Centralized and Distributed Semi-Parametric Compression of Piecewise Smooth Functions

Varit Chaisinthop*, *Member, IEEE,* and Pier Luigi Dragotti, *Member, IEEE*

EDICS: DSP-WAVL (Wavelets theory and applications) and SEN-DCSC (Distributed channel and source coding)

Abstract

This paper introduces novel wavelet-based semi-parametric centralized and distributed compression methods for a class of 1-D piecewise smooth functions. Classical centralized compression schemes are based on a relatively complex, nonlinear encoder and a simple, linear decoder. Recently, a new paradigm in compression called distributed source coding has emerged. This setup involves multiple encoders, where each one partially observes the source, and a centralized decoder. First, we focus on the dual situation of the centralized compression with a simple encoder and a complex decoder. We show that, by incorporating parametric estimation into the decoding procedure, it is possible to achieve the same rate-distortion performance as that of a conventional wavelet-based compression scheme. Second, we consider the distributed compression scenario, where each independent encoder partially observes the 1-D piecewise smooth function. We propose a new wavelet-based distributed compression scheme that uses parametric estimation to perform joint decoding. Our analysis shows that it is possible for the proposed scheme to achieve the same compression performance as that of a joint encoding scheme.

Index Terms

Wavelets, piecewise smooth functions, approximations, compression, distributed source coding, parametric estimation.

The material in this paper was presented in part at the European Conference on Signal Processing (EUSIPCO), Glasgow, UK, August 2009 [1].

V. Chaisinthop is with the Electrical and Electronic Engineering Department, Imperial College London, London SW7 2AZ, U.K. (e-mail: varit.chaisinthop@gmail.com)

P. L. Dragotti is with the Electrical and Electronic Engineering Department, Imperial College London, London SW7 2AZ, U.K. (e-mail: p.dragotti@imperial.ac.uk).

Draft June, 2010

I. INTRODUCTION

Over the last two decades, transform coding has emerged as the dominating compression strategy due to its high efficiency and low complexity. Given an observed source $f \in \mathbf{L_2}(\mathbb{R})$, where $\mathbf{L_2}(\mathbb{R})$ indicates the space of square integrable function, the linear transform decomposes f over the basis $\mathcal{B} = \{g_m\}$ of $\mathbf{L_2}(\mathbb{R})$. The transform coefficients are then quantized and entropy coded. This process is simply reversed at the decoder. The key problem then is to minimize the mean-squared error (MSE) distortion given by $D = \mathrm{E}\left[\|f - \widehat{f}\|^2\right] = \mathrm{E}\left[\int_{-\infty}^{\infty}(f(t) - \widehat{f}(t))^2 dt\right]$ for a given rate R. Many important results and optimality conditions of transform coding have been derived (see [2]). For example, when the source is Gaussian, it is well known that the Karhunen-Loève Transform (KLT) is the optimal transform [2], [3].

Recently, a new paradigm in compression called distributed source coding (DSC) (see [4]–[10]) has started to emerge. In contrast to the centralized scenario, the source is partially observed by a number of independent encoders. The observations, however, can be jointly decoded. Such scenario imposes a new set of requirements for compression, which typically includes low-complexity encoders, robustness and high compression efficiency [4]. It is then natural to wonder how the classical centralized transform coding strategy is going to change under this new scenario. Recent research has provided us with some precise answers. For example, the KLT has been shown to be the best transform for Gaussian sources [11] and the optimality conditions for transforms in high bit-rate regimes are given in [12], [13]. If, however, the Gaussian and high bit rate assumptions are relaxed then the problem of distributed transform coding remains largely open. It is also of interest to note that the shift in complexity from the encoder to the decoder is also central to a set of new sampling theories such as compressed sensing (CS) [14], [15] and sampling signals with finite rate of innovation (FRI) [16], [17]. Compressed sensing has also led to new DSC results [18], [19]. The work in this paper has been inspired by these recent sampling results; it is however closer to FRI theory since the signals considered here are closer to those studied in [17].

In this paper, we will focus on the wavelet transform, which has had a profound impact on modern signal processing theory (see [20]). In particular, we focus on the theoretical study of the performance of wavelets in compression schemes with simple encoders and a complex decoder. Our study is based on the compression of a class of piecewise smooth functions, which is usually used as a simplified model of a row (or a column) of an image [21]. For a conventional centralized transform coding setting, it has been shown that a wavelet-based compression strategy that employs a nonlinear encoder produces the best R-D performance [21]. The problem of finding the best strategy for a wavelet-based distributed compression, however, is still open.

In order to investigate the impact of the structural change in complexity, we first focus on the centralized scenario with a simple encoder and a non-linear decoder. By including a form of nonlinear parametric estimation technique in the decoding process, we will show that it is indeed possible to achieve a comparable R-D performance to that of the traditional scheme. We then investigate how the concept of nonlinear decoding can be applied in the DSC setting. In particular, we focus on the case where the disparity between each observed signal can be described by a form of geometric transformation. This is, in spirit, similar to the use of motion compensated prediction algorithm in video compression. We will also propose a distributed compression scheme that uses the FRI-based parametric estimation technique. Finally, we briefly discuss how the theoretical results in this paper can be extended to real images.

The paper is organized as follows. In the next section, we review the notion of piecewise smooth functions including their wavelet-based approximation and compression results. Section III presents our proposed semi-parametric compression algorithm for the centralized case, where a detailed R-D analysis is given. The concept of semi-parametric compression is then extended to the distributed compression scenario in Section IV. A constructive parametric estimation algorithm based on the sampling theory of signals with finite rate of innovation is presented in Section V. We give our simulation results in Section VI. A brief discussion on application to real images is then presented in Section VII. Finally, the conclusion is given in Section VIII.

II. Signal model, approximation and compression

A. Signal Model

Throughout this paper, we will assume that the support of the continuous function $f(t)$ is normalized to $t \in [0,1[$ where $t \in [a,b[$ denotes $a \leq t < b$. We define the regularity of a function with the Lipschitz exponent [22]. Recall that a function $f(t)$ restricted to $[a,b]$ is said to be uniformly α-Lipschitz over $[a,b]$, with $\alpha \geq 0$, if it can be written as $f(t) = p(t) + \epsilon(t)$, where $p(t)$ is a polynomial of degree $m = \lfloor \alpha \rfloor$ and there exists a constant $K > 0$ such that $\forall t \in (a,b)$ and $\forall \nu \in [a,b]$, $|\epsilon(t)| \leq K|t-\nu|^{\alpha}$. A piecewise smooth function $f(t)$, $t \in [0,1[$ with $K+1$ pieces is then defined as

$$f(t) = \sum_{i=0}^{K} f_i(t) \mathbf{1}_{[t_i, t_{i+1}[}(t) \quad \text{with} \quad \mathbf{1}_{[a,b[}(t) = \begin{cases} 1 & \forall t \in [a,b[, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where $t_0 = 0$, $t_{K+1} = 1$ and $f_i(t)$ is uniformly α-Lipschitz over $[t_i, t_{i+1}]$.

Given a function $f(t)$ as defined in (1), it was shown in [23] that $f(t)$ can be decomposed into two functions: a piecewise polynomial function $f_p(t)$ with pieces of maximum degree $\lfloor \alpha \rfloor$ and a globally

α-Lipschitz smooth function $f_\alpha(t)$. Hence, our piecewise smooth functions can be written as follows:

$$f(t) = f_p(t) + f_\alpha(t). \tag{2}$$

The piecewise polynomial signal can be written as follows:

$$f_p(t) = \sum_{k=0}^{K} \sum_{r=0}^{\lfloor \alpha \rfloor} a_{r,k}(t - t_k)_+^r \tag{3}$$

with $t_+ = \max(t, 0)$. This signal model will be used in our analysis in the forthcoming sections.

B. Wavelet Representation and Approximation

Let us recall that the wavelet decomposition of a continuous function $f(t)$, $t \in [0, 1[$, is formally expressed as

$$f(t) = \sum_{n=0}^{L-1} c_{J,n}\varphi_{J,n}(t) + \sum_{j=-\infty}^{J} \sum_{n=0}^{2^{-j}-1} d_{j,n}\psi_{j,n}(t), \tag{4}$$

where $J < 0$, $L = 2^{-J}$, $\varphi_{J,n}(t) = 2^{-J/2}\varphi\left(2^{-J}t - n\right)$ and $\psi_{j,n}(t) = 2^{-j/2}\psi\left(2^{-j}t - n\right)$. The low-pass $\{c_{J,n}\}$ and high-pass $\{d_{j,n}\}$ coefficients are given by the following inner products: $c_{J,n} = \langle f(t), \tilde{\varphi}_{J,n}(t)\rangle$ and $d_{j,n} = \left\langle f(t), \tilde{\psi}_{j,n}(t)\right\rangle$, where $\tilde{\varphi}_{J,n}$ and $\tilde{\psi}_{j,n}$ are the dual of $\varphi_{J,n}$ and $\psi_{j,n}$ respectively such that $\langle \tilde{\varphi}_{j,m}, \varphi_{j,n}\rangle = \delta_{m,n}$ and similarly for $\tilde{\psi}_{j,n}$.

One of the most well known properties of the wavelet transform is the *vanishing moments* property. The wavelet transform is said to have $(P + 1)$ vanishing moments if its analysis wavelet $\tilde{\psi}(t)$ satisfies $\int_{-\infty}^{\infty} t^p \tilde{\psi}(t)dt = 0$, $\forall p \in \{0, 1, ..., P\}$. Moreover, given a function that is uniformly α-Lipschitz around v and a wavelet with at least $\lfloor \alpha + 1 \rfloor$ vanishing moments, the standard result in wavelet theory states that the wavelet coefficients in the cone of influence of v decays as $d_{j,n} \sim A2^{j(\alpha+1/2)}$ across scales with a constant $A > 0$. We refer to [20], [22], [24]–[26] for a detailed treatment on wavelet theory.

The N-term linear approximation of $f(t)$ can then be obtained by representing $f(t)$ with only N coefficients where the choice of these coefficients is fixed a priori. Normally, the first N coefficients are retained. If we assume that $N \sim 2^{J_N}$ then the following approximation can be obtained:

$$f_N(t) = \sum_{n=0}^{L-1} c_{J,n}\varphi_{J,n}(t) + \sum_{j=J-J_N+1}^{J} \sum_{n=0}^{2^{-j}-1} d_{j,n}\psi_{j,n}(t), \tag{5}$$

with $J_N \geq 0$ and $J < 0$. Note that this is also referred to as linear multiresolution approximation. Let us denote the squared approximation error with $\varepsilon_l(N, f) = \|f(t) - f_N(t)\|^2$. Given that $f(t)$ is uniformly α-Lipschitz over $[0, 1]$ and the wavelet has at least $\lfloor \alpha + 1 \rfloor$ vanishing moments, it can be shown that the error decays as $\varepsilon_l(N, f) \sim N^{-2\alpha}$ [22]. On the contrary, if $f(t)$ is piecewise α-Lipschitz smooth with

K pieces as given in (1) then $\varepsilon_l(N, f) \sim KN^{-1}$ [22]. If we now approximate $f(t)$ with the N largest amplitude coefficients instead, then the best approximation of $f(t)$ is given by

$$f_{\mathcal{I}_N}(t) = \sum_{n \in \mathcal{I}_N} c_{J,n}\varphi_{J,n}(t) + \sum_{(j,n) \in \mathcal{I}_N} d_{j,n}\psi_{j,n}(t),$$

where $\mathcal{I}_N$ is the index set of the N largest amplitude coefficients. This is a form of nonlinear approximation since the index set $\mathcal{I}_N$ depends on $f(t)$. Given the same piecewise α-Lipschitz smooth function, it now follows that $\varepsilon_n(N, f) \sim N^{-2\alpha}$ [22]. Therefore, nonlinear approximation is superior to linear approximation when the function is piecewise smooth.

C. Wavelet Compression

One can think of compression as a process of approximation followed by quantization. Suppose that we are to compress a smooth α-Lipschitz function. Given that the wavelet has at least $\lfloor \alpha + 1 \rfloor$ vanishing moments, the distortion-rate function $D(R)$ of a compression scheme that allocates the bits to the first N coefficients is $D(R) \leq c_1 R^{-2\alpha}$ [21]. If, however, a function is piecewise smooth as described by (1), the achieved $D(R)$ becomes $D(R) \leq c_2 R^{-2\alpha} + c_3 R^{-1}$. On the other hand, a compression scheme that allocates bits to the N largest coefficients achieves $D(R) \leq c_4 R^{-2\alpha} + c_5 \sqrt{R} 2^{-c_6 \sqrt{R}}$ [21]. Thus, at high rates, the distortion of a nonlinear compression scheme decays as $R^{-2\alpha}$ whereas a scheme with a linear approximation strategy has a slower decay of R^{-1}.

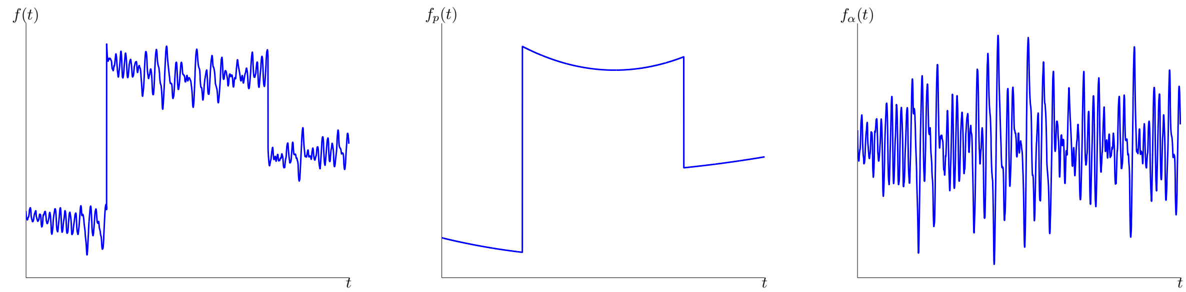
III. Centralized Semi-Parametric Compression

We now consider the setup where the encoder is based on linear approximation and the decoder is nonlinear, which is the dual of the traditional compression.

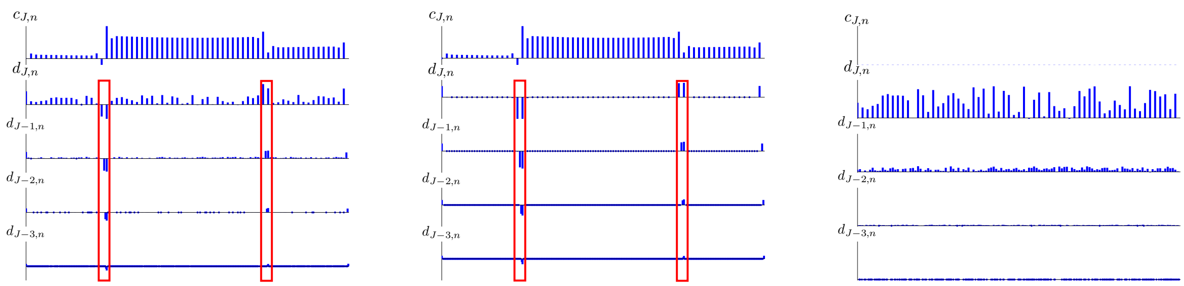
A. Semi-Parametric Compression Strategy

Consider a piecewise smooth function $f(t)$ given by the signal model in (2). Intuitively, one can recover $f(t)$ by reconstructing $f_p(t)$ and $f_\alpha(t)$ separately. Since $f_\alpha(t)$ is uniformly α-Lipschitz, a compression method based on the linear approximation shown in (5) can be used to compress $f_\alpha(t)$ with $D(R) \sim R^{-2\alpha}$. On the other hand, we can reconstruct $f_p(t)$ by estimating the locations t_i and the polynomial coefficients $a_{r,k}$, which is a parametric estimation problem.

First, consider the wavelet decomposition of $f(t)$ as shown in (4) where we assume that the wavelet has at least $\lfloor \alpha + 1 \rfloor$ vanishing moments. We denote with $\mathcal{I}_p$ a set of indices such that $\mathcal{I}_p = \{(j, n) \in \mathbb{Z} : |\langle f_p(t), \psi_{j,n}(t) \rangle > 0\}$. In other words, the coefficients $\{d_{j,n}\}_{(j,n) \in \mathcal{I}_p}$ are in the *cone of influence* of



(a) (b) (c)



(d) (e) (f)

Fig. 1. (a) A piecewise smooth function $f(t) = f_p(t) + f_\alpha(t)$; (b) the piecewise polynomial function $f_p(t)$; (c) the smooth α-Lipschitz function $f_\alpha(t)$; (d) coefficients of $f(t)$; (e) coefficients of $f_p(t)$; (f) coefficients of $f_\alpha(t)$. Note that, the boxes represent the cone of influence of discontinuities as denoted by the index set $\mathcal{I}_p$.

discontinuities of $f_p(t)$. In contrast, the coefficients in $\{d_{j,n}\}_{(j,n)\notin\mathcal{I}_p}$ are outside the cone of influence and decay as $d_{j,n} \sim 2^{j(\alpha+1/2)}$ This is illustrated in Fig. 1.

Let us now address the linear approximation-based quantization strategy in details. From (5), $f_\alpha(t)$ is approximated with only the coefficients in decomposition level $j = J - J_N + 1, ..., J$. Since the wavelet coefficients of $f_\alpha(t)$ decays as $d_{j,n} \leq A2^{j(\alpha+1/2)}$, this is equivalent to setting the quantizer step size to $\Delta = A2^{(J-J_N+1)(\alpha+1/2)}$. Therefore, the number of bits allocated to each coefficient at resolution 2^{-j} is given by $R_{j,\alpha} = \left\lceil \log_2 \left(A2^{j(\alpha+1/2)}/\Delta \right) \right\rceil + 1 = \lceil (J_N - J + j)(\alpha + 1/2) + 1 \rceil$, $j = J - J_N + 1, ..., J$, where an extra bit is needed to code the sign. Hence, the total rate allocated to the high-pass coefficients $d_{j,n}$ for the reconstruction of $f_\alpha(t)$ is $R_1 = \sum_{j=J-J_N+1}^{J} 2^{-j} R_{j,\alpha}$.

In reality, the encoder does not have a direct access to the smooth component $f_\alpha(t)$. This is not a problem for the coefficients $\{d_{j,n}\}_{(j,n)\notin\mathcal{I}_p}$ outside the cone of influence of discontinuities as

$$d_{j,n} = \langle f(t), \psi_{j,n}(t) \rangle = \langle f_\alpha(t), \psi_{j,n}(t) \rangle + \underbrace{\langle f_p(t), \psi_{j,n}(t) \rangle}_{=0} = \langle f_\alpha(t), \psi_{j,n}(t) \rangle, \quad (j,n) \notin \mathcal{I}_p.$$

Hence the above quantization strategy can be applied directly. For the coefficients $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$ in the cone of influence of discontinuities, it follows that

$$d_{j,n} = \langle f_\alpha(t), \psi_{j,n}(t)\rangle + \langle f_p(t), \psi_{j,n}(t)\rangle\,, \quad (j,n) \in \mathcal{I}_p.$$

Since $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$ do not decay as $2^{j(\alpha+1/2)}$ across scales, the values of $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$ are outside the range of the quantizer i.e. for the same Δ, more than $R_{j,\alpha}$ bits are required to code the coefficients in the set $\mathcal{I}_p$. However, it is true that the information of the wavelet coefficients of $f_\alpha(t)$ is fully contained within the first $R_{j,\alpha}$ bits (starting from the less significant bits (LSB)). Therefore, for $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$, only the first $R_{j,\alpha}$ LSBs are transmitted to the decoder and the rest of the bits can be discarded.

As mentioned earlier, the reconstruction of the piecewise polynomial component $f_p(t)$ can be done parametrically. We recall that the wavelet coefficients of $f(t)$ are given by $d_{j,n} = \langle f_\alpha(t), \psi_{j,n}(t)\rangle + \langle f_p(t), \psi_{j,n}(t)\rangle$, $(j,n) \in \mathcal{I}_p$. Let us denote with $\hat{f}_p(t)$, the reconstructed version of $f_p(t)$, which is obtained from parametric estimation, and let $\hat{d}_{j,n} = \langle \hat{f}_p(t), \psi_{j,n}(t)\rangle$ be the corresponding wavelet coefficients. In addition, by following the above quantization strategy, the decoder receives the first $R_{j,\alpha}$ LSBs of $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$. We denote with $\bar{d}_{j,n}$, the quantized version of $d_{j,n}$. We will now show that $\{d_{j,n}\}_{(j,n)\in\mathcal{I}_p}$ can be decoded using the concept of error correction code. Let us demonstrate this concept with the following example.

Consider the case where $R_{J,\alpha} = 2$ bits for a given J. Suppose that the binary representation of $\bar{d}_{J,n}$ is 1111 and that, at the decoder, parametric estimation gives $\hat{d}_{J,n} = 1110$. The encoder sends the first $R_{J,\alpha}$ LSBs of $\bar{d}_{J,n}$, which is 11, to the decoder. Since $d_{J,n} = \langle f_\alpha(t), \psi_{j,n}(t)\rangle + \langle f_p(t), \psi_{j,n}(t)\rangle$, this means that the quantized version of $\langle f_\alpha(t), \psi_{J,n}(t)\rangle$ can either take a positive value of 1 or a negative value of 111 (this is without the sign bit attached), which would lead to $\bar{d}_{J,n} = 1111$ or $\bar{d}_{J,n} = 111$ respectively. However, since the decoder knows that the magnitude of $\langle f_\alpha(t), \psi_{j,n}(t)\rangle$ is bounded by $R_{J,\alpha} - 1 = 1$ bit, it can select $\bar{d}_{J,n} = 1111$ as the correct decoded value. This is, in spirit, similar to the use channel coding technique in Wyner-Ziv coding [6], [9], where the first $R_{J,\alpha}$ bits is the equivalent of the coset.

The proposed 'semi-parametric' compression algorithm for a piecewise smooth function can now be outlined as follows:

Algorithm 1 (centralized semi-parametric compression):

Encoding:

1: N**-term linear approximation**: the encoder approximates $f(t)$ as shown in (5);
2: **Quantization**: the coefficients $\{c_{J,n}\}$ and $\{d_{j,n}\}_{J-J_N+1\leq j\leq J}$ are quantized using a linear approximation-based quantization strategy as discussed in this section to obtain the quantized coefficients $\{\bar{c}_{J,n}\}$

and $\{\bar{d}_{j,n}\}_{J-J_N+1\leq j\leq J}$.

Decoding:

1: **Parametric estimation**: the decoder approximates $f_p(t)$ by estimating the locations $\{t_i\}_{0\leq i\leq K+1}$ and the polynomial coefficients $\{a_{r,k}\}_{0\leq r\leq\lfloor\alpha\rfloor,0\leq k\leq K}$ of $f_p(t)$ from the received quantized coefficients $\{\bar{c}_{J,n}\}$ and $\{\bar{d}_{j,n}\}_{J-J_N+1\leq j\leq J}$ to obtain $\hat{f}_p(t)$;

2: **Cone of influence prediction**: the coefficients $\{d_{j,n}\}_{-\infty<j\leq J,n\in\mathcal{I}_p}$ in the cone of influence are predicted as follows:

$$\hat{d}_{j,n}=\langle\hat{f}_p(t),\psi_{j,n}(t)\rangle,\quad j=-\infty,...,J-J_N,\quad (j,n)\in\mathcal{I}_p;$$

3: **Error correction decoding**: the decoder uses the received $R_{j,\alpha}$ LSBs of quantized coefficients $\bar{d}_{j,n}$ together with the predicted coefficients $\hat{d}_{j,n}$ to decode $\{d_{j,n}\}_{J-J_N+1\leq j\leq J,(j,n)\in\mathcal{I}_p}$, where we denote with $\tilde{d}_{j,n}$, the decoded version of $d_{j,n}$;

4: **Final reconstruction**: $f(t)$ is reconstructed from the inverse wavelet transform of the following set of coefficients: $\{\bar{c}_{J,n}\}$, $\{\bar{d}_{j,n}\}_{J-J_N+1\leq j\leq J,(j,n)\notin\mathcal{I}_p}$, $\{\tilde{d}_{j,n}\}_{J-J_N+1\leq j\leq J,(j,n)\in\mathcal{I}_p}$ and $\{\hat{d}_{j,n}\}_{-\infty<j\leq J-J_N}$.

By using parametric estimation, the decoder is able to *predict* the coefficients in the cone of influence from the reconstructed function $\hat{f}_p(t)$ (see Fig. 1 (b) and (e)). Moreover, the encoder in the proposed algorithm is low in complexity as it is based on linear wavelet approximation.

B. Cramér-Rao Bound of Parametric Estimation

One of the core elements of the proposed algorithm is the parametric estimation step. Let us now assess the efficiency of using scaling and wavelet coefficients in parametric estimation with the Cramér-Rao Bound (CRB). Note that this is also an important step in determining the R-D performance of the proposed algorithm. Given a function $f(\Theta,t)$ where $\Theta=(\theta_1,\theta_2,...,\theta_K)^{\mathrm{T}}$ is a vector of K deterministic parameters, the CRB gives us the lower bound on the variance of any unbiased estimator i.e. $CRB(\Theta)\leq\mathrm{E}\left[\left(\hat{\Theta}-\Theta\right)\left(\hat{\Theta}-\Theta\right)^{\mathrm{T}}\right]$, where $\hat{\Theta}$ is obtained from any unbiased estimation procedure. The CRB can be calculated from the inverse of the Fisher Information Matrix $I(\Theta)$ as $CRB(\Theta)=I^{-1}(\Theta)=\left(\mathrm{E}\left[\nabla l(\Theta)\nabla l(\Theta)^{\mathrm{T}}\right]\right)^{-1}$, where $l(\Theta)$ is the log-likelihood function and $\nabla=\left(\frac{\partial}{\partial\theta_1},\frac{\partial}{\partial\theta_2},...,\frac{\partial}{\partial\theta_K}\right)$.

Consider now the problem of estimating Θ from a set of noisy transform coefficients given by $\hat{y}_n=y_n+\epsilon_n=y_n(f(\Theta,t))+\epsilon_n$, $n\in\mathcal{I}_L$, where ϵ_n is i.i.d. additive Gaussian noise with zero mean and variance σ_ϵ^2, and $\mathcal{I}_L$ denotes the index set of the received coefficients. It can be shown that $CRB(\Theta)$ is as follows:

$$CRB(\Theta)=\sigma_\epsilon^2\left(\sum_{n\in\mathcal{I}_L}\nabla y_n\nabla y_n^{\mathrm{T}}\right)^{-1}. \tag{6}$$

The full derivation of this result can be found in [27], [28].

In order to gain some intuition, we consider the following simplified estimation problem. The parametric function of interest $s(t)$ is assumed to be piecewise constant with a single discontinuity:

$$s(t) = \begin{cases} 0 & t < t_0, \\ A & t_0 \leq t < 1. \end{cases} \tag{7}$$

The estimator then has to retrieve $\Theta = (t_0, A)^{\mathrm{T}}$ from a set of L noisy scaling coefficients $\hat{y}_n = y_n + \epsilon_n = \langle s(t), \varphi_{J,n}(t) \rangle + \epsilon_n$, $n = 0, 1, ..., L-1$. By applying the formula in (6), we have that $CRB(\Theta) = \sigma_\epsilon^2 J_{t_0,A}$, where $J_{t_0,A}$ is

$$J_{t_0,A} = \begin{bmatrix} \sum_{n=0}^{L-1} \left(\frac{\partial y_n}{\partial t_0} \right)^2 & \sum_{n=0}^{L-1} \frac{\partial y_n}{\partial t_0} \frac{\partial y_n}{\partial A} \\ \sum_{n=0}^{L-1} \frac{\partial y_n}{\partial A} \frac{\partial y_n}{\partial t_0} & \sum_{n=0}^{L-1} \left(\frac{\partial y_n}{\partial A} \right)^2 \end{bmatrix}^{-1}.$$

Therefore, the CRBs for the estimation of the location t_0 and the amplitude A of $s(t)$ are given by

$$CRB(t_0) = \sigma_\epsilon^2 \left(J_{t_0,A} \right)_{11} \quad \text{and} \quad CRB(A) = \sigma_\epsilon^2 \left(J_{t_0,A} \right)_{22} \tag{8}$$

where $(J_{t_0,A})_{ij}$ denotes the entry in the i-th row and j-th column of the matrix $J_{t_0,A}$.

C. Rate-Distortion Analysis

The distortion-rate bound $D(R)$ of the proposed semi-parametric compression algorithm is derived in this section. For simplicity, we will use the simplified model of $f(t)$ given by

$$f(t) = s(t) + f_\alpha(t), \tag{9}$$

where $s(t)$ is the step function in (7) and $f_\alpha(t)$ is such that $0 \leq \alpha < 1$. In addition, let us assume that the decoder uses L low-pass coefficients to estimate the t_0 and A. With this set up, we will show that, at high rates, the distortion $D(R)$ decays as $R^{-2\alpha}$ for a wide range of rates.

Firstly, let us revisit the low-pass representation of $f(t)$, which can be written as follows:

$$y_n = \langle f(t), \varphi_{J,n}(t) \rangle \overset{(a)}{=} \langle s(t), \varphi_{J,n}(t) \rangle + \langle f_\alpha(t), \varphi_{J,n}(t) \rangle = y_n^s + y_n^\alpha,$$

where (a) follows from (2) and the linearity of the inner product. We can then write the quantized coefficients as

$$\bar{y}_n = y_n + \epsilon_n^q = y_n^s + y_n^\alpha + \epsilon_n^q = y_n^s + \epsilon_n^s, \tag{10}$$

where ϵ_n^q represents the quantization noise, which is assumed to be additive and Gaussian. Thus, we have written the quantized coefficients $\bar{y}_n$ as the sum of the coefficients of the step function y_n^s and the *noise* term $\epsilon_n^s = \epsilon_n^q + y_n^\alpha$. Suppose that a uniform scalar quantizer is used, at high rates, it follows that the variance σ_q^2 of the quantization noise $\{\epsilon_n^q\}_{0\leq n\leq L-1}$ is given by

$$\sigma_q^2 = C2^{-2\frac{R_2}{L}}, \tag{11}$$

where C is a constant and R_2 is the total rate allocated to represent $\{y_n\}_{0\leq n\leq L-1}$.

In the derivation of the $D(R)$ bound that follows, we assume the following:

- the wavelet basis has at least $\lfloor \alpha+1 \rfloor$ vanishing moments;
- the decoder uses L quantized low-pass coefficients to estimate the parameters t_0 and A;
- the estimators of t_0 and A are minimum-variance unbiased estimator;
- the quantization noise, ϵ_n^q, is additive Gaussian;
- the probability density function (PDF) of the low-pass coefficients of the smooth function $f_\alpha(t)$ denoted by y_n^α is zero-mean Gaussian [1] with variance σ_α^2;
- the quantization noise ϵ_n^q and the coefficients y_n^α are independent, which imply that ϵ_n^s is Gaussian distributed with zero mean and variance σ_ϵ^2, where

$$\sigma_\epsilon^2 = \left(\sigma_q^2 + \sigma_\alpha^2\right), \tag{12}$$

We begin by noting that the proposed decoding algorithm essentially reconstructs $s(t)$ and $f_\alpha(t)$ separately. The total distortion D can, therefore, be written as

$$D = D_\alpha + D_s, \tag{13}$$

where D_α and D_s are the distortion for the reconstruction of $f_\alpha(t)$ and $s(t)$ respectively. Since $f_\alpha(t)$ is uniformly α-Lipschitz smooth, an encoder in Algorithm 1 whose rate allocation follows the N-term wavelet linear approximation strategy achieves

$$D_\alpha(R_1) \leq c_7 R_1^{-2\alpha}, \tag{14}$$

where R_1 is the total rate (in bits) allocated for the compression of $f_\alpha(t)$.

Our next step then is to derive the expression for D_s. The reconstructed step function $\hat{s}(t)$ can be written as

$$\hat{s}(t) = \begin{cases} 0 & t < t_0 + \epsilon_t, \\ A + \epsilon_A & t_0 + \epsilon_t \leq t \leq 1. \end{cases}$$

[1]The PDF of y_n^α and ϵ_n^q are arbitrarily assumed to be zero-mean Gaussian as this allows us to use the analytical expression of the CRB, given by (6). The derived distortion-rate bound is then verified with simulations.

Here, the errors in the estimation of t_0 and A are represented by ϵ_t and ϵ_A. It then follows that the average distortion $D_s = \mathrm{MSE}\left(s(t) - \hat{s}(t)\right)$ is given by

$$D_s = \mathrm{E}\left[\int \left(s(t) - \hat{s}(t)\right)^2 dt\right] = \mathrm{E}\left[\int_{t_0}^{t_0+|\epsilon_t|} A^2 dt + \int_{t_0+\epsilon_t}^{1} \epsilon_A^2 dt\right] = \mathrm{E}\left[A^2|\epsilon_t| + \epsilon_A^2 c_\tau - \epsilon_A^2 \epsilon_t\right]$$

with a constant $0 \leq c_\tau \leq 1$. By assuming that ϵ_t and ϵ_A are independent, we have that

$$D_s = \mathrm{E}\left[A^2|\epsilon_t|\right] + \mathrm{E}\left[\epsilon_A^2 c_\tau\right] - \mathrm{E}\left[\epsilon_A^2\right] \underbrace{\mathrm{E}\left[\epsilon_t\right]}_{=0} = A^2 \mathrm{E}\left[|\epsilon_t|\right] + c_\tau \mathrm{E}\left[\epsilon_A^2\right].$$

Let us denote with σ_t^2 and σ_A^2 the variances of ϵ_t and ϵ_A respectively. Our assumption that the estimators are minimum variance estimators means that both σ_t^2 and σ_A^2 are given by their respective CRBs shown in (8) where, from (10), σ_ϵ^2 is given by (12). Using Jensen's inequality for concave functions[2], we have that

$$\mathrm{E}[|\epsilon_t|] = \mathrm{E}\left[\sqrt{\left(\epsilon_t - \mathrm{E}[\epsilon_t]\right)^2}\right] \leq \sqrt{\mathrm{E}\left[\left(\epsilon_t - \mathrm{E}[\epsilon_t]\right)^2\right]} = \sigma_t = \sqrt{CRB(t_0)}$$

as $\mathrm{E}[\epsilon_t] = 0$. Clearly, $\mathrm{E}\left[\epsilon_A^2\right] = \sigma_A^2 = CRB(A)$. Therefore, the expected distortion can be written as

$$D_s \leq A^2 \sigma_t + \sigma_A^2 = A^2 \sqrt{CRB(t_0)} + CRB(A).$$

By using the expression for the CRBs in (8) together with the relationship given in (10), we obtain the following R-D bound for the estimation of the step function:

$$\begin{aligned} D_s(R_2) &\leq A^2 \sigma_\epsilon \left(J_{t_0,A}\right)_{11}^{\frac{1}{2}} + \sigma_\epsilon^2 \left(J_{t_0,A}\right)_{22} \\ &\overset{(a)}{=} c_8 \left(\sigma_q^2 + \sigma_\alpha^2\right)^{\frac{1}{2}} + c_9 \left(\sigma_q^2 + \sigma_\alpha^2\right) \\ &\overset{(b)}{=} c_8 \left(c_{10} 2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right)^{\frac{1}{2}} + c_9 \left(c_{10} 2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right), \end{aligned} \tag{15}$$

where we have replaced $(J_{t_0,A})_{11}$ and $(J_{t_0,A})_{22}$ with c_8 and c_9 respectively and (a) and (b) follow from substituting in (12) and (11).

The expression for the total distortion-rate bound can now be obtained by substituting (14) and (15) into (13), which gives

$$D(R) \leq c_7 R_1^{-2\alpha} + c_8 \left(c_{10} 2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right)^{\frac{1}{2}} + c_9 \left(c_{10} 2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right) \tag{16}$$

with R equal to $R = R_1 + R_2$. Given the total rate R, we now need to allocate the bits among R_1 and R_2 so that the distortion in (16) is minimized. This is a well known constrained optimization problem,

[2]Jensen's inequality: given a random variable X and a concave function $f(x)$, it follows that $\mathrm{E}[f(X)] \leq f(\mathrm{E}[X])$.

which can be solved using Lagrange multipliers. One necessary condition for the optimal bit allocation is that the derivatives of the distortion D with respect to R_1 and R_2 are equal i.e.

$$\frac{\partial D}{\partial R_1} = \frac{\partial D}{\partial R_2}. \tag{17}$$

First, let us consider the case where the variance of $\{y_n^\alpha\}$ is negligible i.e. $\sigma_\alpha^2 \approx 0$. The $D(R)$ function now becomes

$$D(R) \leq c_7 R_1^{-2\alpha} + c_8\sqrt{c_{10}} 2^{\frac{-R_2}{L}} + c_9 c_{10} 2^{\frac{-2R_2}{L}}. \tag{18}$$

By applying the condition given in (17) to (18) and assuming a high-rate regime, we have that the bits can be approximately allocated as

$$R_2 \approx L(2\alpha+1)\log_2 R_1 + C' \tag{19}$$

and the total rate is then given by

$$R = R_1 + L(2\alpha+1)\log_2 R_1 + C' \approx R_1. \tag{20}$$

From the substitution of (19) into (18), together with the approximation in (20), we have that the total distortion is

$$D(R) \leq c_7 R^{-2\alpha} + c_{11} R^{-(2\alpha+1)} + c_{12} R^{-2(2\alpha+1)}. \tag{21}$$

From (21), we can see that both terms $c_{11}R^{-(2\alpha+1)}$ and $c_{12}R^{-2(2\alpha+1)}$ represent the distortion due to the discontinuity, which decay faster than $c_7 R^{-2\alpha}$. Therefore, given that $\sigma_\alpha^2 \approx 0$, the $D(R)$ curve of our proposed scheme follows $D(R) \sim R^{-2\alpha}$ at high R.

If we now consider the case where $\sigma_\alpha^2 > 0$ and assume that $c_{10}2^{\frac{-2R_2}{L}} < \sigma_\alpha^2$, the distortion given in (16) can then be approximated with a Taylor series expansion of the square root term $\left(c_{10}2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right)^{\frac{1}{2}}$ to obtain

$$D(R) \leq c_7 R_1^{-2\alpha} + c_8\left(\frac{c_{10}2^{\frac{-2R_2}{L}}}{2\sigma_\alpha} + \sigma_\alpha\right) + c_9\left(c_{10}2^{\frac{-2R_2}{L}} + \sigma_\alpha^2\right). \tag{22}$$

By solving the equal gradient condition in (17), where D is approximately given by (22), we obtain the following rate allocation: $R_2 = \frac{L}{2}(2\alpha+1)\log_2 R_1 + C'$ with a constant C'. Therefore, the overall $D(R)$ function becomes

$$D(R) \leq c_7 R^{-2\alpha} + c_{11} R^{-(2\alpha+1)} + \left(c_8\sigma_\alpha + c_9\sigma_\alpha^2\right).$$

Note that the term $c_{11}R^{-(2\alpha+1)}$ now represents the distortion caused by the discontinuity, which still decays faster than the distortion from the encoding of the smooth function. It is straightforward to extend the results of this R-D analysis to the case where the signal is piecewise smooth as given in (2), which is summarized as follows:

Summary 1: Consider a piecewise smooth function $f(t)$ as given in (2). The semi-parametric compression of $f(t)$, which employs a linear approximation strategy at the encoder and reconstructs the piecewise polynomial component $f_p(t)$ and the uniformly smooth component $f_\alpha(t)$ separately, achieves the following $D(R)$ function:

$$D(R) \leq c_{13}R^{-2\alpha} + c_{14}R^{-(2\alpha+1)} + c_{15}R^{-2(2\alpha+1)}$$

when the variance of the coefficients of $f_\alpha(t)$ is close to zero. Otherwise, the achievable $D(R)$ function is

$$D(R) \leq c_{13}R^{-2\alpha} + c_{14}R^{-(2\alpha+1)} + \left(c_{16}\sigma_\alpha + c_{17}\sigma_\alpha^2\right).$$

Therefore, given that σ_α^2 is sufficiently small, the proposed scheme can achieve the dominating decay rate of $R^{-2\alpha}$ for a wide range of rates. Such performance is comparable to that of a compression scheme based on nonlinear approximation.

IV. Distributed Semi-Parametric Compression

We now investigate the problem of distributed transform coding of piecewise smooth functions. Our aim here is to devise a suitable compression strategy for this new scenario. We then determine the differences in R-D performance among the distributed, joint and independent compression schemes.

A. Signal and Disparity Models

We consider the scenario where N piecewise smooth signals are independently encoded but are decoded jointly. We denote a set of N functions with $\{f_i(t)\}_{1\leq i\leq N}$, where the subscript i indicates the signal that is observed by Encoder i. Each function is given by the model in (2). Hence, $f_i(t) \in \mathbf{L_2}([0,1])$ can be written as

$$f_i(t) = f_{i_p}(t) + f_{i_\alpha}(t), \quad i = 1, ..., N. \tag{23}$$

We consider a disparity model where the main difference between the two observed signals is described by a shift. For simplicity, we write each function $f_i(t)$ as

$$f_i(t) = f_1(t - \tau_i) + \epsilon_{i_\alpha}(t), \quad i = 2, ..., N. \tag{24}$$

This is illustrated in Fig. 2. The term $\epsilon_{i_\alpha}(t)$ represents the prediction error (or the residual), which is assumed to be uniformly α-Lipschitz. The construction of this model is motivated by the block-based prediction or disparity estimation in applications such as video compression algorithms, multi-view images and image-based rendering. Fig. 3 shows an example of the scan lines taken from two images where the signals and the residual follow the model in (23) and (24) quite closely.

B. Distributed Semi-Parametric Compression Strategy

In the analysis that follows, we will assume that a reconstructed version of $f_1(t)$ is available at the decoder by means of conventional wavelet nonlinear approximation-based compression strategy. Thus, $f_i(t)$, $i = 2, ..., N$, can be reconstructed by first estimating the shift parameter τ_i and then the residual $\epsilon_{i_\alpha}(t)$ as follows: $\hat{f}_i(t) = \hat{f}_1(t - \hat{\tau}_i) + \hat{\epsilon}_{i_\alpha}(t)$, where $\hat{f}_i(t)$ and $\hat{\epsilon}_{i_\alpha}(t)$ denote the reconstructed versions, and $\hat{\tau}_i$ is the estimated shift parameter. One of the challenges here is that Encoders 2 to N have no access to $f_1(t)$ and, hence, the prediction error $\epsilon_{i_\alpha}(t)$ cannot be directly calculated and transmitted.

From the model of the piecewise smooth function in (23), let us define the locations of the discontinuities in $f_{i_p}(t)$ with $\{t_{i_k}\}_{1\leq k\leq K}$ such that $t_{i_k} = t_{1_k} - \tau_i$. Therefore, by retrieving $\{t_{i_k}\}_{1\leq k\leq K}$, the shift parameters τ_i can be estimated by taking the average:

$$\hat{\tau}_i = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{t}_{1_k} - \hat{t}_{i_k}\right), \quad i = 2, ..., N. \tag{25}$$

In the following analysis, we will assume for simplicity that the decoder is able to retrieve the locations $\{t_{1_k}\}_{1\leq k\leq K}$ of $f_{1_p}(t)$. Hence, the problem of estimating τ_i becomes the problem of estimating the locations $\{t_{i_k}\}_{1\leq k\leq K}$, $i = 2, ..., N$. In addition, we will assume that the decoder only uses L low-pass coefficients $c_{i_{J,n}} = \langle f_i(t), \varphi_{J,n}(t)\rangle$, $n = 0, ..., L-1$, of $f_i(t)$ to estimate $\{t_{i_k}\}_{1\leq k\leq K}$.

Since the reconstructed version of $f_1(t)$ is available at the decoder, the prediction of $f_i(t)$ can be formed by $\tilde{f}_i(t) = \hat{f}_1(t - \hat{\tau}_i)$. Assuming that the range of the amplitude of $\epsilon_{i_\alpha}(t)$ can be estimated a priori, we can adopt a similar quantization strategy as the one discussed in Section III, where the encoder

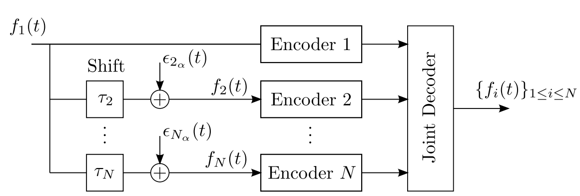
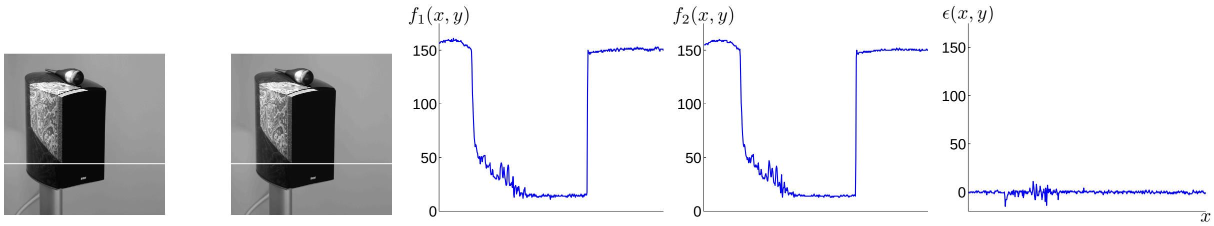


Fig. 2. Distributed compression problem setup using the disparity-by-translation model with prediction error. Each function is piecewise smooth and $f_i(t) = f_1(t - \tau_i) + \epsilon_{i_\alpha}(t)$, $i = 2, ..., N$.



(a) (b) (c) (d) (e)

Fig. 3. Examples of the scan lines taken from stereo images shown in (a) and (b). (c) A scan line of the first image $f_1(x, y')$; (d) a scan line from the second image $f_2(x, y')$; (e) the prediction error given by $f_2(x, y') - f_1(x - \tau_x, y')$, where τ_x denotes the shift parameter.

only transmits the required LSBs to the decoder. We, therefore, propose the following semi-parametric distributed compression algorithm:

Algorithm 2 (distributed semi-parametric compression):

Encoding and Decoding of $f_1(t)$:

1: **Nonlinear approximation-based compression**: $f_1(t)$ is encoded and decoded with a conventional wavelet nonlinear approximation-based compression scheme;

2: **Extracting locations of discontinuities**: the locations $\{t_{1_k}\}_{k=0,..,K}$ are extracted from $\hat{f}_1(t)$.

Encoding of $f_i(t)$, $i = 2, ..., N$:

1: **N-term linear approximation**: the encoder approximates $f_i(t)$ as shown in (5);

2: **Quantization**: the coefficients $\{c_{i_{J,n}}\}$ and $\{d_{i_{j,n}}\}_{J-J_N+1\leq j\leq J}$ are quantized using a linear approximation-based quantization strategy as discussed in Section III-A to obtain $\{\bar{c}_{i_{J,n}}\}$ and $\{\bar{d}_{i_{j,n}}\}_{J-J_N+1\leq j\leq J}$. For $\{\bar{d}_{i_{j,n}}\}$, only the required LSBs are transmitted. The analysis that determines the number of required LSBs will be given in the next section.

Joint Decoding of $f_i(t)$, $i = 2, ..., N$:

1: **Parametric estimation**: the decoder estimates the locations t_{i_k}, $k = 0, .., K$, from the L quantized low-pass coefficients $\bar{c}_{i_{J,n}}$ and the shift parameter τ_i is calculated using (25);

2: **Prediction by translation**: a predicted version of $f_i(t)$ is formed by $\tilde{f}_i(t) = \hat{f}_1(t - \hat{\tau})$ and the coefficients $\tilde{d}_{i_{j,n}} = \left\langle \tilde{f}_i(t), \psi_{j,n}(t) \right\rangle$, $J - J_N + 1 \leq j \leq J$ are obtained;

3: **Error correction decoding**: the error correction decoding technique discussed in Section III-A can also be applied at this stage where the received LSBs of quantized coefficients $\{\bar{d}_{i_{j,n}}\}_{J-J_N+1\leq j\leq J}$ are used together with the predicted coefficients $\{\tilde{d}_{i_{j,n}}\}_{J-J_N+1\leq j\leq J}$ to decode $d_{i_{j,n}}$. The decoded coefficient is denoted with $\hat{d}_{i_{j,n}}$;

4: **Final reconstruction**: the signal $f_i(t)$ is reconstructed by taking the inverse wavelet transform of the following set of coefficients: $\{\bar{c}_{i_{J,n}}\}$, $\{\hat{d}_{i_{j,n}}\}_{J-J_N+1\leq j\leq J}$ and $\{\tilde{d}_{i_{j,n}}\}_{-\infty<j\leq J-J_N}$.

C. Rate-Distortion Analysis

As with the centralized case, we consider a simplified model of the piecewise smooth signal, which consists of a step function and a uniformly smooth α-Lipschitz function. We also make the same set of assumptions as in the previous section here. For the sake of clarity, we begin by assuming that there are only two functions $f_1(t)$ and $f_2(t)$ where

$$f_i(t) = s_i(t) + f_{i_\alpha}(t), \;\; s_i(t) = \begin{cases} 0 & t < t_i, \\ A & t_i \leq t < 1, \end{cases}$$

for $i = 1, 2$. The generalization of the analysis to the case of N signals will be given at the end of this section. We divide the derivation of the $D(R)$ function into two parts; the first part will look at disparity by translation only and the second part will look at the residual. We will show that the overall $D(R)$ function decays as $R^{-2\alpha}$ for a wide range of rates and that the rate-distortion performance of the proposed distributed compression scheme is better than an independent encoding scheme and, in some cases, comparable to the joint encoding scheme.

1) Disparity by Translation: Let us begin by assuming that there is no prediction error i.e. $\epsilon_{2_\alpha}(t) = 0$. Hence, the function $f_2(t)$ can be written as $f_2(t) = f_1(t-\tau)$. We have that the total distortion is

$$D = D_1 + D_2, \tag{26}$$

where D_1 and D_2 are the distortion due to the reconstruction of $f_1(t)$ and $f_2(t)$ respectively. Since a conventional wavelet nonlinear approximation-based compression scheme is used to encode $f_1(t)$, D_1 is given by

$$D_1(R_1) \leq c_1 R_1^{-2\alpha}, \tag{27}$$

where R_1 is the total number of bits allocated to compress $f_1(t)$. Our next task is then to derive D_2.

Since the decoder reconstructs $f_2(t)$ by estimating the shift parameter $\tau = t_1 - t_2$, the reconstructed function can be written as $\hat{f}_2(t) = \hat{f}_1(t-\hat{\tau}) = \hat{s}_1(t-\hat{\tau}) + \hat{f}_{1_\alpha}(t-\hat{\tau})$, where $\hat{\tau}$ is the estimated shift parameter. Let $\epsilon_{f_1}(t)$ denote the compression error of $f_1(t)$ such that $f_1(t) = \hat{f}_1(t) + \epsilon_{f_1}(t)$. It then follows that D_2 is given by $D_2 = \mathrm{E}\left[\int \left(f_2(t) - \hat{f}_2(t)\right)^2 dt\right] = \mathrm{E}\left[\int \left(f_1(t-\tau) - (f_1(t-\hat{\tau}) + \epsilon_{f_1}(t-\hat{\tau}))\right)^2 dt\right]$.

Let ϵ_τ denote an error in the estimation of τ i.e. $\epsilon_\tau = \tau - \hat{\tau}$. Given that R_1 is sufficiently high and that ϵ_τ is small, the distortion D_2 can be approximated as

$$D_2 \approx \mathrm{E}\left[\int (s_1(t-\tau) - s_1(t-\hat{\tau}))^2\, dt + \int \epsilon_{f_1}^2(t-\hat{\tau})dt\right] = A^2 \mathrm{E}\left[|\epsilon_\tau|\right] + D_1.$$

Given that the location t_1 can be retrieved by the decoder, the error in the parametric estimation becomes $\epsilon_\tau = t_2 - \hat{t}_2 = \epsilon_{t_2}$. From the analysis of the centralized case in Section III, this leads to the distortion of the form:

$$D_2 \approx A^2 \mathrm{E}\left[|\epsilon_{t_2}|\right] + D_1 = A^2\sqrt{CRB(t_2)} + D_1.$$

Using the same set of assumptions as in Section III, it follows that $\sqrt{CRB(t_2)} = c_2\left(\sigma_q^2 + \sigma_\alpha^2\right)^{\frac{1}{2}}$ with $\sigma_q^2 = C2^{-\frac{2R_2}{L}}$. Here, σ_q^2 is the variance of the quantization noise and σ_α^2 represents the variance of the term $\langle f_{2_\alpha}(t), \varphi_{J,n}(t)\rangle$. Hence, we have that

$$D_2 \approx c_2\left(c_3 2^{-\frac{2R_2}{L}} + \sigma_\alpha^2\right)^{\frac{1}{2}} + D_1. \tag{28}$$

Therefore, the total distortion can be approximated by substituting (27) and (28) into (26), which gives

$$D \approx 2c_1 R_1^{-2\alpha} + c_2\left(c_3 2^{-\frac{2R_2}{L}} + \sigma_\alpha^2\right)^{\frac{1}{2}}. \tag{29}$$

Assuming a high rate regime, where $c_3 2^{\frac{-2R_2}{L}} < \sigma_\alpha^2$, the distortion given in (29) can be approximated with a Taylor series expansion of the square root function as follows:

$$D(R) \approx 2c_1 R_1^{-2\alpha} + \frac{c_2 c_3}{2\sigma_\alpha} 2^{\frac{-2R_2}{L}} + c_2\sigma_\alpha. \tag{30}$$

By solving the Lagrange multiplier method, we have that the optimal rate allocation for R_2 is

$$R_2 = \frac{L}{2}(2\alpha+1)\log_2 R_1 + C \tag{31}$$

with a constant C. The total rate R is thus given by

$$R = R_1 + \frac{L}{2}(2\alpha+1)\log_2 R_1 + C \approx R_1. \tag{32}$$

Therefore, by substituting (31) into (30) together with the approximation in (32), the overall $D(R)$ function of the proposed scheme is as follows:

$$D(R) \leq 2c_1 R^{-2\alpha} + c_4 R^{-(2\alpha+1)} + c_2\sigma_\alpha.$$

If the term σ_α is sufficiently small, then $D(R)$ decays as $R^{-2\alpha}$ for a wide range of rates. Finally, we note here that even though Encoder 2 employs a linear compression strategy, the overall $D(R)$ function has a decay characteristic of a nonlinear scheme for a wide range of rates.

2) Disparity by Translation with Prediction Error: Let us now include the prediction error $\epsilon_\alpha(t)$ to the distortion-rate analysis such that $f_2(t) = f_1(t-\tau) + \epsilon_\alpha(t) = s_1(t-\tau) + f_{1_\alpha}(t-\tau) + \epsilon_\alpha(t)$. It follows that the total distortion of the proposed scheme is now given by $D = D_1 + D_2 = D_1 + D_\tau + D_\epsilon$, where D_τ is due to the reconstruction of the prediction $\tilde{f}_2(t) = \hat{f}_1(t-\hat{\tau})$ and D_ϵ is due to the reconstruction of the prediction error. Our next step is to determine D_ϵ.

First, we examine the approximation property of the prediction error. Since the function $\epsilon_\alpha(t)$ is uniformly α-Lipschitz, the wavelet linear approximation based compression gives the following distortion:

$$D_\epsilon(R_\epsilon) = \beta D_1(R_\epsilon) = \beta c_1 R_\epsilon^{-2\alpha} \quad \text{with} \quad \beta \geq 0, \tag{33}$$

where R_ϵ is the total rates allocated to represent $\epsilon_\alpha(t)$. The constant β is used to relate the energy of the prediction error to $f_1(t)$. In terms of the wavelet coefficients, (33) implies $\max_{j,n\in\mathbb{Z}} |d_{\epsilon_{j,n}}| \leq \sqrt{\beta}\max_{j,n\in\mathbb{Z}} |d_{1_{j,n}}|$, where $\{d_{\epsilon_{j,n}}\}$ and $\{d_{1_{j,n}}\}$ denote the wavelet coefficients of $\epsilon_\alpha(t)$ and $f_1(t)$ respectively.

If we are to compress $\epsilon_\alpha(t)$ directly, a linear approximation-based compression would achieve the $D(R)$ given in (33) with a quantizer step size

$$\Delta_\epsilon \leq \sqrt{\beta} \max_{n\in\mathbb{Z}} \left(|d_{1_{J,n}}|\right) 2^{-J_N(\alpha+1/2)} \tag{34}$$

such that the coefficients in decomposition level $J - J_N + 1 \leq j \leq J$ are retained. Note that $\max\limits_{n\in\mathbb{Z}} \left(|d_{1_{J,n}}|\right) \approx \max\limits_{n\in\mathbb{Z}} \left(|d_{2_{J,n}}|\right)$. Let $R_\epsilon(j)$ be the number of bits per coefficient required to directly quantize $d_{\epsilon_{j,n}}$, it follows that $R_\epsilon(j) = \lceil (J_N - J + j)(\alpha + 1/2)\rceil + 1$ with $J_N > 0$, $J < 0$ and $j \leq J$. Note that one extra bit has been included for the sign.

In our setup, however, Encoder 2 does not have access to $\epsilon_\alpha(t)$. Intuitively, one can still quantize the coefficients of $f_2(t)$ with a step size Δ_ϵ as given in (34). If we assume that the best possible prediction $f_1(t-\tau)$ is available at the decoder, then the information of the wavelet coefficients of $\epsilon_\alpha(t)$ is fully contained within the first $R_\epsilon(j)$ LSBs. This means that the quantized coefficients of the prediction error can be retrieved from the first $R_\epsilon(j)$ LSBs of each coefficient and any additional distortion is due to the error in the prediction.

In reality, β has to be estimated prior to compression as $\epsilon_\alpha(t)$ is not accessible. We denote with β^*, the value of β estimated by Encoder 2. Thus, the actual quantizer step size used by Encoder 2 is given by $\Delta_\epsilon^* = \sqrt{\beta^*}\max_{n\in\mathbb{Z}} \left(|d_{2_{J,n}}|\right) 2^{-J_N(\alpha+1/2)}$. If $f_1(t-\tau)$ is available at the decoder, then setting $\beta^* = \beta$ gives us the $D(R)$ equivalent to that of a joint encoding scheme. Instead, if $\beta^* > \beta$, then the step size Δ_ϵ^* will be too large and the added redundancy will result in an inferior compression performance. Lastly, if $\beta^* < \beta$ then not enough bits will be transmitted for $\epsilon_\alpha(t)$ to be correctly decoded. In the analysis that follows, we will assume that $\beta^* \geq \beta$.

With the above quantization strategy, we can be certain that D_ϵ decays as $D_\epsilon \leq \beta^* c_1 R_{2_\epsilon}^{-2\alpha}$ with a linear approximation-based compression strategy both inside and outside the cone of influence of discontinuities. This is because the encoder transmits enough bits to carry the information of $\epsilon_\alpha(t)$ and any additional error is due to the parametric estimation of the prediction $\tilde{f}_2(t)$. Here, R_{2_ϵ} denotes the total number of bits allocated to represent $\epsilon_\alpha(t)$. Note also that we have used the estimated β^* instead of β. From the analysis in the previous setup, where the prediction error was absent, it directly follows from (28) and (30) that D_τ is given by

$$D_\tau(R_{2_\tau}) \approx c_1 R_1^{-2\alpha} + \frac{c_2 c_3}{2\sigma_\alpha} 2^{\frac{-2R_{2_\tau}}{L}} + c_2 \sigma_\alpha,$$

where R_{2_τ} is the number of bits allocated to the L low-pass coefficients, which are used to estimate $\tilde{f}_2(t)$.

We can now express the total distortion as follows:

$$D(R) \leq 2c_1 R_1^{-2\alpha} + \frac{c_2 c_3}{2\sigma_\alpha} 2^{\frac{-2R_{2_\tau}}{L}} + c_2 \sigma_\alpha + \beta^* c_1 R_{2_\epsilon}^{-2\alpha} \tag{35}$$

with $R = R_1 + R_{2_\tau} + R_{2_\epsilon}$. The optimal bit allocation can then be obtained using the Lagrange multiplier method, which gives

$$\begin{aligned} R_{2_\tau} &= \tfrac{L}{2}(2\alpha+1)\log_2 R_1 + C, \\ R_{2_\epsilon} &= \left(\tfrac{\beta^*}{2}\right)^{\frac{1}{(2\alpha+1)}} R_1. \end{aligned} \tag{36}$$

The total rate R can then be approximated as

$$\begin{aligned} R &= R_1 + \tfrac{L}{2}(2\alpha+1)\log_2 R_1 + C + \left(\tfrac{\beta^*}{2}\right)^{\frac{1}{(2\alpha+1)}} R_1 \\ &\approx \left(1 + \left(\tfrac{\beta^*}{2}\right)^{\frac{1}{(2\alpha+1)}}\right) R_1 \end{aligned} \tag{37}$$

for high R. By substituting (36) and (37) into (35), we have that the $D(R)$ curve for the proposed distributed semi-parametric compression scheme is given by

$$D(R) \leq \left(1 + \left(\frac{\beta^*}{2}\right)^{\frac{1}{(2\alpha+1)}}\right)^{2\alpha+1} \left(2c_1 R^{-2\alpha} + c_4 R^{-(2\alpha+1)}\right) + c_2 \sigma_\alpha.$$

If we assume that R is high and that the term σ_α is negligible, then the distortion-rate behavior at high rates follows

$$D(R) \leq 2\left(1 + \left(\frac{\beta^*}{2}\right)^{\frac{1}{2\alpha+1}}\right)^{2\alpha+1} c_1 R^{-2\alpha}. \tag{38}$$

3) Extension to N Signals: It is straight forward to extend the R-D analysis for the setup in (24) to N signals. It is assumed here that $\sigma_{i_\alpha} = \sigma_\alpha$ for $i = 2, ..., N$. In addition, we can assume that Encoder 2 to N use the same β^*. From (35), it follows that the total distortion is now given by

$$D(R) \leq Nc_1 R_1^{-2\alpha} + \frac{c_2 c_3}{2\sigma_\alpha} \sum_{i=2}^{N} 2^{\frac{-2R_{i_\tau}}{L}} + (N-1)c_2\sigma_\alpha + \beta^* c_1 \sum_{i=2}^{N} R_{i_\epsilon}^{-2\alpha},$$

which gives the following optimal rate allocation:

$$R_{i_\tau} = \frac{L}{2}(2\alpha+1)\log_2 R_1 + C \text{ and } R_{i_\epsilon} = \left(\frac{\beta^*}{N}\right)^{\frac{1}{2\alpha+1}} R_1$$

with $i = 2, ..., N$. The total rate can then be approximated as $R \approx \left(1 + (N-1)\left(\frac{\beta^*}{N}\right)^{\frac{1}{2\alpha+1}}\right) R_1$. Finally, the resulting $D(R)$ bound obtained is given as follows:

$$D(R) \leq \left(1 + (N-1)\left(\frac{\beta^*}{N}\right)^{\frac{1}{(2\alpha+1)}}\right)^{2\alpha+1} \left(Nc_1 R^{-2\alpha} + (N-1)c_{10} R^{-(2\alpha+1)}\right) + (N-1)c_2\sigma_\alpha.$$

D. Comparison with Independent and Joint Compression

Let us now compare the R-D performance of the proposed compression scheme with that of an independent scheme and an ideal joint encoding scheme. For the sake of clarity in the following analysis, we first assume that there are only two signals, namely $f_1(t)$ and $f_2(t)$.

1) Independent compression: The independent compression scheme encodes and decodes each observed signal independently. We denote the $D(R)$ curve of such scheme with $D_{ind}(R)$, which can be written as $D_{ind}(R) \leq c_1 R_1^{-2\alpha} + c_1 R_2^{-2\alpha}$ with $R = R_1 + R_2$. Clearly, the optimal rate allocation is given by $R_1 = R_2$. This gives the following distortion-rate function:

$$D_{ind}(R) \leq 2c_1 \left(\frac{R}{2}\right)^{-2\alpha} = 2^{2\alpha+1} c_1 R^{-2\alpha}.$$

In comparison to the proposed distributed scheme, assuming that σ_α is sufficiently small, the distortion of the independent scheme is higher by a factor of

$$\frac{D_{ind}(R)}{D(R)} = \frac{2^{2\alpha}}{\left(1 + \left(\frac{\beta^*}{2}\right)^{\frac{1}{2\alpha+1}}\right)^{2\alpha+1}}.$$

Interestingly, the gain in the compression performance increases with the smoothness of the function. One can easily show that the independent compression scheme for N signals achieves $D_{ind}(R) = 2c_1 N^{2\alpha} R^{-2\alpha}$.

2) Centralized compression: Since an ideal centralized joint encoding scheme has access to all of the observed functions, the prediction error can be encoded with optimal bit allocation and transmitted directly to the decoder along with the quantized shift parameter τ. The decoder can then reconstruct $f_2(t)$ as $\hat{f}_2(t) = \bar{f}_1(t - \bar{\tau}) + \bar{\epsilon}_\alpha(t)$.

Let ϵ_τ^q be the quantization error of τ, assuming that a uniform quantizer is used, the distortion in the reconstruction of the prediction $\tilde{f}_2(t) = f_1(t - \tau)$ can be approximated by $D_\tau(R_\tau) \approx A^2 E[|\epsilon_\tau^q|] \leq \frac{A^2}{\sqrt{12}} 2^{-R_\tau}$, where we have used the Jensen's inequality and R_τ denotes the number of bits allocated to quantize τ. By following a similar analysis as shown in the distributed case, the total distortion can be shown to be

$$D_{joint}(R) \leq 2c_1 R_1^{-2\alpha} + \frac{A^2}{\sqrt{12}} 2^{-R_\tau} + \beta c_1 R_\epsilon^{-2\alpha},$$

where R_ϵ is the total bits allocated to the compression of $\epsilon_\alpha(t)$. It then follows that the optimal bit allocation is given by $R_\tau = (2\alpha + 1)\log_2 R_1 + C$ and $R_\epsilon = \left(\frac{\beta}{2}\right)^{\frac{1}{(2\alpha+1)}} R_1$. In a high-rate regime, we can approximate the total rate R with

$$R = R_1 + R_\tau + R_\epsilon \approx \left(1 + \left(\frac{\beta}{2}\right)^{\frac{1}{(2\alpha+1)}}\right) R_1.$$

It then follows that the joint encoding scheme achieves the following $D(R)$ at high rates:

$$D(R) \leq \left(1 + \left(\frac{\beta^*}{2}\right)^{\frac{1}{2\alpha+1}}\right)^{2\alpha+1} \left(2c_1 R^{-2\alpha} + c_9 R^{-(2\alpha+1)}\right),$$

which has the same form as the distributed case given in (38). The only difference is in the values of β^* and β. This means that the closer the value of β^* can be to the actual β (i.e. the better the quality of the prediction of $f_2(t)$), the closer the performance of the distributed compression scheme is to that of an ideal joint encoding scenario. The extension of this analysis to the case with N signals gives

$$D_{joint}(R) \leq \left(1 + (N-1)\left(\frac{\beta}{N}\right)^{\frac{1}{(2\alpha+1)}}\right)^{2\alpha+1} \left(Nc_1 R^{-2\alpha} + (N-1)c_{11} R^{-(2\alpha+1)}\right).$$

As with before, the analysis in this section can be generalized to the case where each function is piecewise smooth as given in (2). We summarize the findings in this section as follows:

Summary 2: Consider a set of N piecewise smooth functions, $\{f_i(t)\}_{1\leq i\leq N} \in \mathbf{L_2}([0,1])$, where each consists of a piecewise polynomial function and a uniformly α-Lipschitz function and $f_i(t) = f_1(t - \tau) + \epsilon_{i_\alpha}(t)$. The function $\epsilon_{i_\alpha}(t)$ is uniformly α-Lipschitz. Given that the $D(R)$ function corresponding to a linear compression of $\epsilon_{i_\alpha}(t)$ follows $D_{i_\epsilon}(R_{i_\epsilon}) = \beta D_1(R_{i_\epsilon})$, at high rates, the proposed semi-parametric distributed compression scheme achieves

$$D(R) \leq \left(1+(N-1)\left(\frac{\beta^*}{N}\right)^{\frac{1}{(2\alpha+1)}}\right)^{2\alpha+1}\left(Nc_{18}R^{-2\alpha}+(N-1)c_{19}R^{-(2\alpha+1)}\right)+(N-1)c_{20}\sigma_\alpha,$$

where β^* is the estimate of β. At high rates, provided that σ_α is sufficiently small, if $\beta^* = \beta$ then the achieved $D(R)$ performance is comparable to that of a joint encoding scheme and is better by a factor of

$$\frac{2N^{2\alpha-1}}{\left(1+(N-1)\left(\frac{\beta^*}{N}\right)^{\frac{1}{2\alpha+1}}\right)^{2\alpha+1}}$$

when compared to an independent compression scheme.

V. Constructive Parametric Estimation Algorithms

One of the key elements of the proposed algorithms is the use parametric estimation during the decoding stage. In this section, we introduce a practical parametric estimation technique based around the recently developed concept of sampling of signals with finite rate of innovation (FRI) [16]. FRI signals are, loosely speaking, a class of signals or functions $f(t)$ that can be described by a finite number of free parameters over a given interval $t \in [t_a, t_b]$. The definition and sampling theory of FRI signals are given in details in [16], [17]. It is easy to see that a piecewise polynomial signal as shown in (3) also belongs to this class of functions as there are a finite number of discontinuities and each polynomial piece can be described by at most $\lfloor \alpha \rfloor$ polynomial coefficients.

Let us present one of the key results from the sampling schemes of FRI signals described in [16], [17]. Given a function $f(t)$ and a scaling function $\varphi(t)$, in the context of sampling, the samples or the coefficients are given by $y_n = \langle f(t), \varphi(t/T - n)\rangle$, where T is the sampling period. Assume that $\varphi(t)$ together with its shifted versions can reproduce polynomials of maximum degree P i.e. $\varphi(t)$ satisfies

$$\sum_{n\in\mathbb{Z}} w_n^p \varphi(t/T - n) = t^p, \quad p = 0, 1, ..., P \tag{39}$$

for a proper set of coefficients w_n^p. The polynomial reproduction coefficients can be calculated as $w_n^p = \langle t^p, \tilde{\varphi}(t/T - n)\rangle$, where $\tilde{\varphi}(t)$ is the dual of $\varphi(t)$. It then follows that the continuous moment m_p of order p of the signal $f(t)$ can be obtained as follows:

$$M_p = \int f(t)t^p dt = \left\langle f(t), \sum_{n\in\mathbb{Z}} w_n^p \varphi(t/T - n)\right\rangle = \sum_{n\in\mathbb{Z}} w_n^p \langle f(t), \varphi(t/T - n)\rangle = \sum_{n\in\mathbb{Z}} w_n^p y_n. \tag{40}$$

Therefore, given w_n^p, one can retrieve the continuous moments of $f(t)$ from the coefficients y_n provided that $f(t)$ lies in the region where the condition given by (39) is satisfied. In fact, the standard result in wavelet theory states that if the wavelet function has $(P+1)$ vanishing moments then the corresponding scaling function reproduces polynomial up to an order of P.

In [17], a sampling scheme that allows a perfect reconstruction of a piecewise polynomial signal was presented. First, the authors developed a sampling framework for a stream of Diracs and differentiated Diracs based on the annihilating filter method (a.k.a. Prony's method [29]), which uses the retrieved continuous moments to calculate the locations and amplitudes of each Dirac. Since the R-th derivative $f_p^{(R)}(t)$ of a piecewise polynomial function $f_p(t)$ with pieces of maximum degree R is also a stream of differentiated Diracs, the same sampling scheme can be applied to reconstruct $f_p^{(R)}(t)$. Let us denote with $z_n^{(R)}$, the R-th order finite difference of y_n. It was shown in [17] that $z_n^{(1)} = y_{n+1} - y_n = \left\langle \frac{df(t)}{dt}, \varphi(t/T-n) * \beta_0(t/T-n) \right\rangle$, which leads to the following identity:

$$z_n^{(R)} = \left\langle f^{(R)}(t), \varphi(t/T-n) * \beta_{R-1}(t/T-n) \right\rangle,$$

where $\beta_p(t)$ denotes a p-th order B-spline function. Therefore, from (40), we have that

$$M_p^{(R)} = \sum_n w_n'^p z_n^{(R)}, \tag{41}$$

where $M_p^{(R)}$ is the p-th order continuous moment of $f_p^{(R)}(t)$ and $w_n'^p$ are the polynomial reproduction coefficients of the new scaling function $\varphi(t/T-n) * \beta_{R-1}(t/T-n)$. One can, therefore, obtain $M_p^{(R)}$ directly from the coefficients y_n using (41), which then allows the function $f_p^{(R)}(t)$ and, hence, $f_p(t)$ to be reconstructed.

Consider now a set of noisy quantized low-pass coefficients $\bar{c}_{M,n}$ of a piecewise polynomial function $f_p(t) = \sum_{k=0}^{1} \sum_{r=0}^{r=R-1} a_{k,r}(t-t_k)_+^r$, where $t_0 = 0$, $t_1 \in]0,1[$ and $\bar{c}_{J,n} = \langle f_p(t), \varphi_{J,n}(t) \rangle + \epsilon_n$. Assuming that $\varphi_{J,n}(t)$ reproduces polynomials of maximum order $P \geq R-1$, we now present the following parametric estimation algorithm, which is based on the sampling theory of FRI signals, for the estimation of $f_p(t)$:

Algorithm 3 (FRI-based parametric estimation algorithm):

1: **Finite difference**: the R-order finite difference of $\bar{c}_{M,n}$ is obtained by $\bar{z}_n^{(R)} = \bar{z}_{n+1}^{(R-1)} - \bar{z}_n^{(R-1)}$ with $\bar{z}_n^{(1)} = \bar{c}_{J,n+1} - \bar{c}_{J,n}$.

2: **Thresholding**: in order to reduce the effect of noise, thresholding is applied as

$$\tilde{z}_n^{(R)} = \begin{cases} \bar{z}_n^{(R)} & \bar{z}_n^{(R)} \geq z_{th} \\ 0 & \text{otherwise.} \end{cases}$$

where z_{th} is a constant, which can be chosen empirically.

3: **Moments estimation**: the continuous moments of $f_p^{(R)}(t)$ are estimated as:

$$\tilde{M}_p^{(R)} = \sum_n w_n'^{(p)} \tilde{z}_n^{(R)} \quad p = 0, 1...., P + R - 1.$$

4: **Annihilating filter method**: the locations t_k of the discontinuities are estimated from $\tilde{M}_p'$, $p = 0, ..., P + R - 1$, with the annihilating filter method described in [17].

5: **Solving Vandermonde system**: the amplitudes of the stream of differentiated Diracs $f_p^{(R)}(t)$ are estimated by solving the Vandermonde system of equations of $\tilde{M}_p^{(R)} = \sum_{k=0}^{K-1} \sum_{r=0}^{R_k-1} a_{k,r}(-1)^r \frac{p!}{(p-r)!} t_k^{p-r}$, $p = 0, ..., P + R - 1$, which is derived from the identity in (41) (see [17]).

6: **Integration**: the function $f_p(t)$ is estimated by integrating the reconstructed R-th derivative of $f_p(t)$.

Given that $\varphi_{J,n}(t)$ has a compact support L, it was shown in [17] that the above algorithm can reproduce polynomials with pieces of maximum degree $R-1$ if there are at most K discontinuities in an interval of size $2K(L+R)2^J$ and $P+R \geq 2KR-1$. Note however that since $\varphi(t)$ has compact support, if groups of K discontinuities are sufficiently distant, we can reconstruct them independently [17] and therefore we can still reconstruct piecewise polynomial signals with more than K discontinuities. Finally, it is worth noting the stability of the proposed algorithm deteriorates as the number of discontinuities K per time interval increases (as demonstrated in [17]). In particular, we find that the algorithm is stable for one discontinuity per time interval even for piecewise polynomials with high order R. The algorithm becomes noticeably less stable when $K \geq 2$.

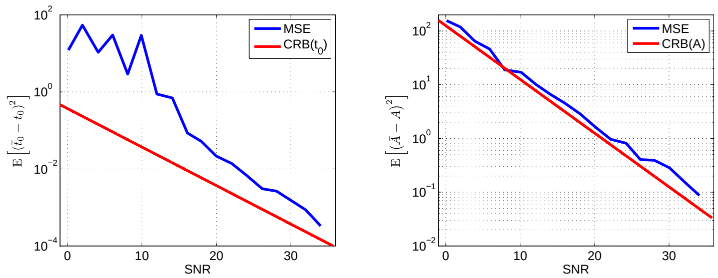
A practical centralized and distributed semi-parametric compression schemes with a low-complexity encoder can now be constructed from Algorithm 1 and Algorithm 2, where the parametric estimation step can be implemented with Algorithm 3. This allows the decoder to approximate the piecewise polynomial function from the quantized low-pass coefficients $\bar{c}_{J,n}$.

VI. SIMULATION RESULTS

This section presents the simulation results of the proposed centralized and distributed semi-parametric compression schemes, where the parametric estimation step is implemented with the FRI-based estimation algorithm presented in the previous section.

A. Parametric Estimation Algorithm

First, let us start by comparing the variance of an estimator that uses Algorithm 3 against the CRB described in Section III-B. In this simulation, the coefficients y_n of a step function $s(t)$ given by (7) are



(a) (b)

Fig. 4. (a) MSE in the estimation of t_0 using Algorithm 3 and the corresponding CRB; (b) MSE in the estimation of A using Algorithm 3 and the corresponding CRB.

obtained as $y_n = \langle s(t), \varphi_{J,n}(t) \rangle$ with $J = 6$, where $\varphi(t)$ is the first order B-spline scaling function. Note that a B-spline function of order $P \geq 0$ is given by

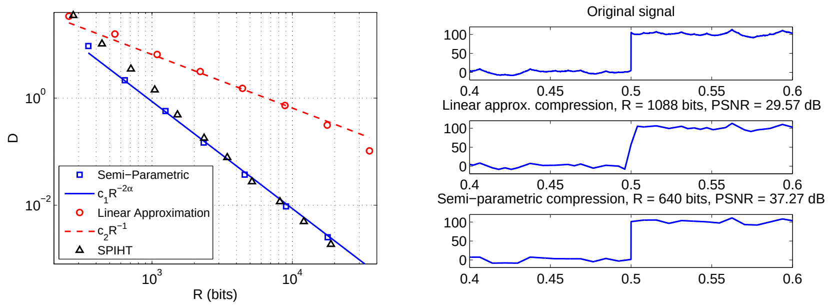
$$\beta(t) = \frac{1}{P!} \sum_{l=0}^{P+1} \binom{P+1}{l} (-1)^l (t-l)_+^P \quad \text{with} \quad (t)_+^P = \begin{cases} 0 & t < 0, \\ t^P & t \geq 0. \end{cases}$$

Gaussian noise with variance σ_ϵ^2 is then added to y_n. The values of the amplitude A and the location t_0 are estimated using Algorithm 3. Note that in our simulations, the parameter for the threshold was chosen empirically.

Fig. 4 shows the plots of the MSE and the corresponding CRB for the retrieval of t_0 and A. Note that the signal-to-noise ratio (SNR) is calculated as $10 \log_{10} \left(\frac{\text{var}[y_n]}{\sigma_\epsilon^2} \right)$ where var $[y_n]$ is the variance of y_n. For the estimation of t_0, our proposed algorithm exhibits the same decay as the CRB when the SNR reaches approximately $15dB$ even though the estimator does not achieve the lower bound. The MSE for the estimation of A follows the CRB even at low SNR.

B. Centralized Semi-Parametric Compression Algorithm

We now present the simulation results of the centralized semi-parametric compression scheme described in Algorithm 1, where the parametric estimation step is implemented with Algorithm 3. Note that in all our simulation, the function $f(t)$ is generated by adding the piecewise polynomial component $f_p(t)$ directly to the smooth component $f_\alpha(t)$. The smooth function $f_\alpha(t)$ is generated in the wavelet domain as follows: first, a set of coefficients $\{d_{J,n}\}$ at the coarsest decomposition level is generated with a random number generator; the rest of the coefficients at j-th decomposition level are then created by scaling the maximum value of the random number generator as $d_{j,n} \sim A2^{j(\alpha+1/2)}$ where $A = \max_n |d_{J,n}|$.



(a) (b)

Fig. 5. (a) $D(R)$ plots (log scale) for the compression of piecewise smooth function with one discontinuity and $\alpha = 1$; (b) plots of the original signal, the reconstructed signals with a linear approximation based scheme and a semi-parametric scheme.

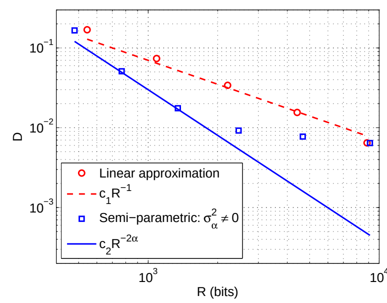
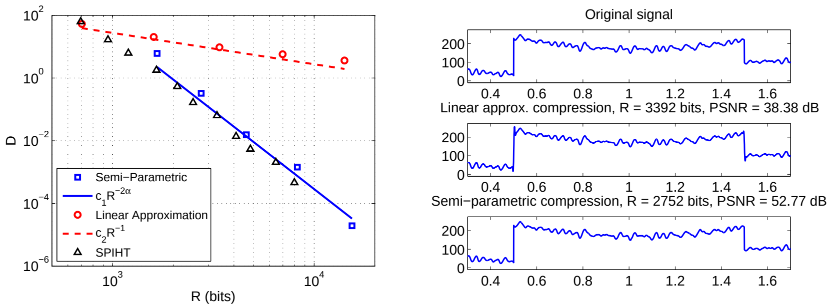


Fig. 6. $D(R)$ plots (log scale) for the compression of piecewise smooth function with one discontinuity and $\alpha = 0.95$.

First, we present the simulation results of the simplified signal model in (9) where we set $\alpha = 1$. A ten-level wavelet transform with a first order B-spline scaling function was used to decompose $f(t)$. Fig. 5 (a) shows the distortion-rate plot of the proposed scheme in comparison with the $D(R)$ curves of a linear and nonlinear approximation-based scheme. Note that we employ a 1-D version of the Set Partitioning In Hierarchical Trees (SPIHT) algorithm [30] to implement nonlinear approximation-based compression in all of our simulations. Our scheme achieves $D(R) \sim R^{-2\alpha}$, which is in line with our analysis in Section III-C, and the performance is comparable with SPIHT. The reconstructed functions are illustrated in Fig. 5 (b). Note that, in this simulation, the term σ_α^2 is insignificant.

Fig. 6 shows the $D(R)$ plot of the proposed scheme where σ_α^2 is not negligible. Here, $f(t)$ is given by (9) with $\alpha = 0.95$. The behavior of $D(R)$ changes from $D(R) \sim R^{-2\alpha}$ to $D(R) \sim C_1\sigma_\alpha + C_2\sigma_\alpha^2$ after a certain rate point. The value of C_1 and C_2 depend largely on the performance of the parametric estimation algorithm. Thus, a better parametric estimation algorithm allows a wider range of R where $D(R) \sim R^{-2\alpha}$.



(a) (b)

Fig. 7. (a) $D(R)$ plots (log scale) for the compression of piecewise smooth function with two discontinuities and $\alpha = 2.5$. (b) Plots of the original signal, the reconstructed signals with a linear approximation based scheme and a semi-parametric scheme.

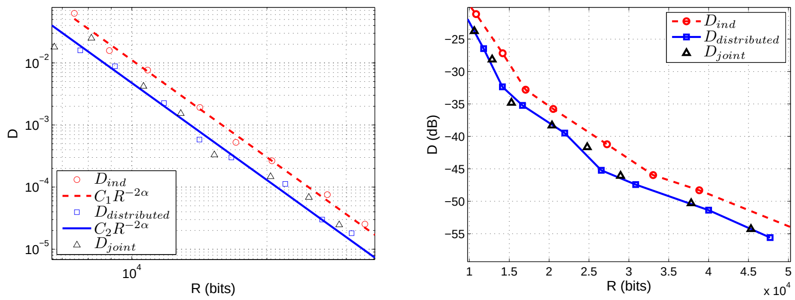
The simulation results for a function $f(t)$ described by (2) are given next. Here, $f_p(t)$ is a piecewise quadratic function with three pieces and $\alpha = 2.5$. We used a six-level wavelet decomposition with a second order B-spline scaling function. The parametric estimation for the two discontinuities in the function are done locally. The $D(R)$ curves are given in Fig. 7 (a), where the proposed scheme also achieves $D(R) \sim R^{-2\alpha}$. The reconstructed functions are shown in Fig. 7 (b).

C. Distributed Semi-Parametric Compression Algorithm

We now present the simulation result of the proposed distributed semi-parametric compression schemes. The parametric estimation step is implemented with Algorithm 3. Two piecewise smooth functions, $f_1(t)$ and $f_2(t)$, $t \in [0,1[$, are generated, where $f_2(t) = f_1(t-\tau) + \epsilon_\alpha(t)$ such that $\sup_{t\in[0,1[} |\epsilon_\alpha(t)| \leq \sup_{t\in[0,1[} \sqrt{\beta}|f_1(t)|$ with $\beta = 0.04$. Both functions contain two smooth pieces with $\alpha = 2.6$. We use a second order B-spline scaling function to decompose the signals up to six decomposition levels. Encoder 2 sets $\beta^* = 0.06$. Fig. 8 (a) shows $D(R)$ plots of the proposed scheme. The distortion of the proposed scheme is approximately 2.33 times lower when compared to the independent coding scheme, which is in line with the predicted gain of 2.26 times from Summary 2. The result also shows that the achieved distortion is very close to that of the joint encoding scheme. This is because β^* is well calibrated to be close to β. From Fig. 8 (b), at high rates, our scheme outperforms the independent scheme by approximately 3 dB where the gain predicted in our analysis is 3.5 dB.

VII. Application to real images

In this section, we briefly discuss how the distributed semi-parametric compression scheme presented in this paper can be extended to a set of real images or video. Firstly, we acknowledge that some textures



(a) (b)

Fig. 8. $D(R)$ plots of compression schemes based on disparity-by-translation model with prediction error (log-log scale in (a) and decibel scale in (b)).

of an image do not entirely fit the piecewise smooth model but are closer to noise. The encoding strategy presented in algorithm 2, therefore, has to be slightly modified by using a non-linear approximation based quantization strategy rather than a linear one. The disparity between each image can be described with a motion vector. As an alternative to the FRI-based method, here the parametric estimation at the decoder is performed using a typical maximum likelihood block-based motion estimation. Fig. 9 presents a set of preliminary results for the distributed compression of the stereo images in Fig. 3. We used a SPIHT-based algorithm to compress both images. At the joint decoder, the prediction of the second image is obtained from the first image by using a simple block-based motion estimation. Error correction code as the one discussed in Section III-A is then applied to fully decode the residual of the second image, where the power of the residual is estimated a priori. For this set of images, we found that the performance in terms of PSNR is better than an independent compression scheme.

VIII. Conclusion

The objective of this paper is to develop a new approach to centralized and distributed compression using wavelets. First, a new centralized semi-parametric compression algorithm for piecewise smooth functions has been proposed. The encoder of the proposed algorithm uses a wavelet-based linear approximation strategy. The decoder is, instead, nonlinear and employs a parametric estimation technique to reconstruct the singular structure of the observed function. Our analysis shows that the $D(R)$ function of the proposed scheme achieves a dominating decay rate of $R^{-2\alpha}$ for a wide range of rates, which is comparable to that of a conventional compression scheme with a nonlinear encoder and a linear decoder. The concept of semi-parametric compression has then been extended to a distributed compression scenario,

(a) (b) (c)

Fig. 9. Preliminary results of the extension of the proposed distributed compression scheme to real images: (a) plot of the PSNRs of the independent compression scheme versus our distributed semi-parametric scheme; (b) the reconstructed second image using the semi-parametric distributed compression at 0.07 bpp with the PSNR of 31.9 dB; (c) the reconstructed second image using independent SPIHT compression 0.066 bpp with the PSNR of 28.35 dB.

where we modeled the disparity between each observed signal with a shift and a prediction error, which is uniformly smooth. By using this model, we have been able to provide precise answers on how the total rate has to be split between the different encoders, and devise an encoding and decoding strategy that achieves the same performance of a joint encoding scheme for a wide range of rates.

REFERENCES

[1] V. Chaisinthop and P. L. Dragotti, "Semi-parametric compression of piecewise smooth functions," *Proceedings of European Conference on Signal Processing (EUSIPCO)*, August 2009.

[2] V. K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18(5), pp. 9–21, September 2001.

[3] J. Huang and P. Schultheiss, "Block quantization of correlated gaussian random variables," *IEEE Transactions on Communications Systems*, vol. 11, no. 3, pp. 289 –296, September 1963.

[4] R. Puri and K. Ramchandran, "PRISM: a video coding architecture based on distributed compression principles," in *40th Allerton Conference on Communication, Control and Computing*, Allerton, IL, October 2002.

[5] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436 –2448, oct. 2007.

[6] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[7] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, "Distributed monoview and multiview video coding," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 67–76, 2007.

[8] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[9] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[10] P. L. Dragotti and M. Gastpar, *Distributed Source Coding: Theory, Algorithms and Applications*. Academic Press, January 2009.

[11] M. Gastpar, P. L. Dragotti, and M. Vetterli, “The distributed Karhunen Loève transform,” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5177–5196, December 2006.

[12] M. Flierl and P. Vandergheynst, “Distributed coding of highly correlated image sequences with motion-compensated temporal wavelets,” *Eurasip Journal Applied Signal Processing*, vol. 2006, September 2006.

[13] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, “High-rate quantization and transform coding with side information at the decoder,” *EURASIP Journal on Signal Processing*, vol. 86 (11), pp. 3160–3179, November 2006.

[14] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.

[15] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, 2006.

[16] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[17] P. L. Dragotti, M. Vetterli, and T. Blu, “Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix,” *IEEE Transactions on Signal Processing*, vol. 55(5), pp. 1741–1757, May 2007.

[18] M. F. Duarte, S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, “Joint sparsity models for distributed compressed sensing,” in *Online Proceedings of the Workshop on Signal Processing with Adaptative Sparse Structured Representations (SPARS)*, Rennes, France, 2005.

[19] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, “Distributed compressive sensing,” *CoRR*, vol. abs/0901.3403, 2009.

[20] M. Vetterli, “Wavelets, approximation, and compression,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 59–73, 2001.

[21] A. Cohen, I. Daubechies, O. Guleryuz, and M. Orchard, “On the importance of combining wavelet-based nonlinear approximation with coding strategies,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1895–1921, July 2002.

[22] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, December 2008.

[23] P. L. Dragotti and M. Vetterli, “Wavelet footprints: theory, algorithms, and applications,” *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1306–1323, May 2003.

[24] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice Hall PTR, April 1995.

[25] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.

[26] M. Unser and T. Blu, “Wavelet theory demystified,” *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 470–483, 2003.

[27] L. Coulot, M. Vetterli, T. Blu, and P. L. Dragotti, “Sampling signals with finite rate of innovation in presence of noise,” *École Polytechnique Fédérale de Lausanne (EPFL) Technical Report*, 2007.

[28] P. L. Dragotti and F. Homann, “Sampling signals with finite rate of innovation in the presence of noise,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, 2009, pp. 2941 –2944.

[29] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[30] A. Said and W. Pearlman, “A new, fast, and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.