2$$

$$= \frac{1}{R} \sum_{r=1}^{R} \| P_{\mathbf{V}_k}(\widehat{y}_r) \|^2 \qquad (32)$$

so that

$$P_{\mathcal{M}}(\widehat{\Gamma}) = \sum_{k=1}^{K} \frac{1}{I_k^{\#}} \left(\frac{1}{R} \sum_{r=1}^{R} \| P_{\mathbf{V}_k}(\widehat{y}_r) \|^2 \right) P_{\mathbf{V}_k}.$$

We use the orthogonality of the spaces $\mathbf{V}_k$ and the independence of the R realizations $\widehat{y}_r$, $r = 1, \ldots, R$ to show that since $\widehat{\epsilon} = \widehat{\Gamma} - \Gamma$

$$\mathbf{E}\{\| P_{\mathcal{M}}(\widehat{\epsilon}) \|^2\} = \mathbf{E}\{\| P_{\mathcal{M}}(\widehat{\Gamma}) - P_{\mathcal{M}}(\Gamma) \|^2\}$$

$$= \frac{1}{R} \sum_{k=1}^{K} \frac{1}{I_k^{\#}} \text{Var}(\| P_{\mathbf{V}_k}(\widehat{y}) \|^2) \qquad (33)$$

where $\widehat{y}$ is a Gaussian vector whose covariance matrix Γ admits λ as a largest eigenvalue. For each $k \in [1, K]$, $P_{\mathbf{V}_k}(\widehat{y})$ is a Gaussian random vector of dimension $I_k^{\#}$. The eigenvalues of its covariance are clearly smaller than or equal to λ. Rewriting it in the Karhunen–Loève basis of its covariance gives

$$\| P_{\mathbf{V}_k}(\widehat{y}) \|^2 = \sum_{l=1}^{I_k^{\#}} \mu_l \, \widehat{w}_l^2$$

where $\{\widehat{w}_l\}_{1 \le l \le I_k^{\#}}$ are independent Gaussian random variables of variance 1, and $|\mu_l| \le \lambda$ for each l. Therefore

$$\text{Var}(\| P_{\mathbf{V}_k}(\widehat{y}) \|^2) = \sum_{l=1}^{I_k^{\#}} |\mu_l|^2 \, \text{Var}(\widehat{w}_l^2) = 2 \sum_{l=1}^{I_k^{\#}} |\mu_l|^2 \le 2 \, \lambda^2 \, I_k^{\#}. \qquad (34)$$

Combining (33) and (34) proves (8).

REFERENCES

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

[2] ——, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comput. Rend. Acad. Sci. Paris Ser A*, vol. 319, pp. 1317–1322, 1994.

[3] S. G. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Statist.*, vol. 26, pp. 1–47, 1998.

[4] D. L. Donoho, S. Mallat, and R. von Sachs, "Estimating covariances of locally stationary processes: Rates of convergence of best basis methods," Stanford Univ., Statistics Dept., 517, 1998.

[5] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 214–224, Feb. 1995.

[6] M. B. Priestley, "Evolutionary spectra and nonstationary processes," *J. R. Statist. Soc. B*, vol. 27, pp. 204–237, 1965.

[7] Y. Samuelides, "Macrotile estimation and market model with jumps," Ph.D. dissertation, Ecole Polytechnique, Paris, France, 2001.

[8] Y. Samuelides and S. Mallat, "Penalized macrotile estimation of locally stationary covariances," Ann. Statist., 2002, submitted for publication.

[9] A. R. Barron, L. Birgé, and P. Massart, "Risk bounds for model selection via penalization," *Probab. Theory Relat. Fields*, vol. 113, pp. 301–413, 1999.

[10] A. R. Barron and T. M. Cover, "Minimum complexity density estimation," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1034–1054, July 1991.

[11] R. R. Coifman and Y. Meyer, "Remarques sur l'analyze de Fourier à fenêtre," *Comput. Rend. Acad. Sci. Paris Ser.A*, vol. 312, pp. 259–261, 1991.

[12] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without block effects," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 553–559, Apr. 1989.

[13] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.

[14] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. London, U.K.: Chapman & Hall, 1993.

[15] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best-basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.

[16] R. Dahlhaus, "Fitting time series models to nonstationary processes," *Ann. Statist.*, vol. 25, pp. 1–37, 1997.

[17] M. H. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra," *Ann. Statist.*, vol. 25, pp. 38–76, 1997.

[18] M. B. Priestley, *Spectral Analysis and Time Series*. Boston: Academic, 1981.

[19] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1461–1470, 1985.

[20] K. S. Riedel, "Optimal data-based kernel estimation of evolutionary spectra," *IEEE Trans. Signal Process.*, vol. 41, pp. 2439–2447, July 1993.

David L. Donoho received the A.B. degree (summa cum laude) in statistics from Princeton University, Princeton, NJ, and the Ph.D. degree in statistics from Harvard University, Cambridge, MA.

He is a Professor of statistics at Stanford University, Stanford, CA. He was a Professor at the University of California, Berkeley, and a Visiting Profeoosr at the Université de Paris, Paris, France, as well as a Sackler Fellow at Tel Aviv University, Tel Aviv, Israel. His research interests are in harmonic analysis, image representation, and mathematical statistics.

Dr. Donoho is a member of the U.S. National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences.

Stéphane Mallat was born in Paris, France. He graduated from Ecole Polytechnique, Paris, in 1984 and from Ecole Nationale Supérieure des Télécommunications, Paris, in 1985. He received the Ph.D. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1988.

In 1988, he joined the Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, NY, became Associate Professor in 1993, and is now a Research Professor. In the fall 1994, he was a Visiting Professor with the Electrical Engineering Department, Massachusetts Institute of Technology, Cambridge, and in the spring 1994 with the Applied Mathematics Department, University of Tel Aviv, Tel Aviv, Israel. Since 1995, he has been a Professor with the Applied Mathematics Department at Ecole Polytechnique, Paris, and is now Chairman of the department. His research interest include computer vision, signal processing, and diverse applications of wavelet transforms. He is the author of the book *A Wavelet Tour of Signal Processing* (New York: Academic, 1998).

Dr. Mallat received the 1990 IEEE Signal Processing Society's paper award, the 1993 Alfred Sloan fellowship in mathematics, the 1997 Outstanding Achievement Award from the SPIE Optical Engineering Society, and the 1997 Blaise Pascal Prize in applied mathematics from the French Academy of Sciences.

Rainer von Sachs received the Diploma in mathematics from the University of Freiburg, Freiburg, Germany, in 1987 and the Ph.D. degree in mathematics (statistics) from the University of Heidelberg, Heidelberg, Germany, in 1991.

From 1988 to 1992, he was a Research Fellow with the University of Heidelberg and, from 1992 to 1998, Assistant Professor of mathematics at the University of Kaiserslautern, Kaiserslautern, Germany. He is currently Professor of statistics at the Université Catholique de Louvain, Louvain-la-Neuve, Belgium. His research interests include mathematical statistics, nonparametric curve estimation, analysis and modeling of nonstationary time series, spectral density estimation, and wavelet methods.

Yann Samuelides was born in Paris, France. He graduated from Ecole Polytechnique, Paris, in 1995 and from Ecole des Mines de Paris in 1998. He received the Ph.D degree in applied mathematics from Ecole Polytechnique in 2001.

In 1997, he was a visiting scholar with the Mathematics Department, Stanford University, Stanford, CA. From 1998 to 2000, he was a research fellow with the Applied Mathematics Department at Ecole Polytechnique. From 2000 to 2002, he was a vice president with the Fixed Income Research and Strategy Team at BNPParibas, London, U.K. and New York, NY. Since 2002, he has been a Director, Global Head of Credit, Options, and Swaps Research at Credit Agricole Indosuez, Paris. His research interests include applications of probabilistic models in mathematical finance, statistical signal processing, and geophysics.