

## Influence of Unknown Exterior Samples on Interpolated Values for Band-limited Images

Loïc Simon, Jean-Michel Morel

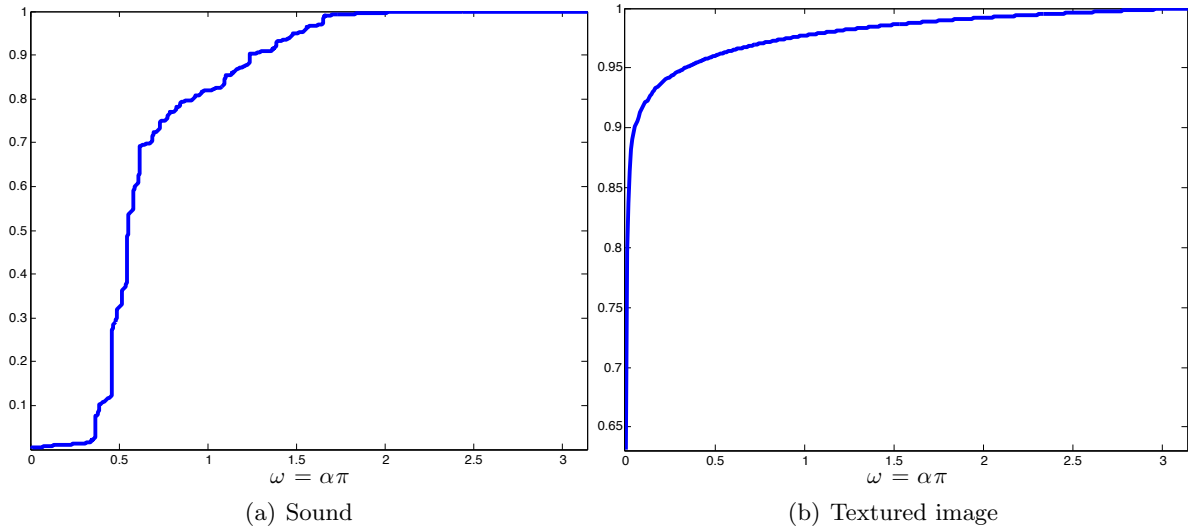
**Abstract.** The growing size of digital images and their increasing information content in terms of bits per pixel (or SNR) leads to ask to which extent the known samples permit to restore the underlying continuous image. In the context of band-limited data the Shannon-Whittaker theory gives an adequate theoretical answer provided that infinitely many samples are measured. Yet, we show that the current accuracy of digital images will be limited in the future by the truncation error, which is no-more negligible with respect to other perturbations such as quantization or aliasing. To do so, we propose a method to estimate the truncation error. All of our results are expressed in terms of Root Mean Squared Error (RMSE) under the common hypothesis of *band-limited weakly stationary random processes*. As a first contribution, we present a general expression of the truncation RMSE involving the spectral content of the image. We then derive a simple and generic scheme to evaluate bounds on the truncation error. The actual computation of error bounds is conducted for two standard interpolation schemes, namely the Shannon-Whittaker and the DFT interpolators. These theoretical bounds reveal a specific decay of the truncation error as a function of the distance from the sample to the image boundary. The tight estimates obtained and validated on a set of experiments confirm that the truncation error can become the main error term in HDR images. In classic 8 bit images it is bound by the quantization error at a moderate distance from the image boundary, but still requires large images to become manageable.

**Key words.** Truncation Error, RMSE, Image Interpolation, Band-Limited Random Processes

**AMS subject classifications.** 68U10, 65D05, 65T40

**1. Introduction.** Even if digital image acquisition (capture, storage, transmission) is in essence discrete, image processing remains deeply continuous. Continuous resampling operations constitute a key element of the vast majority of high-level processing tasks (such as image registration, stitching, or stereo reconstruction). To succeed in the computation of an image transformation, one must be able to reconstruct the underlying continuous image from the available samples. In fact, it is never possible to truly recover the continuous data, since errors of several kinds occur. Generally speaking one can refer to at least three major sources of errors: photon counting, quantization and aliasing.

The first two relate to the way CCD captors operate individually, while the last one depends on a relation between the step of the grid of captors and the highest frequencies in the signal (after optical filtering). It is well known that the photon count is well approximated by a Poisson distribution. As a consequence, after normalization in the image dynamic range (for instance  $[0, 255]$  for a 8-bit camera) the standard deviation of the pixel value is of the form  $\sqrt{\lambda X}$  where  $X$  is the error-free response and  $\lambda$  is a factor depending on various capture settings (*e.g.* the exposure time, the brightness of the scene). Quantization errors however are independent of these settings: the error follows a uniform distribution on  $[-0.5, 0.5]$ . Therefore it may act as a standard reference. For comparison purposes, we provide the



**Figure 1.1.** Estimation of the cumulative spectral power for a sound signal (a) and a textured image (b).

standard deviation of the quantization error:

$$\sqrt{\mathbb{E}[(q(X) - X)^2]} = \frac{1}{2\sqrt{3}} \simeq 0.29,$$

where  $X$  is the original signal and  $q(X)$  the quantized one.

Aliasing relates merely to the fact that an infinite number of signals (referred to as aliases) correspond to the same collection of sample values. Such a definition is very general and simply states that interpolation is doomed by an intrinsic ambiguity. In the common usage, the term aliasing is dedicated to the more particular situation where an infinite number of samples are available. Even in such circumstances, the ambiguity remains serious since, given a particular reconstruction, one can build aliases by applying frequency shifts that are multiples of the sampling frequency  $\omega_s := 2\pi/a$  ( $a$  being the step of the sampling grid). The ambiguity is therefore totally harnessed if one restricts the possible signals to those with a spectrum included in a predefined range of length  $\omega_s$ . Typically, one considers the class of Nyquist band-limited signals, that is those which highest frequency modulus is  $\omega_{Nyq} := \frac{\omega_s}{2}$ .

These three classical sources of error are well studied [15] and one could claim that in current good-quality cameras they are becoming rather negligible. Aliasing can be reduced thanks to the combination of very small captors and optical anti-aliasing filters. Photon counting errors can be lessened by increasing the exposure time or by high dynamic range fusion techniques. Last, since the advent of 16-bit cameras, the quantization error is alleviated relatively to the dynamic range (which passes from 256 to  $2^{16}$ ). Given these facts, it should be virtually possible to obtain a perfect reconstruction using the Shannon-Whittaker reconstruction (recalled later on). Nonetheless the real situation in which images are captured is more complicated because only a finite number of samples are known. In this article, we will focus specifically on this additional source of error commonly known as the *truncation error*.

The truncation error has been thoroughly studied, ever since the sixties. However, the published results are scarcely known by the image processing community. And while the truncation error is often happily ignored, we shall see that it should not be overlooked given the average size of current images. This statement is all the more valid for 16-bit images. In order to motivate our study, let us consider quickly a few early works on the truncation error, such as [18], [14], [26], [9] and [10]. These articles already provide instructive information on the truncation error. Unfortunately, many of them employ deterministic models such as bounded signals [26] or finite-energy signals [18, 14, 9]. Those are perfectly valid for images but they lead to upper-bound expressions that cannot be evaluated from the known samples. A better approach consists in relaxing the deterministic assumption and requiring that the signal is statistically stationary [10]. This model is commonly accepted with respect to images and presents the great advantage that global properties can be well estimated from a local knowledge of the signal.

Along the same line, many articles ([13, 26, 11, 9, 19]) assume that the signal is not only band-limited but is also oversampled so that its highest frequency is strictly less than the Nyquist rate. By plotting the signal cumulative spectral power  $\int_{|\omega| \leq \alpha\pi} d\Psi(\omega)$  against bandwidth  $\alpha\pi$ , one can visually check the amount of oversampling. For example Figure 1.1 suggests that the audio signal (left plot) is oversampled since the full intensity is reached around  $\alpha = 2/3$ . Differently, the textured image (right plot) contains power up to the Nyquist bandwidth. This behavior is typical of textures because they involve fast local variations similar to white-noise (*c.f.* Figure 1.2). Published upper bounds that do not rely on oversampling [18, 10] are few and they imply that the truncation error decreases more slowly with the size of the signal than in the case of oversampled signals. In this study, we confirm this observation in a theoretical context adapted to images. We also give experimental hindsight of the significance of the truncation error for image processing applications.

Most of the published works consider the truncation error involved in interpolation methods that are not common practice (at least not in image processing). The typical case concerns the truncated Shannon-Whittaker series or variants referred to as self-truncating series. In practice, other interpolation methods for instance based on FFT are preferred. The most classical one is certainly the DFT zero-padding. To the best of our knowledge, upper-bounds concerning this interpolation method have only been studied in [25]. Their approach is very similar to [26] and likewise their results are restricted to oversampled signals. In the supplementary file, we present a complementary bibliographic study in order to place the truncation error in a broader context (such as approximation theory [20, 16]) as well as to discuss more recent works (*e.g.* [27, 19, 17, 4]) on the subject.

Even in recent works, truncation error is never considered for popular interpolators such as B-splines. Several papers [1, 23] do provide a deep insight concerning the aliasing error. A comprehensive summary of the corresponding results can be found in [22] where the problem is presented in the context of approximating a continuous signal in the space  $V(h)$  generated by the integer shifts of a predefined interpolation kernel  $h$ . One admirable finding (*c.f.* [22, Theorem 2]) states that several projectors on  $V(h)$  can be obtained by combining a pre-filter  $h_a$ , an infinite sampling, a digital corrective filter  $p$  and the filter  $h$ . Usually, the pre-filter is fixed by the device, so that only the digital filter  $p$  can be chosen freely. By doing so, one can achieve the projection of the true continuous signal onto  $V(h)$  perpendicularly to  $V(h_a)$ .



**Figure 1.2.** A textured patch of an image looks alike uncorrelated random variables (a.k.a white-noise).

In this paper, we admit that current cameras are set up to produce images that are practically band-limited. In other words, the anti-aliasing prefilter  $h_a$  is assumed to act as an ideal low-pass filter. As a result, we consider the so-obtained band-limited signal as the reference signal for all our error computations. In doing so we ignore the error that was already introduced by the pre-filter. In that context, we would like to estimate the behavior and the typical magnitude of the truncation error. More precisely, given a band-limited signal  $X_t$  dependent on the continuous variable  $t$ , we consider the error between  $X_t$  and any reconstruction  $\tilde{X}_t$  obtained from the samples  $(X_k)_{|k| \leq K}$ . On the one hand, our context is more limited than the one considered by Unser et al. because we concentrate on a specific pre-filter that is sinc and because we only consider errors occurring after prefiltering. But on the other hand, we are considering a more realistic situation where only finitely many samples are acquired. In summary, the error that we wish to estimate can be expressed as follows<sup>1</sup>:

$$\epsilon_t = \left| \sum_{k \in \mathbb{Z}} X_k \text{sinc}(t - k) - \sum_{|k| \leq K} X_k h(t - k) \right|.$$

As hinted above, our effort is led by two key specific aspects of image processing which are

---

<sup>1</sup>Beware that truncation error name could falsely convey the following non equivalent expression  $\left| \sum_{k \in \mathbb{Z}} X_k h(t - k) - \sum_{|k| \leq K} X_k h(t - k) \right|$  where the filter  $h$  appears in the finite and the infinite expansion.

the potential absence of oversampling and the seldom use of truncated Shannon-Whittaker interpolation. Section 2 provides the necessary background to understand the problem at stake. In particular, a few classical results on Random Processes are recalled in order to introduce several generic tools that are useful in the rest of our derivations. The principal one is a spectral-representation of the reconstruction error (see Theorem 2.12). Starting from section 3 until the end, we narrow our perspectives to band-limited processes and the associated truncation error (cf Theorem 3.1). After giving a grasp of the expected behavior of the truncation error, we provide an effective way to decompose a signal into spectral components that shall be practical for accurately estimating the error (see Figure 3.1). Then, two technical sections are dedicated to finding accurate upper-bounds for the truncation error with and without over-sampling. On the one hand, Section 4 deals with the classical truncated Shannon-Whittaker interpolation. And on the other hand, Section 5 tackles the more wide-spread zero-padding interpolation. The most important results are recapped in Theorem 4.9 and 5.15. Given the high technicality of the upper-bound derivations, they were first obtained for univariate Random Processes. Section 6 provides a generic way to transform such univariate results into higher-dimensions. This operation allows us to verify that the truncation error remains mostly influenced by the so-called white-noise component of the signal.

In addition to the previous theoretical contributions, we have also designed a generic experimental protocol. This protocol is presented in Section 7 along with selected examples that demonstrate the accuracy of our bounds. Besides their validation character, the experiments also give a down-to-earth feeling of the degree of errors due to truncation for real images. For the sake of comparison, we provide empirical estimates of the truncation error for a naive implementation of B-splines (that ignores the procedures recommended by [22]). This study is of interest of course only when noise conditions and aliasing do not constitute the major cause of error. In the opposite case, the use of local interpolators and of wavelets is fully justified ([7, 8]), as a more global interpolation would only propagate errors.

## 2. General context of linear shift-invariant interpolation.

**2.1. Useful notations.** Throughout the paper,  $t \in \mathbb{R}^d$  will refer to a generic location in the domain of a multivariate signal, while  $k \in \mathbb{Z}^d$  will stand for an integer location in the same domain. We will also use the standard notation  $\omega \in \mathbb{R}^d$  to represent angular frequencies. The Fourier transform of a deterministic signal  $x_t$  will be denoted by  $\mathcal{F}(x)$  and defined as  $\hat{x}_\omega := \mathcal{F}(x)(\omega) := \int e^{-i\omega t} x_t dt$ . With such conventions, the inverse Fourier transform is expressed as

$$\mathcal{F}^{-1}(\hat{x})(t) := \frac{1}{(2\pi)^d} \int e^{i\omega t} \hat{x}_\omega d\omega.$$

In both formulas,  $\omega t$  denotes the scalar product of  $\omega$  and  $t$ . We will also consider the discrete Fourier transforms:

$$\mathcal{DFT}(x)[l] := \sum_{|k| \leq K} e^{-i\frac{2\pi l}{N} k} x[k]$$

and

$$\mathcal{DFT}^{-1}(\hat{x})[k] := \frac{1}{N} \sum_{|l| \leq K} e^{i\frac{2\pi l}{N} k} \hat{x}[l],$$

where  $K \in (\mathbb{Z}^+)^d$ ,  $N := (2K_1 + 1) \times \cdots \times (2K_d + 1)$  and to simplify the notations, we have assumed that the signal was sampled symmetrically. This simplifying assumption will be kept throughout the paper.

Many results in the paper will be presented in the multivariate case  $d > 1$ . In order to keep the derivations simple it is useful to choose multivariate conventions that extend the univariate notations in a lightweight manner. In particular, we have chosen the same notation for the scalar product when  $d > 1$  and classical product when  $d = 1$ . Similarly, we adopt appropriate notations for comparison operators. If  $t, t' \in \mathbb{R}^d$  and  $T \in (R^+)^d$  then

- $t \leq t' \Leftrightarrow t' \geq t \Leftrightarrow \forall i \in \{1, \dots, d\}, t_i \leq t'_i$
- $|t| \leq T \Leftrightarrow \forall i \in \{1, \dots, d\}, |t_i| \leq T_i$
- $|t| > T \Leftrightarrow \neg(|t| \leq T) \Leftrightarrow \exists i \in \{1, \dots, d\}, |t_i| > T_i$
- in all other contexts,  $|t| := \|t\|_2$  represents the Euclidean norm of  $t$ .

In this fashion, the notations concerning  $|t|$  are context dependent. In particular, to ensure that  $\leq$  and  $>$  remain consistent some symmetry was broken in the quantifiers. These conventions are kept unchanged for integer locations  $k$  and for angular frequencies  $\omega$ .

In the univariate context  $d = 1$ , the previous notations are of course consistent with their typical meaning. They are actually mere separable extensions of univariate notations. We will use separable extensions in a more systematic way, when dealing with functions of multivariables. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ , we will continue to denote  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  the function such that  $f(t) = \prod_{i=1}^d f(t_i)$ . For example  $\sin(t) := \prod_{i=1}^d \sin(t_i)$ .

In the same vein, given a law  $r$  acting on two multivariate variables  $t \in \mathbb{R}^d$  and  $t' \in \mathbb{R}^d$ , we will sometimes extend the definition of  $r$  to one multivariate and one univariate argument. To do so, starting from variables  $u \in \mathbb{R}$  and  $u' \in \mathbb{R}$ , we can create vectors of the right size by replicating the univariate variables  $d$  times:  $t_u := (u, \dots, u)$ ,  $t'_u := (u', \dots, u')$ . Then, we define  $r(u, t') := r(t_u, t')$  and  $r(t, u') := r(t, t'_u)$ . This convention shall be used for arithmetic laws and comparison operators. For instance, the notation  $|\omega| \leq \pi$  means  $\forall i \in \{1, \dots, d\}, |\omega_i| \leq \pi$ .

**2.2. Classical results on Random Processes.** As usual when dealing with missing data, we will use stochastic models to account for the unknown values of an image. Therefore, in what follows,  $X_t$  stands for a random process, where  $t \in \mathbb{R}^d$  is either a time or space variable. Our definitions follow [12].

**Definition 2.1 (Random processes).** *Let  $X_t$  be a random process (RP). We refer to  $\mathbb{E}[X_t]$  and  $\mathbb{E}[X_\tau X_{\tau+t}]$  as respectively the first and second order statistics of  $X_t$ .*

*A random process  $X_t$  is said to be weakly stationary if its first and second order statistics are shift invariant*

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[X_{t'}] \\ \mathbb{E}[X_\tau X_{\tau+t}] &= \mathbb{E}[X_{\tau'} X_{\tau'+t}]. \end{aligned}$$

*We will generically denote by  $\mu := \mathbb{E}[X_t]$  its average and by  $R_X(t) := \mathbb{E}[(X_\tau - \mu)(X_{\tau+t} - \mu)]$  its auto-correlation function (which is easily shown to be symmetric positive semi-definite).*

**Remark 2.1.** *Unless specified otherwise, all random processes will be assumed weakly stationary.*

We now recall the definition of the notion of convergence that will be most useful.



**Definition 2.2 (Convergence in quadratic mean).** Let  $X_1, \dots, X_n, \dots$  a sequence of random variables and  $X$  another random variable. We say that  $X_n \rightarrow X$  in quadratic mean (or q.m.) iff

$$\mathbb{E} \left[ |X_n - X|^2 \right] \rightarrow 0.$$

By extension, a sequence of RPs  $X_t^n$  converges to  $X_t$  at  $t$  iff

$$\mathbb{E} \left[ |X_t^n - X_t|^2 \right] \rightarrow 0.$$

Furthermore, a RP  $X_t$  is continuous in quadratic mean iff  $\forall t X_s \rightarrow X_t$  in quadratic mean when  $s \rightarrow t$ .

**Proposition 2.3.** Let  $X_t$  be a weakly stationary RP. Then the following assertions are equivalent:

1.  $X_t$  is continuous in quadratic mean
2.  $R_X$  is continuous at  $t = 0$
3.  $R_X$  is uniformly continuous on  $\mathbb{R}$

*Proof.* See supplementary file or [6] (p.143). ■

**Definition 2.4 (Power spectral distribution).** Let  $X_t$  be a weakly stationary RP continuous in quadratic mean. We denote  $d\Psi_X(\omega) := \mathcal{F}(R_X)(\omega)$  and refer to it as the power spectral distribution (P.S.D) of  $X_t$ . Moreover,

- $X_t$  is said to be Nyquist band-limited iff  $\text{supp}(d\Psi_X) \subset \{|\omega| \leq \pi\} := \Omega_\pi$ .
- If in addition,  $d\Psi_X(\partial\Omega_\pi) = 0$ , then  $X_t$  is said strictly Nyquist band-limited.

**Remark 2.2.** Given the assumption in the previous definition,  $d\Psi_X$  is necessarily well-defined because the auto-covariance is a continuous (thanks to Proposition 2.3) and bounded function. Besides, according to Bochner's theorem it is a positive and finite Borel measure, hence the notation  $d\Psi$ .

**2.3. Sampling and reconstruction.** The action of extracting samples of a random process is mathematically equivalent to the multiplication with a Dirac comb.

**Definition 2.5 (Dirac Combs).** Let  $0 < K \leq \infty$ . We will refer to the following distribution as the Dirac comb of half-length  $K$

$$\Delta_K = \sum_{k \in \mathbb{Z}^d, |k| \leq K} \delta_k.$$

Whenever it makes sense, we call the sampled version of a random process  $X_t$  the following random distribution

$$X.\Delta_K := \sum_{|k| \leq K} X_k \delta_k.$$

**Remark 2.3.** The sampled version is mathematically well defined for any random process without any continuity assumption. Of course, some operations on the sampled version, such as taking limits in the distribution sense, should not be permitted unless  $X_t$  is almost surely continuous. This is the case for band-limited random processes which are almost surely analytic

[3]. In this article, we do not rely on such operations, and the sampled version can be thought as a convenient formal notation to encode the whole sequence of samples.

The action of recovering a continuous signal from its samples is known as a reconstruction. Among all possible reconstruction schemes, the class of linear and shift invariant reconstructions is often favored for practical reasons.

**Definition 2.6 (Reconstruction Systems).** We call:

- reconstruction kernel, any function  $h : \mathbb{R} \mapsto \mathbb{R}$
- reconstruction system, any sequence  $(h_K)_{K>0}$  of reconstruction kernels.

Besides, a reconstruction system

- is said interpolating iff

$$\forall K > 0, \forall |k| \leq K, h_K(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases},$$

- reproduces constants iff

$$\forall K > 0, \forall |t| \leq K, \sum_{|k| \leq K} h_K(t - k) = 1.$$

Finally, we say that a reconstruction system  $(h_K)$  satisfies the reconstruction systems dominated condition if  $\exists h_\infty$  such that

- $\forall t, \forall |\omega| \leq \pi, \lim_{K \rightarrow \infty} \sum_{|k| \leq K} e^{i\omega k} h_K(t - k) = \sum_k e^{i\omega k} h_\infty(t - k)$  and
- $\forall t, \exists M(t) > 0$  such that

$$\forall K > 0, \forall |\omega| \leq \pi, \left| \sum_{|k| \leq K} e^{i\omega k} h_K(t - k) \right| \leq M(t).$$

**Definition 2.7 (Linear shift-invariant reconstruction).** Let  $X_t$  be a random process,  $(h_K)$  a reconstruction system and  $0 < K < \infty$ . We define the reconstruction of  $X_t$  from its samples  $X_k, k \in \{-K, \dots, K\}$  by<sup>2</sup>

$$\tilde{X}_t^{[h_K, K]} := (X \cdot \Delta_K) * h_K(t) := \sum_{|k| \leq K} X_k h_K(t - k).$$

Then, if the system verifies the reconstruction systems dominated condition, we can extend the previous definition for  $K = \infty$

$$\tilde{X}_t^{[h_\infty, \infty]} := (X \cdot \Delta_\infty) * h_\infty(t) = \lim_{K \rightarrow \infty} (X \cdot \Delta_K) * h_K(t) = \lim_{K \rightarrow \infty} \sum_{|k| \leq K} X_k h_K(t - k),$$

where convergence holds in quadratic mean.

**Remark 2.4.**

- If  $h_K$  is interpolating, then the reconstruction is flawless at sample positions  $|k| \leq K$ .

---

<sup>2</sup>The notation  $(X \cdot \Delta_K) * h_K(t)$  is purely formal.



- If  $h_K$  reproduces constants, then the reconstruction is perfect for constant inputs (the bias is null).
- In the previous definition, the limit case ( $K = \infty$ ) requires proving the convergence in quadratic mean (see the supplementary file).

**Definition 2.8 (RMSE).** Let  $X_t$  be a random process and  $\tilde{X}_t$  an arbitrary reconstruction. We will evaluate the quality of this reconstruction in terms of RMSE:

$$RMSE_{[\tilde{X}]}(t)^2 := \mathbb{E}[(X_t - \tilde{X}_t)^2]$$

The ideal low-pass filter plays a central role in the theory of sampling and reconstruction.

**Proposition 2.9.** The Shannon-Whittaker reconstruction system ( $h_K(t) = \text{sinc}(t) := \frac{\sin(\pi t)}{\pi t}$ ) is interpolating and verifies the reconstruction systems dominated condition. Note that it reproduces constants only asymptotically as  $K \rightarrow \infty$ .

*Proof.* See the supplementary file. ■

**Definition 2.10 (Shannon-Whittaker interpolation).** We will call ideal Shannon-Whittaker interpolation of  $X_t$ ,

$$\tilde{X}_t^{[\text{sinc}, \infty]} := \lim_{K \rightarrow \infty} \sum_{|k| \leq K} X_k \text{sinc}(t - k)$$

in quadratic mean, that is to say

$$\lim_{K \rightarrow \infty} \mathbb{E} \left[ \left( \tilde{X}_t^{[\text{sinc}, \infty]} - \sum_{|k| \leq K} X_k \text{sinc}(t - k) \right)^2 \right] = 0.$$

For the sake of completeness we provide the sampling theorem for random processes. A proof can be found in [2], but we will obtain it as a by-product of our RMSE estimations.

**Theorem 2.11 (Sampling theorem).** Let  $X_t$  be a (not necessarily strictly) Nyquist band-limited random process. Then, uniformly when  $t$  is bounded and in quadratic mean,

$$X_t = \tilde{X}_t^{[\text{sinc}, \infty]}.$$

**2.4. Generalized aliasing error.** The reconstruction error can be considered under the angle of a generalized aliasing. Given a certain number of samples, there are an infinite number of continuous signals called aliases that agree with these measurements. Any interpolated reconstruction is but a particular alias of the unknown continuous signal. It is useful to make a clear distinction between two specific types of aliasing. On the one hand, what we shall call the *classical aliasing*, corresponds to situations where an infinite number of samples are known but the signal spectrum is not bounded into the Nyquist domain. On the other hand, what we shall call *truncation aliasing* corresponds to Nyquist band-limited signals sampled only at a finite number of locations.

We now narrow our analysis to linear shift-invariant reconstruction. In such a case, we will derive a general representation of the generalized aliasing RMSE. This representation will be referred to as the ‘‘spectral representation’’ due to the central role of the spectral content in its expression. The next theorem has an intuitive interpretation. It merely states that the mean squared error is the sum of the squared errors with respect to the average value of  $X$  and with respect to every pure harmonic  $e^{i\omega t}$  (weighted by the spectrum of  $X$ ).

**Theorem 2.12 (Spectral representation of the generalized aliasing RMSE).** *Let  $X_t$  be a random process of average  $\mu$  and power spectrum  $d\Psi_X$ ,  $K < \infty$  and  $h_K$  a reconstruction system ( $h_K$  is not required to be interpolating nor to preserve constants). Then the mean squared error of the reconstruction  $\tilde{X}_t^{[h_K, K]} = (X \cdot \Delta_K) * h_K(t)$  is*

$$RMSE_{[X \cdot \Delta_K * h_K]}(t)^2 = |\mu - \mu \Delta_K * h_K(t)|^2 + \frac{1}{(2\pi)^d} \int |e^{i\omega t} - [(e^{i\omega} \cdot \Delta_K) * h_K](t)|^2 d\Psi_X(\omega) \quad (2.1)$$

The formula remains true if  $K = \infty$  under the reconstruction systems dominated condition (given in Definition 2.6).

*Proof.* See the supplementary file. ■

**Definition 2.13 (Average and power spectral MSE).** *In what follows, we set*

$$\begin{aligned} MSE_\mu(t)(h_K) &:= \mu^2 |1 - \Delta_K * h_K(t)|^2, \\ MSE_{d\Psi}(t)(h_K) &:= \frac{1}{(2\pi)^d} \int |e^{i\omega t} - (e^{i\omega} \cdot \Delta_K) * h_K(t)|^2 d\Psi(\omega). \end{aligned}$$

We will refer to them respectively as the average (or constant) MSE component and the power spectral MSE component. They correspond to the squared bias and the variance.

We open a short parenthesis on the classical situation of aliasing with an infinite number of samples, and we consider the ideal reconstruction biased towards recovering low-frequencies (that is to say that among all the ‘‘aliases’’ that are consistent with the samples, we chose the one whose spectrum is in the Nyquist band by using  $h = \text{sinc}$ ). We obtain the following result.

**Corollary 2.14.** *Let  $X_t$  be a random process, then the mean squared error of the ideal Shannon-Whittaker reconstruction is*

$$RMSE_{[(X \cdot \Delta_\infty) * \text{sinc}]}(t)^2 = \frac{1}{(2\pi)^d} \int_{|\omega| \geq \pi} |e^{i\omega t} - [(e^{i\omega} \cdot \Delta_\infty) * \text{sinc}](t)|^2 d\Psi_X(\omega)$$

*Proof.* This is a direct consequence of Theorem 2.12 (we can use it for  $K = \infty$  since  $\text{sinc}$  verifies the reconstruction systems dominated condition) and the fact that the Shannon-Whittaker reconstruction ‘‘asymptotically’’ preserves constants (i.e.  $\sum_{k \in \mathbb{Z}^d} \text{sinc}(t - k) \equiv 1$ ) which in turn implies that the average MSE is null. ■

### 3. Case of band-limited processes.

**3.1. Spatial truncation error.** Note that the formula derived in the previous part for the power spectral MSE component (Definition 2.13) is not prone to practical evaluations, because the spectrum of the signal outside the Nyquist band cannot be estimated from a digital signal. This is precisely due to the classical aliasing ambiguity itself. Several safeguards can be adopted to rule out this ambiguity. Given the current trend of CCD cameras, we choose to assume that the signal spectrum vanishes outside the Nyquist band. In other words, we will assume that the process under consideration can be reconstructed perfectly with the ideal Shannon-Whittaker formula:  $X_t = (X \cdot \Delta_\infty) * \text{sinc}$ .

**Theorem 3.1 (Spectral representation of the truncation RMSE).** *Let  $X_t$  be a strictly Nyquist band-limited random process of average  $\mu$ , and  $(h_K)$  a reconstruction system. Then*

$$\begin{aligned} \text{RMSE}_{[X \cdot \Delta_K * h_K]}(t)^2 &= |\mu \Delta_\infty * \text{sinc}(t) - \mu \Delta_K * h_K(t)|^2 \\ &\quad + \frac{1}{(2\pi)^d} \int_{|\omega| \leq \pi} |(e^{i\omega \cdot} \Delta_\infty) * \text{sinc}(t) - (e^{i\omega \cdot} \Delta_K) * h_K(t)|^2 d\Psi_X(\omega). \end{aligned}$$

*Proof.* This theorem is in fact a corollary of Theorem 2.12. Only we have emphasized that  $\mu = \mu \Delta_\infty * \text{sinc}(t)$ , and if  $|\omega| < \pi$   $e^{i\omega t} = [(e^{i\omega \cdot} \Delta_\infty) * \text{sinc}](t)$ . The previous equality is not true for  $|\omega| = \pi$  hence the strict band-limited hypothesis. ■

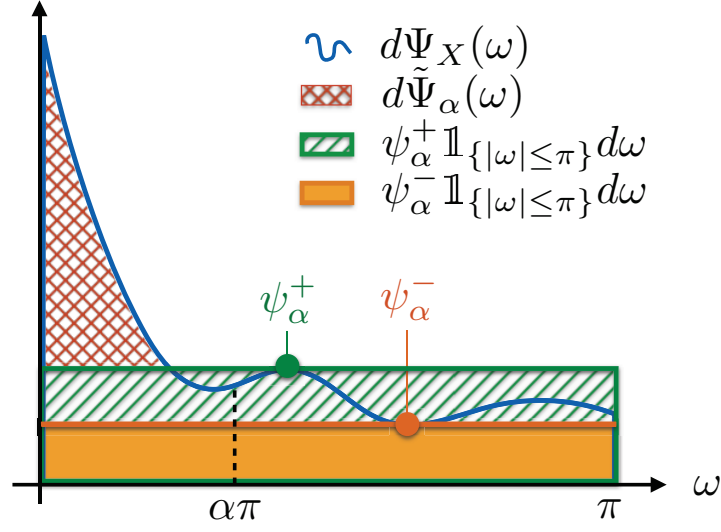
**3.2. Expected general behavior.** The major obstacle encountered when dealing with interpolation of Nyquist band-limited signals relates to the lack of knowledge about the samples outside the signal domain ( $X_k$  for  $|k| > K$ ). One simple approach is to use the Shannon-Whittaker reconstruction, by extending our knowledge about the samples in an arbitrary way. In this paper, we consider two such ways:

- the null extension, where the  $X_k$ 's are assumed null outside the domain:  $X_k = 0$  if  $|k| > K$ ,
- and the periodic extension, where, the  $X_k$ 's (and hence the whole signal  $X_t$ ) are assumed periodic:  $X_k = X_{k \% N}$  if  $|k| > K$ , where  $N = 2K + 1$  is the digitized signal length, and  $k \% N$  is the only integer  $k' \in \{-K, \dots, K\}$  such that  $k - k' \in N\mathbb{Z}$ .

In any case, the error should increase as we get closer to the border of the signal domain. Indeed any arbitrary extension brings but false contributions outside the domain. In addition, since interpolation kernels are bound to yield perfect reconstruction of the samples, we should expect the MSE to oscillate, vanishing at any sample location while reaching a maximum approximately midway between successive samples. In both extension scenarios, we will derive a decomposition of the mean squared error as a product of a modulation and of an envelope:  $\text{RMSE}_{[X \cdot \Delta_K * h_K]}(t)^2 = \frac{\sin^2(\pi t)}{\pi^2} A(t)$ . We will assess the behavior of the envelope as  $t$  varies. As expected, we will see that this behavior depends mostly on the distances from  $t$  to both borders of the signal domain,  $\pm(K + \frac{1}{2})$ .

**Definition 3.2 (border distances).** *Let  $|t| \leq K + \frac{1}{2}$ . We set*

$$\begin{aligned} \delta(t) &:= \min(K + \frac{1}{2} - t, K + \frac{1}{2} + t), \\ \Delta(t) &:= \max(K + \frac{1}{2} - t, K + \frac{1}{2} + t). \end{aligned}$$



**Figure 3.1.** Power spectrum decomposition for a given  $\alpha$  (the illustration actually refers to densities).

**3.3. A scheme to evaluate the power spectral MSE component.** In this part, we expose a general scheme to evaluate the power spectral MSE component. It can be applied to any reconstruction system  $h_K$  provided we can compute the power spectral MSE for signals of constant spectra and oversampled ones.

The starting point of this approach derives from a trivial statement. If one can find a spectrum  $d\tilde{\Psi}$  such that  $d\Psi \leq d\tilde{\Psi}$  (in the sense that  $d\tilde{\Psi} - d\Psi$  is a positive measure), then  $MSE_{d\Psi} \leq MSE_{d\tilde{\Psi}}$ . For instance, to come up with an upper bound, one has to find a  $d\tilde{\Psi}^+$  such that  $d\Psi \leq d\tilde{\Psi}^+$  and for which an upper bound can be computed. Similarly, for a lower bound, one shall find  $d\tilde{\Psi}^-$  such that  $d\tilde{\Psi}^- \leq d\Psi$  and derive a lower bound for  $d\tilde{\Psi}^-$ .

In this paper, we will consider signals of bounded spectrum. In that case,  $d\Psi_X(\omega) = \psi_X(\omega)d\omega$  satisfying  $\exists \psi^- \geq 0, \psi^+ \geq 0$  such that  $\forall \omega, \psi^- \leq \psi_X(\omega) \leq \psi^+$ . Then, using  $d\tilde{\Psi}^+ := \psi^+ \mathbb{1}_{|\omega| \leq \pi} d\omega$  and  $d\tilde{\Psi}^- := \psi^- \mathbb{1}_{|\omega| \leq \pi} d\omega$ , both being constant on  $[-\pi, \pi]$ , the spectral power MSE component bounds reduce to the same computation for a Nyquist band-limited white-noise. In fact, the bounds obtained in this way will not be sharp because  $\psi^+$  and  $\psi^-$  fall apart. Consequently, we will introduce finer bounds on the spectrum as illustrated in Figure 3.1.

**Proposition 3.3.** *Let  $0 \leq \alpha < 1$  and assume that  $\mathbb{1}_{|\omega| \geq \alpha\pi} d\Psi_X(\omega)$  is bounded. Denote by  $\psi_\alpha^+$  and  $\psi_\alpha^-$  the corresponding bounds and  $d\tilde{\Psi}_\alpha$  the positive component of the measure  $d\Psi_X(\omega) - \psi_\alpha^+ \mathbb{1}_{|\omega| \leq \pi} d\omega$ . Then,*

$$\begin{aligned} MSE_{d\Psi_X} &\leq MSE_{d\tilde{\Psi}_\alpha} + \psi_\alpha^+ MSE_{\mathbb{1}_{|\omega| \leq \pi} d\omega}, \\ MSE_{d\Psi_X} &\geq MSE_{d\tilde{\Psi}_\alpha} + \psi_\alpha^- MSE_{\mathbb{1}_{|\omega| \leq \pi} d\omega}. \end{aligned}$$

**Remark 3.1.** *With the previous proposition, we have reduced the bounds computation to two types of signals:*

- *Nyquist band-limited white-noise ( $d\Psi_X(\omega) = \mathbb{1}_{|\omega| \leq \pi} d\omega$ ), and*
- *over-sampled signals ( $\text{supp}(d\Psi_X(\omega)) \subset \{|\omega| \leq \alpha\pi\}$ ).*

All the derivations needed to achieve such a decomposition are valid in the case of multivariate band-limited RPs. In the next two sections, we focus on two reconstruction systems closely related to the Shannon-Whittaker formula. For each of them, we derive an accurate closed-form upper-bound that builds upon the previous decomposition. Given the technicality of the corresponding derivations, we consider univariate signals in a first stage and extend the results to multivariate RPs in a subsequent section.

**4. Truncated Shannon-Whittaker interpolation RMSE.** In this section, we consider the case of the truncated Shannon-Whittaker interpolation, where the interpolation kernel is sinc in the univariate case ( $d = 1$ ). It is interesting to point out that this method corresponds to the ideal Shannon-Whittaker reconstruction under the assumption that all unknown samples are null. This assumption does not imply that the signal vanishes outside the acquisition domain since this would contradict with Heisenberg's principle (in the context of band-limited signals). As such, it is hardly sustainable in practice, since it would be a striking coincidence that the signal vanishes exactly at the sample locations. The weakly stationary hypothesis that we have made is clearly much more satisfactory.

**4.1. Spectral representation.** Applying Theorem 3.1 with  $h_K = \text{sinc}$  yields the following corollary.

**Corollary 4.1.** *Let  $X_t$  be a strictly Nyquist band-limited random process of average  $\mu$ , then*

$$RMSE_{[X.\Delta_K*\text{sinc}]}(t)^2 = \frac{\sin^2(\pi t)}{\pi^2} \left( \begin{array}{c} \mu^2 \left| \sum_{|k|>K} \frac{(-1)^k}{t-k} \right|^2 \\ + \\ \frac{1}{2\pi} \int \left| \sum_{|k|>K} e^{i\omega k} \frac{(-1)^k}{t-k} \right|^2 d\Psi_X(\omega) \end{array} \right)$$

*Proof.* We have merely used that  $\sin(\pi(t - k)) = \sin(\pi t)(-1)^k$ . ■

**Definition 4.2 (Average and power spectral envelopes).** *In what follows, we set  $\forall |t| < K + \frac{1}{2}$*

$$A_\mu(t) := \mu^2 \left| \sum_{|k|>K} \frac{(-1)^k}{t-k} \right|^2,$$

$$A_{d\Psi}(t) := \frac{1}{2\pi} \int \left| \sum_{|k|>K} e^{i\omega k} \frac{(-1)^k}{t-k} \right|^2 d\Psi(\omega).$$

*Whenever these notions are involved, it will be implicitly assumed that  $|t| < K + \frac{1}{2}$ .*

**Proposition 4.3.** *Let  $X_t$  be a random process, then*

$$A_{d\Psi_X}(t) = \sum_{|k|>K, |m|>K} R_X(k-m) \frac{(-1)^{k+m}}{(t-k)(t-m)}$$

*Proof.*

$$\begin{aligned}
A_{d\Psi_X}(t) &= \frac{1}{2\pi} \int \left| \sum_{|k|>K} e^{i\omega k} \frac{(-1)^k}{t-k} \right|^2 d\Psi_X(\omega) \\
&= \frac{1}{2\pi} \int \sum_{|k|>K, |m|>K} e^{i\omega(k-m)} \frac{(-1)^{k+m}}{(t-k)(t-m)} d\Psi_X(\omega) \\
&= \sum_{|k|>K, |m|>K} \frac{(-1)^{k+m}}{(t-k)(t-m)} \frac{1}{2\pi} \int e^{i\omega(k-m)} d\Psi_X(\omega) \\
&= \sum_{|k|>K, |m|>K} R_X(k-m) \frac{(-1)^{k+m}}{(t-k)(t-m)}.
\end{aligned}$$

In the previous derivation, exchanging the sum and the integral is allowed thanks to the reconstruction systems dominated condition proven in Proposition 2.9. Indeed, this condition implies that:

$$\begin{aligned}
\left| \sum_{|k|\leq K, |m|\leq K} \frac{(-1)^{k+m}}{(t-k)(t-m)} \right| &= \frac{\pi^2}{\sin^2(\pi t)} \left| \sum_{|k|\leq K, |m|\leq K} \text{sinc}(t-k) \text{sinc}(t-m) e^{i\omega(k-m)} \right| \\
&= \frac{\pi^2}{\sin^2(\pi t)} \left| \sum_{|k|\leq K} \text{sinc}(t-k) e^{i\omega k} \right|^2 \leq \left( \frac{\pi M(t)}{\sin(\pi t)} \right)^2
\end{aligned}$$

■

We will now evaluate the behavior of each of the envelope terms.

**Proposition 4.4 (Average envelope component).** *Let  $X_t$  be a Nyquist band-limited random process of average value  $\mu$ , then*

$$A_\mu(t) \leq \mu^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2,$$

where  $\delta(t)$  and  $\Delta(t)$  are the distances defined in Definition 3.2.

*Proof.* This derives easily from the fact that  $A_\mu(t)$  is expressed in Definition 4.2 via an alternating series when  $|t| < K + 1$ . ■

## 4.2. Band-limited white-noise RMSE.

**Definition 4.5 (Nyquist band-limited white noise).**

A random process  $W_t$  is a Nyquist band-limited white noise of variance  $\sigma^2$  if  $d\Psi_W(\omega) = \sigma^2 \mathbb{1}_{|\omega|\leq\pi} d\omega$ .

**Theorem 4.6.** *Let  $W_t$  be a Nyquist band-limited white-noise of unit variance. Then,  $\forall |t| <$*



$K + \frac{1}{2}$ ,

$$\begin{aligned} A_{d\Psi_W}(t) &\leq \frac{1}{\delta(t)} + \frac{1}{\Delta(t)}, \\ A_{d\Psi_W}(t) &\geq \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}}. \end{aligned}$$

*Proof.* Using Proposition 4.3,

$$A_{d\Psi}(t) = \sum_{|k|>K, |m|>K} R_W(k-m) \frac{(-1)^{k+m}}{(t-k)(t-m)}.$$

Then, given that  $W_t$  is a Nyquist band-limited white-noise,  $R_W(t) = \text{sinc}(t)$ . Therefore,

$$A_{d\Psi}(t) = \sum_{|k|>K} \frac{1}{(t-k)^2}.$$

Then a monotonicity argument ensures that  $\frac{1}{(t-k)^2} \geq \int_k^{k+1} \frac{1}{(t-y)^2} dy$ , which implies that

$$\begin{aligned} A_{d\Psi}(t) &\geq \int_{|y|\geq K+1}^{\infty} \frac{1}{(t-y)^2} dy \\ &= \frac{1}{K+1-t} + \frac{1}{K+1+t} \\ &= \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}}. \end{aligned}$$

A similar monotonicity argument would not suffice to derive the desired upper bound. Instead we shall use the convexity of  $g_t : y \mapsto \frac{1}{(t-y)^2}$  on intervals of the form  $]k - \frac{1}{2}, k + \frac{1}{2}[$  whenever  $\max(|k - \frac{1}{2}|, |k + \frac{1}{2}|) > t$ . Then Jensen's inequality applied to a uniform random variable on  $[k - \frac{1}{2}, k + \frac{1}{2}]$  implies

$$\frac{1}{(t-k)^2} = g_t\left(\frac{k + \frac{1}{2} + k - \frac{1}{2}}{2}\right) \leq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} g_t(y) dy = \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{(t-y)^2} dy.$$

Therefore, we obtain

$$\begin{aligned} A_{d\Psi}(t) &\leq \int_{|y|\geq K+\frac{1}{2}}^{\infty} \frac{1}{(t-y)^2} dy \\ &= \frac{1}{K+\frac{1}{2}-t} + \frac{1}{K+\frac{1}{2}+t} \\ &= \frac{1}{\delta(t)} + \frac{1}{\Delta(t)}. \end{aligned}$$

■

**4.3. Oversampled signal RMSE.** In this part we will consider signals oversampled with a rate  $\alpha < 1$ . To derive an upper bound we will need the following technical lemma.

**Lemma 4.7.**  $\forall |\omega| < \pi$  and  $\forall |t| < K + 1$ ,

$$\left| \sum_{|k| > K} e^{i\omega k} \frac{(-1)^k}{t - k} \right| \leq \frac{2}{|1 + e^{i\omega}|} \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right).$$

*Proof.* In what follows, we will use a summation by parts (a.k.a. Abel transformation). Given two sequences,  $a_k, b_k, k \geq 0$ , we can denote  $S(b)_K^k = \sum_{K < k' \leq k} b_{k'}$ . Then we have,

$$\sum_{K < k \leq K'} a_k b_k = a_{K'} S(b)_K^{K'} - \sum_{K < k < K'} (a_{k+1} - a_k) S(b)_K^k.$$

Using this formula with  $a_k := \frac{1}{t-k}$  and  $b_k = (-e^{i\omega})^k$ , one obtains for  $|t| < K + 1$

$$\begin{aligned} \left| \sum_{K < k \leq K'} e^{i\omega k} \frac{(-1)^k}{t - k} \right| &\leq |a_{K'} S(b)_K^{K'}| + \sum_{K < k < K'} |a_{k+1} - a_k| |S(b)_K^k| \\ &= \frac{1}{K' - t} |S(b)_K^{K'}| + \sum_{K < k < K'} \left( \frac{1}{k - t} - \frac{1}{k + 1 - t} \right) |S(b)_K^k|. \end{aligned}$$

Noting that

$$\begin{aligned} |S(b)_K^k| &= \left| (-e^{i\omega})^{K+1} \frac{1 - (-e^{i\omega})^{k-K}}{1 + e^{i\omega}} \right| \\ &\leq \frac{2}{|1 + e^{i\omega}|} \end{aligned}$$

and letting  $K' \rightarrow \infty$ , we obtain

$$\begin{aligned} \left| \sum_{k > K} e^{i\omega k} \frac{(-1)^k}{t - k} \right| &\leq \frac{2}{|1 + e^{i\omega}|} \sum_{k > K} \left( \frac{1}{k - t} - \frac{1}{k + 1 - t} \right) \\ &= \frac{2}{|1 + e^{i\omega}|} \frac{1}{K + 1 - t}. \end{aligned}$$

By symmetry,

$$\left| \sum_{k < -K} e^{i\omega k} \frac{(-1)^k}{t - k} \right| \leq \frac{2}{|1 + e^{-i\omega}|} \frac{1}{K + 1 + t} = \frac{2}{|1 + e^{i\omega}|} \frac{1}{K + 1 + t}.$$

■

**Proposition 4.8.** Let  $\alpha < 1$  and  $d\Psi$  a spectrum supported on  $\{|\omega| \leq \alpha\pi\}$ , then

$$A_{d\Psi}(t) \leq \sigma_\alpha^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2,$$

where

$$\sigma_\alpha^2 := \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{4}{|1 + e^{i\omega}|^2} d\Psi(\omega).$$

*Proof.* This is a direct consequence of Definition 4.2 and of the previous lemma. ■

**Remark 4.1.** Since  $\frac{1}{|1+e^{i\omega}|^2} \underset{\omega \rightarrow \pi}{\sim} \frac{1}{(\pi-\omega)^2}$  the condition on the support could be replaced by one with respect to the total variation of  $\frac{1}{(\omega-\pi)^2} d\Psi_X(\omega)$  near  $\pi$ . A similar variation was proposed in [5]. None of these refinements can tackle the case of band-limited white-noise.

**4.4. General case.** The next theorem summarizes some of the previous results. For the white-noise component, only the upper-bound is alluded to. For the record, the lower-bound served to demonstrate the sharpness of the upper-bound.

**Theorem 4.9.** Let  $X_t$  be a random process and  $0 \leq \alpha < 1$  such that  $\mathbf{1}_{|\omega| \geq \alpha\pi} d\Psi_X(\omega)$  is bounded. We have shown a decomposition of the mean squared error  $RMSE_{[X, \Delta_K * \text{sinc}]}(t)^2$  into the following terms (in order of appearance in the formula below):

- a **modulation**,
- an **average envelope component**,
- a **low-frequency envelope component** and
- a **white-noise envelope component**.

This decomposition leads to the following estimate valid  $\forall |t| < K + \frac{1}{2}$

$$RMSE_{[X, \Delta_K * \text{sinc}]}(t)^2 \leq \frac{\sin^2(\pi t)}{\pi^2} \times \left( \begin{array}{c} \mu_X^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2 \\ + \\ \sigma_\alpha^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2 \\ + \\ \sigma'_\alpha{}^2 \left( \frac{1}{\delta(t)} + \frac{1}{\Delta(t)} \right) \end{array} \right),$$

where

$$\sigma'_\alpha{}^2 := \|\mathbf{1}_{|\omega| \geq \alpha\pi} d\Psi_X(\omega)\|_\infty$$

and

$$\sigma_\alpha^2 := \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{4}{|1 + e^{i\omega}|^2} d\tilde{\Psi}_\alpha(\omega) = \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{2}{1 + \cos(\omega)} d\tilde{\Psi}_\alpha(\omega).$$

*Proof.* Nothing remains to be proved since the theorem only summarizes previous developments (namely Proposition 3.3, Corollary 4.1, Definition 4.2, Theorem 4.6 and Proposition 4.8). Note that  $d\tilde{\Psi}_\alpha$  was defined in Proposition 3.3, where  $\sigma'_\alpha{}^2$  was referred to as  $\psi_\alpha^+$ . ■

**5. Periodic Shannon-Whittaker interpolation RMSE.** In the previous section, we have looked upon the interpolation method most classically considered in the context of truncation error. However, this method is hardly used in practice mainly because its implementation is very inefficient. In the class of linear shift invariant methods, the ones based on rapidly decaying kernels (such as B-splines) are therefore strongly preferred. Here we consider yet another approach often referred as DFT zero-padding interpolation. While the underlying kernel displays as slow a decay as sinc, efficient implementations are often possible thanks to the Fast Fourier Transform algorithm. Here again we only consider the univariate case ( $d = 1$ ).

**5.1. Periodic extension and DFT interpolation.** We first introduce the natural reconstruction kernel for band-limited periodic signals of length  $N$ . Instead of relying on an explicit formula, we prefer the following definition as a periodized version of the “band-limited Dirac distribution” (*i.e.* of the cardinal sine) since it makes its purpose clearer.

**Definition 5.1 (Discrete cardinal sine).** Let  $N = 2K + 1$ . We define the discrete cardinal sine function of rank  $K$  as

$$\text{sincd}_K(t) := \sum_{q \in \mathbb{Z}} \text{sinc}(t - qN).$$

Based on this definition, it is trivial to show that the truncated interpolation based on the discrete cardinal sine is in fact equivalent to the infinite Shannon-Whittaker expansion, if the signal is periodic. But let us first provide some useful notations.

**Definition 5.2 (Index periodization).** For  $k \in \mathbb{Z}$ , let us define the unique couple of indices  $q_N(k)$  and  $k \% N$  such that  $|k \% N| \leq K$  and

$$k = q_N(k)N + k \% N.$$

**Proposition 5.3 (Periodic extension).** Let  $X_t$  be a random process, then

$$X.\Delta_K * \text{sincd}_K(t) = \sum_{k \in \mathbb{Z}} X_{k \% N} \text{sinc}(t - k).$$

*Proof.*

$$\begin{aligned} X.\Delta_K * \text{sincd}_K(t) &= \sum_{|k| \leq K} X_k \text{sincd}_K(t - k) \\ &= \sum_{|k| \leq K} X_k \sum_{q \in \mathbb{Z}} \text{sinc}(t - k - qN) \end{aligned}$$

Using the substitution  $k' = k + qN$ , we indeed obtain the appropriate formula

$$X.\Delta_K * \text{sincd}_K(t) = \sum_{k'} X_{k' \% N} \text{sinc}(t - k')$$

In the next proposition, we provide two alternate expressions of the discrete cardinal sine.

**Proposition 5.4.** *For all  $t \in \mathbb{R}$ , we have*

1.

$$(\text{sincd}_K(t - k))_{|k| \leq K} = \mathcal{DFT}^{-1} \left( e^{-2i\pi t \frac{1}{N}} \right)_{|l| \leq K} \quad (5.1)$$

2.

$$\text{sincd}_K(t) = \frac{\sin(\pi t)}{N \sin(\frac{\pi}{N} t)}.$$

*Proof.*

1. This point is somehow another formulation of the definition by taking into consideration the link between the Fourier transform and the DFT. More precisely, since  $\text{sinc}(t - \tau) = \mathcal{F}^{-1}(e^{-it\omega} \mathbf{1}_{\{|\omega| < \pi\}})(\tau)$  its periodized version with period  $N$  verifies

$$(\text{sincd}_K(t - k))_{|k| \leq K} = \mathcal{DFT}^{-1}(e^{-it2\pi \frac{1}{N}} \mathbf{1}_{\{|2\pi \frac{1}{N}| < \pi\}})_{|l| \leq K} = \mathcal{DFT}^{-1}(e^{-it2\pi \frac{1}{N}})_{|l| \leq K}$$

2. The second point derives from the first and the following computation

$$\begin{aligned} \text{sincd}_K(t) &= \mathcal{DFT}^{-1} \left( e^{-2i\pi t \frac{1}{N}} \right)_{|l| \leq K} [0] = \frac{1}{N} \sum_{|l| \leq K} e^{-2i\pi t \frac{1}{N}} \\ &= \frac{1}{N} e^{2i\pi t \frac{K}{N}} \frac{1 - e^{-2i\pi t \frac{N}{N}}}{1 - e^{-2i\pi t \frac{1}{N}}} = \frac{1}{N} \frac{\sin(\pi t)}{\sin(\pi \frac{t}{N})}. \end{aligned}$$

The previous proposition gives us the opportunity to make the link with the standard DFT interpolation.

**Corollary 5.5 (DFT interpolation).**  $\forall t$ ,

$$X.\Delta_K * \text{sincd}_K(t) = \frac{1}{N} \sum_{|l| \leq K} \mathcal{DFT}((X_k)_{|k| \leq K}) [l] e^{2i\pi t \frac{1}{N}}.$$

*Proof.* According to the previous proposition,

$$\begin{aligned} X.\Delta_K * \text{sincd}_K(t) &= X.\Delta_K * \mathcal{DFT}^{-1}(e^{2i\pi t \frac{1}{N}}) \\ &= \sum_{|k| \leq K} X_k \mathcal{DFT}^{-1}(e^{2i\pi t \frac{1}{N}})[k]. \end{aligned}$$

Then using Parseval's identity, we get

$$\begin{aligned} X.\Delta_K * \text{sincd}_K(t) &= \frac{1}{N} \sum_{|l| \leq K} \mathcal{DFT}(X_k)[l] \mathcal{DFT}(\mathcal{DFT}^{-1}(e^{2i\pi t \frac{1}{N}}))[l] \\ &= \frac{1}{N} \sum_{|l| \leq K} \mathcal{DFT}(X_k)[l] e^{2i\pi t \frac{1}{N}}. \end{aligned}$$

**5.2. Spectral representation.** Applying Theorem 3.1 with  $h_K(t) = \text{sinc}_K(t)$  yields the following corollary.

**Corollary 5.6.** *Let  $X_t$  be a Nyquist band-limited random process of average value  $\mu$ , then*

$$RMSE_{[X.\Delta_K*\text{sinc}_K]}(t)^2 = \frac{\sin^2(\pi t)}{\pi^2} \frac{1}{2\pi} \int \left| \sum_{|k|>K} (e^{i\omega k} - e^{i\omega k \% N}) \frac{(-1)^k}{t-k} \right|^2 d\Psi(\omega).$$

*Proof.* First, we need to show that the average component is null. As a reminder,

$$MSE_\mu(t)(\text{sinc}_K) = \mu^2 \left| 1 - \sum_{|k|\leq K} \text{sinc}_K(t-k) \right|^2.$$

But according to Equation (5.1),

$$\begin{aligned} \sum_{|k|\leq K} \text{sinc}_K(t-k) &= \mathcal{DFT} \left( (\text{sinc}_K(t-k))_{|k|\leq K} \right) [l=0] \\ &= \mathcal{DFT} \left( \mathcal{DFT}^{-1} \left( (e^{2i\pi t \frac{l}{N}})_{|l|\leq K} \right) \right) [l=0] \\ &= e^{2i\pi t \frac{0}{N}} = 1. \end{aligned}$$

Other than that, Theorem 3.1 and then Proposition 5.3 imply that

$$\begin{aligned} MSE_{d\Psi}(\text{sinc}_K) &= \frac{1}{2\pi} \int \left| \sum_{k\in\mathbb{Z}} e^{i\omega k} \text{sinc}(t-k) - \sum_{|k|\leq K} e^{i\omega k} \text{sinc}_K(t-k) \right|^2 d\Psi(\omega) \\ &= \frac{1}{2\pi} \int \left| \sum_{|k|>K} (e^{i\omega k} - e^{i\omega k \% N}) \text{sinc}(t-k) \right|^2 d\Psi(\omega) \end{aligned}$$

and once again we can use that  $\sin(\pi(t-k)) = \sin(\pi t)(-1)^k$ . ■

**Definition 5.7 (Power spectral envelope).** *In what follows, we set*

$$B_{d\Psi}(t) := \frac{1}{2\pi} \int \left| \sum_{|k|>K} (e^{i\omega k} - e^{i\omega k \% N}) \frac{(-1)^k}{t-k} \right|^2 d\Psi(\omega).$$

*We will refer to it as the power spectral envelope component.*

### 5.3. Band-limited white-noise RMSE.

**5.3.1. Global bounds.** **Lemma 5.8.** *Let  $W_t$  be a Nyquist band-limited white-noise of unit variance. Then, following the notations as in Definitions 4.2 and 5.7*

$$\begin{aligned} A_{d\Psi_W}(t) &= \sum_{|k|\leq K} \frac{\pi^2}{N^2 \sin^2(\frac{\pi}{N}(t-k))} \left( 1 - \text{sinc}^2(\frac{1}{N}(t-k)) \right), \\ B_{d\Psi_W}(t) &= \sum_{|k|\leq K} \frac{2\pi^2}{N^2 \sin^2(\frac{\pi}{N}(t-k))} \left( 1 - \text{sinc}(\frac{1}{N}(t-k)) \right). \end{aligned}$$



*Proof.* See the supplementary file. ■

**Proposition 5.9.** *Let  $W_t$  be a Nyquist band-limited white-noise of unit variance. Then*

$$A_{d\Psi_W}(t) \leq B_{d\Psi_W}(t) \leq 2A_{d\Psi_W}(t).$$

*Proof.* The previous lemma can be reformulated as

$$A_{d\Psi_W} = \sum_{|k| \leq K} a_k \text{ and } B_{d\Psi_W} = \sum_{|k| \leq K} b_k,$$

with  $a_k, b_k \geq 0$  and

$$\frac{b_k}{a_k} = 2 \frac{1 - \operatorname{sinc}(\frac{1}{N}(t-k))}{1 - \operatorname{sinc}^2(\frac{1}{N}(t-k))} = \frac{2}{1 + \operatorname{sinc}(\frac{1}{N}(t-k))} := r\left(\frac{1}{N}(t-k)\right).$$

A simple analysis shows that the function  $r$  is even and increasing on  $[0, 1]$ . As a result,  $\forall |u| \leq 1, r(u) \in [r(0), r(1)] = [1, 2]$ . ■

**Corollary 5.10.** *Let  $W_t$  be a Nyquist band-limited unit white-noise. Then,  $\forall |t| < K + \frac{1}{2}$*

$$\frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \leq B_{d\Psi_W}(t) \leq \frac{2}{\delta(t)} + \frac{2}{\Delta(t)}.$$

*Proof.* These inequalities derive from the previous proposition and from Theorem 4.6. ■

The comparison of the lower and upper bounds reveals that these two estimates of  $B_{d\Psi_W}$  are much less tight than the corresponding bounds on  $A_{d\Psi_W}$ . In the next two sections, we will improve this situation by providing accurate asymptotic estimates of  $B_{d\Psi_W}$ . They should in particular allow us to determine which of the upper and lower bound is tighter when considering limit cases (near the border and at the center of the signal domain).

**5.3.2. Exact equivalent inside the image domain.** In this part and the following one, we derive more accurate estimates of the error based on two complementary asymptotic analyses when  $K$  becomes large. To account for this limiting case, we allow the interpolation location  $t$  to vary with  $K$  which in practice boils down to consider sequences  $t_K$ . Here we tackle the case where  $t_K$  cannot get closer to the image border than by a margin linear in  $K$ .

**Proposition 5.11.** *Let  $W_t$  be a Nyquist band-limited white-noise of unit variance. Then,  $\forall |t_K| \leq \beta K$  with  $\beta < 1$ ,*

$$B_{d\Psi_W}(t_K) \underset{N \rightarrow \infty}{\sim} \frac{2\pi^2}{N} \int_{-\frac{1}{2} - \frac{t_K}{N}}^{\frac{1}{2} - \frac{t_K}{N}} \frac{1}{\sin^2 \pi(u)} (1 - \operatorname{sinc}(u)) du.$$

*Proof.* See the supplementary file. ■

**Remark 5.1.** When  $t_K = o(K)$ , we obtain

$$\begin{aligned} B_{d\Psi_W}(t_K)_{N \rightarrow \infty} &\sim \frac{2\pi^2}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\sin^2(\pi u)} (1 - \text{sinc}(u)) du \\ &\simeq 1.08 \frac{2}{K + \frac{1}{2}} \simeq 1.08 \left( \frac{1}{\delta(t_K)} + \frac{1}{\Delta(t_K)} \right). \end{aligned}$$

Therefore, the global lower bound computed previously is almost tight.

**5.3.3. Tight asymptotic bounds near the image border.** The previous equivalent does not remain true when loosing the constraint  $|t_K| \leq \beta K$ , that is when  $t_K$  approaches the domain border. However, we have a slightly weaker result in this case.

**Proposition 5.12.** Let  $W_t$  be a Nyquist band-limited white-noise of unit variance. Then,  $\forall |t_K| \leq K$  with  $\frac{\delta(t_K)}{K} \rightarrow 0^3$ ,

$$\frac{2}{\delta(t_K) + \frac{1}{2}} + \mathcal{O}\left(\frac{\log(N)}{N}\right) \leq B_{d\Psi_W}(t_K) \leq \frac{2}{\delta(t_K)} + \mathcal{O}\left(\frac{1}{N}\right),$$

where the Landau notation  $a_N = \mathcal{O}(b_N)$  means that  $\exists M > 0, \forall N, |a_N| \leq Mb_N$ .

*Proof.* See supplementary file. ■

**Remark 5.2.** Note that under the theorem's assumptions,  $\frac{1}{\Delta(t_K)} = \frac{1}{N - \delta(t_K)} = \mathcal{O}\left(\frac{1}{N}\right)$  and then this time, it is the global upper bound that is tight.

**5.4. Oversampled signal RMSE.** **Lemma 5.13.**  $\forall |\omega| < \pi$  and  $\forall |t| < K + 1/2$ ,

$$\left| \sum_{|k| > K} (e^{i\omega k} - e^{i\omega k \% N}) \frac{(-1)^k}{t - k} \right| \leq \frac{4}{|1 + e^{i\omega}|} \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right).$$

*Proof.* See the supplementary file. ■

**Proposition 5.14.** Let  $\alpha < 1$  and  $d\Psi$  a spectrum supported on  $\{|\omega| \leq \alpha\pi\}$ , then

$$B_{d\Psi}(t) \leq 4\sigma_\alpha^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2,$$

where  $B_{d\Psi}(t)$  was given in Definition 5.7 and

$$\sigma_\alpha^2 := \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{4}{|1 + e^{i\omega}|^2} d\Psi(\omega).$$

*Proof.* This is a direct consequence of the previous lemma. ■

---

<sup>3</sup>This assumption ensures that the terms in the big  $\mathcal{O}$  are negligible compared to the remaining ones.

**5.5. General case.** The next theorem summarizes some of the previous results. For the white-noise component, it only considers the global upper-bound and does not include the asymptotic analyses that were presented earlier.

**Theorem 5.15.** *Let  $X_t$  be a random process and  $0 \leq \alpha < 1$  such that  $\mathbb{1}_{|\omega| \geq \alpha\pi} d\Psi_X(\omega)$  is bounded. We have shown a decomposition of the mean squared error  $RMSE_{[X, \Delta_K * \text{sincd}_K]}(t)^2$  similar to the one of the truncated Shannon-Whittaker interpolation, that is a decomposition into:*

- a **modulation**,
- an **average envelope component**,
- a **low-frequency envelope component** and
- a **white-noise envelope component**.

$$RMSE_{[X, \Delta_K * \text{sincd}_K]}(t)^2 \leq \frac{\sin^2(\pi t)}{\pi^2} \times \left( \begin{array}{c} 0 \\ + \\ 4\sigma_\alpha^2 \left( \frac{1}{\delta(t) + \frac{1}{2}} + \frac{1}{\Delta(t) + \frac{1}{2}} \right)^2 \\ + \\ 2\sigma'_\alpha{}^2 \left( \frac{1}{\delta(t)} + \frac{1}{\Delta(t)} \right) \end{array} \right),$$

where

$$\sigma'_\alpha{}^2 := \|\mathbb{1}_{|\omega| \geq \alpha\pi} d\Psi_X(\omega)\|_\infty$$

and

$$\sigma_\alpha^2 := \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{4}{|1 + e^{i\omega}|^2} d\tilde{\Psi}_\alpha(\omega) = \frac{1}{2\pi} \int_{|\omega| \leq \alpha\pi} \frac{2}{1 + \cos(\omega)} d\tilde{\Psi}_\alpha(\omega).$$

*Proof.* Here again nothing remains to be proved since the theorem only summarizes previous developments (namely Proposition 3.3, Corollary 5.6, Definition 5.7, Corollary 5.10 and Proposition 5.14). Note that  $d\tilde{\Psi}_\alpha$  was defined in Proposition 3.3, where  $\sigma'_\alpha{}^2$  was referred to as  $\psi_\alpha^+$ . ■

**Remark 5.3.** *The main differences with the truncated Shannon-Whittaker case are that*

- *the average envelope vanishes*
- *multiplicative factors are larger (at most twice as large if speaking of RMSE).*

**6. Multivariate extension.** In section 2, all the derivations were made in the case of multivariate RPs and for a generic reconstruction system  $h_K$ . These derivations have led to a general form of the truncation RMSE (see Theorem 2.12), which is at the basis of the upper bounds derived afterwards. Because of their high degree of technicality the upper-bounds could only be obtained for two specific reconstruction systems and for univariate RPs. The latter limitation is all the more unfortunate since images are two-dimensional. In this section, we partly extend our results in the multivariate case.

**6.1. General behavior of the multivariate MSE.** Considering the common case of separable reconstruction systems, then the reconstruction is itself separable in the sense that  $\hat{X}_t^{[h_K, K]} := \sum_{|k| \leq K} X_k \prod_{j=1}^d h_{K_j}(t_j - k_j)$  can be obtained by applying successive univariate

reconstructions. Indeed,

$$\tilde{X}_t^{[h_K, K]} := \sum_{|k_1| \leq K_1} \cdots \left( \sum_{|k_d| \leq K_d} X_k h_{K_d}(t_d - k_d) \right) \cdots h_{K_1}(t_1 - k_1)$$

Based on this observation, it is tempting to deduce that the errors entailed by each univariate stage accumulate, so that the multivariate error would be the sum of individual univariate errors. This argument is fallacious because the statistical properties of the original RP are not preserved during any of the univariate reconstruction stages.

**Remark 6.1.** *Let us restate the spectral representation of Theorem 2.12:*

$$\begin{aligned} MSE_\mu(t)(h_K) &:= \mu^2 \left| 1 - \prod_{j=1}^d \Delta_{K_j} * h_{K_j}(t_j) \right|^2, \\ MSE_{d\Psi}(t)(h_K) &:= \frac{1}{(2\pi)^d} \int \left| 1 - \prod_{j=1}^d \left( e^{i\omega_j(\cdot - t_j)} \Delta_{K_j} \right) * h_{K_j}(t_j) \right|^2 d\Psi(\omega). \end{aligned}$$

We have used a slightly different (but equivalent) formula for the spectral component in order to highlight a common property of both components, namely the fact that they involve terms of the form

$$\left| 1 - \prod_{j=1}^d a_j \right|.$$

In sections 4 and 5, we have derived univariate upper-bounds of the form  $|1 - a_j| \leq u_j$ . We now turn them into multivariate bounds by using the following algebraic identity

$$1 - \prod_{j=1}^d a_j = \sum_{j=1}^d (1 - a_j) \prod_{m < j} a_m.$$

Therefore,

$$\left| 1 - \prod_{j=1}^d a_j \right| \leq \sum_{j=1}^d |1 - a_j| \prod_{m < j} |a_m| \leq \sum_{j=1}^d u_j \prod_{m < j} (1 + u_m). \quad (6.1)$$

Applying Equation 6.1 in the context of a 2D image leads to the following result.

**Proposition 6.1.** *Let  $h_K$  be either sinc or sincd $_K$  and let  $X_{t_1, t_2}$  a band-limited RP. Then*

$$\begin{aligned} RMSE_{[X, \Delta_K * h_K]}(t_1, t_2)^2 &= \frac{\sin^2(\pi t_1)}{\pi^2} \mathcal{O} \left( \frac{1}{\delta_1(t)^2} + \frac{1}{\delta_1(t)} \right) + \frac{\sin^2(\pi t_2)}{\pi^2} \mathcal{O} \left( \frac{1}{\delta_2(t)^2} + \frac{1}{\delta_2(t)} \right) + \\ &\quad \frac{\sin^2(\pi t_1) \sin^2(\pi t_2)}{\pi^2} \mathcal{O} \left( \frac{1}{\delta_1(t)^2} + \frac{1}{\delta_1(t)} \right) \mathcal{O} \left( \frac{1}{\delta_2(t)^2} + \frac{1}{\delta_2(t)} \right) \end{aligned}$$

where  $\forall j \in \{1, 2\}$

$$\begin{aligned}\delta_j(t) &:= \min(K_j + \frac{1}{2} - t_j, K_j + \frac{1}{2} + t_j), \\ \Delta_j(t) &:= \max(K_j + \frac{1}{2} - t_j, K_j + \frac{1}{2} + t_j).\end{aligned}$$

*Proof.* This is a consequence of Equation 6.1 and either Theorem 4.9 (for sinc) or Theorem 5.15 (for sincd). ■

**Remark 6.2.** *The leading constants were intentionally left out of the upper-bound because they no longer have simple expressions. Neither do they remain sharp at all. Indeed the inequality  $|a_i| \leq (1 + u_i)$  is quite loose.*

*In any case, we are interested here by the general behavior of the multi-variate MSE. The upper-bound shows that far from the border, the behavior is essentially the same as in 1D, or more precisely as the sum of two 1D interpolation errors. Near the border however, the error might behave like the product of two 1D interpolation errors.*

**6.2. White-noise.** Since the white-noise contribution of the MSE was clearly identified as the main one far enough from the border of the image, we develop a more accurate upper-bound for such signals. In this case, the procedure of Remark 6.1 can be refined and yields a much simpler and sharper result. Indeed, we will show that the upper bound for the MSE is truly additive in that case. We will rely on the following expression of the MSE.

**Proposition 6.2.** *Let  $(h_K)$  a reconstruction system, and let  $W_t$  a band-limited white noise of unit variance. Then,*

$$MSE_{d\Psi_W}(t)(h_K) = 1 - \prod_{j=1}^d \sum_{|k_j| \leq K_j} h_{K_j}(t_j - k_j) (2\text{sinc}(t_j - k_j) - h_{K_j}(t_j - k_j))$$

*Proof.* To simplify the notations, we set

$$a_j(\omega_j, t_j) := \left( e^{i\omega_j(\cdot - t_j)} \Delta_{K_j} \right) * h_{K_j}(t_j) = \sum_{|k_j| \leq K_j} h_{K_j}(t_j - k_j) e^{i\omega_j(t_j - k_j)}.$$

With such notations,

$$\begin{aligned}MSE_{d\Psi_W}(t)(h_K) &= \frac{1}{(2\pi)^d} \int \left| 1 - \prod_{j=1}^d \left( e^{i\omega_j(\cdot - t_j)} \Delta_{K_j} \right) * h_{K_j}(t_j) \right|^2 d\Psi_W(\omega) \\ &= \frac{1}{(2\pi)^d} \int \left| 1 - \prod_{j=1}^d a_j(\omega_j, t_j) \right|^2 d\Psi_W(\omega) \\ &= \frac{1}{(2\pi)^d} \int 1 - 2 \prod_{j=1}^d a_j(\omega_j, t_j) + \prod_{j=1}^d a_j(\omega_j, t_j)^2 d\Psi_W(\omega) \\ &= R_W(0) - \prod_{j=1}^d \int_{-\pi}^{\pi} 2a_j(\omega_j, t_j) - a_j(\omega_j, t_j)^2 d\omega_j\end{aligned}$$

Then we can use that  $R_W(0) = 1$ ,  $\int_{-\pi}^{\pi} a_j(\omega_j, t_j) d\omega_j = \sum_{|k_j| \leq K_j} h_{K_j}(t_j - k_j) \text{sinc}(t_j - k_j)$  and

$$\begin{aligned} \int_{-\pi}^{\pi} a_j(\omega_j, t_j)^2 d\omega_j &= \sum_{|k_j| \leq K_j, |m_j| \leq K_j} h_{K_j}(t_j - k_j) h_{K_j}(t_j - m_j) \text{sinc}(m_j - k_j) \\ &= \sum_{|k_j| \leq K_j} h_{K_j}(t_j - k_j)^2. \end{aligned}$$

■

In order to apply similar derivations as in Remark 6.1 in a finer way, we need a tight upper-bound for the terms involved in the product.

**Lemma 6.3.** *Let  $(h_K)$  a reconstruction system subject to  $\forall K, \forall |t| \leq K + \frac{1}{2}$ ,*

$$\sum_{|k| \leq K} h_K(k - t) \text{sinc}(t - k) \geq 0$$

and

$$\sum_{|k| \leq K} |h_K(k - t)|^2 \leq 1$$

Then,

$$\left| \sum_{|k| \leq K} h_K(t - k) (2 \text{sinc}(t - k) - h_K(t - k)) \right| \leq 1.$$

*Proof.* This is a mere application of the Cauchy-Schwarz inequality. ■

It is easy to verify the assumptions of the previous lemma for  $\text{sincd}_K$  and for  $\text{sinc}$  and obtain the following result.

**Theorem 6.4.** *Let  $h_K$  be either  $\text{sinc}$  or  $\text{sincd}_K$ , and let  $W_t$  a band-limited white noise of unit variance. Then,*

$$\text{MSE}_{d\Psi_W}(t)(h_K) \leq C \sum_{j=1}^d \frac{\sin^2(\pi t_j)}{\pi^2} \left( \frac{1}{\delta_j(t)} + \frac{1}{\Delta_j(t)} \right)$$

where  $C = 1$  for  $\text{sinc}$  or  $C = 2$  for  $\text{sincd}_K$  and  $\forall j$

$$\begin{aligned} \delta_j(t) &:= \min\left(K_j + \frac{1}{2} - t_j, K_j + \frac{1}{2} + t_j\right), \\ \Delta_j(t) &:= \max\left(K_j + \frac{1}{2} - t_j, K_j + \frac{1}{2} + t_j\right). \end{aligned}$$



*Proof.* Let  $b_j := \sum_{|k| \leq K} h_K(t-k)(2\text{sinc}(t-k) - h_K(t-k))$ . We know that  $|b_j| \leq 1$ . Then

$$\begin{aligned} MSE_{d\Psi_w}(t)(h_K) &= 1 - \prod_{j=1}^d b_j \\ &= \sum_{j=1}^d (1 - b_j) \prod_{m < j} b_m \\ &\leq \sum_{j=1}^d |1 - b_j| \prod_{m < j} |b_m| \\ &\leq \sum_{j=1}^d |1 - b_j| \end{aligned}$$

Since  $1 - b_j$  corresponds to the MSE in the univariate case, we can use the results in Theorem 4.6 and Corollary 5.10 to conclude the proof. ■

**7. Experiments.** In this section, we will validate our theoretical bounds by comparing them to an empirical estimator of the interpolation RMSE. We will first explain quickly the way we derive this estimator (see Section 7.1). Since our theoretical bounds rely on the image spectrum statistics, we also need a practical way to evaluate them; this point is discussed in Section 7.2. Last, we consider several test images and provide comparisons between the bounds and the estimator in the two interpolation cases that we have tackled in this article.

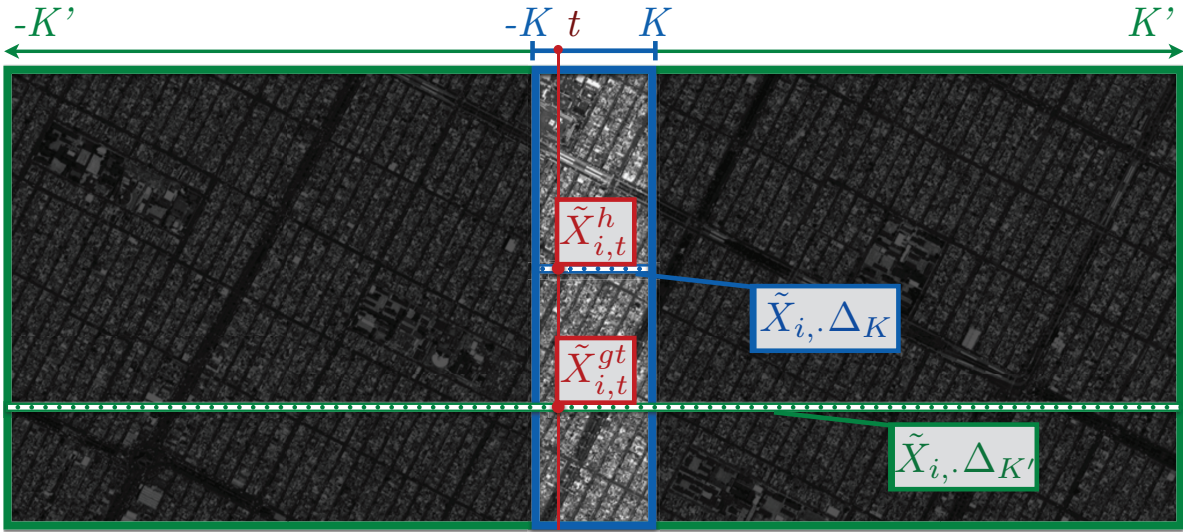
**7.1. Empirical estimator.** The core difficulty to derive an accurate RMSE estimator resides in that we need to know the ground-truth interpolated signal. The only way to truly obtain such a thing is to generate synthetic Nyquist band-limited data analytically and then perform sampling. We have chosen a different path because synthetic signals are not representative of real images.

Instead, starting from an image of large size, we can perform an interpolation using the entire image and use it as a reference value to evaluate the accuracy of an interpolated image obtained by only using a part of the image (see Figure 7.1). We will refer somehow improperly to the reference as the “ground-truth” interpolated signal. When computing the reference, we will always use the truncated Shannon-Whittaker scheme. Let us denote  $\tilde{X}_t^{gt} := X \cdot \Delta_{K'} * \text{sinc}(t)$  the “ground-truth” interpolate and  $\tilde{X}_t^h := X \cdot \Delta_K * h_K(t)$  the evaluated interpolate with  $K \ll K'$ . We want to empirically estimate

$$MSE[\tilde{X}^{gt} - \tilde{X}^h](t) := \mathbb{E} \left[ (\tilde{X}_t^{gt} - \tilde{X}_t^h)^2 \right].$$

To do so, we assume that each of the  $M$  image rows is a realization of the same RP  $X_t$  and that interpolation is carried on along the rows. Then the empirical estimator is

$$MSE_{emp}[\tilde{X}^{gt} - \tilde{X}^h](t) := \frac{1}{M} \sum_{i=1}^M \left( \tilde{X}_{i,t}^{gt} - \tilde{X}_{i,t}^h \right)^2.$$



**Figure 7.1.** The RMSE estimator is based on a comparison between the interpolation method under consideration applied to a partial image (blue rectangle) and the truncated Shannon-Whittaker interpolation applied to the entire image (green rectangle). Using either a partial row  $X_{i,\cdot}\Delta_K$  or a complete row  $X_{i,\cdot}\Delta_{K'}$ , we can compute two interpolated values at position  $t$ :  $\tilde{X}_{i,t}^h$  and  $\tilde{X}_{i,t}^{gt}$ . Last, we can estimate the mean squared difference by empirically averaging over the rows.

**7.2. Spectrum statistics.** Observing carefully Theorems 4.9 and 5.15, one cannot fail to notice their dependence on three statistics:  $\mu_X$ ,  $\sigma_\alpha^2$  and  $\sigma'_\alpha^2$ . In our experimental study case, since we consider real images, we do not have direct access to these statistics. Therefore we must estimate them from the image. On the one hand, the average value  $\mu_X$  can be evaluated by using the classical empirical estimate

$$\mu_X \simeq \frac{1}{MN} \sum_{i,j} X_{i,j}.$$

On the other hand,  $\sigma_\alpha^2$  and  $\sigma'^2_\alpha$  are spectral statistics that require an estimate of the spectrum of the image. A classical theorem due to Wiener and Khintchine [24] states that

$$d\Psi_X(\omega) = d\mathbb{E}[|\mathcal{F}(X)(\omega)|^2].$$

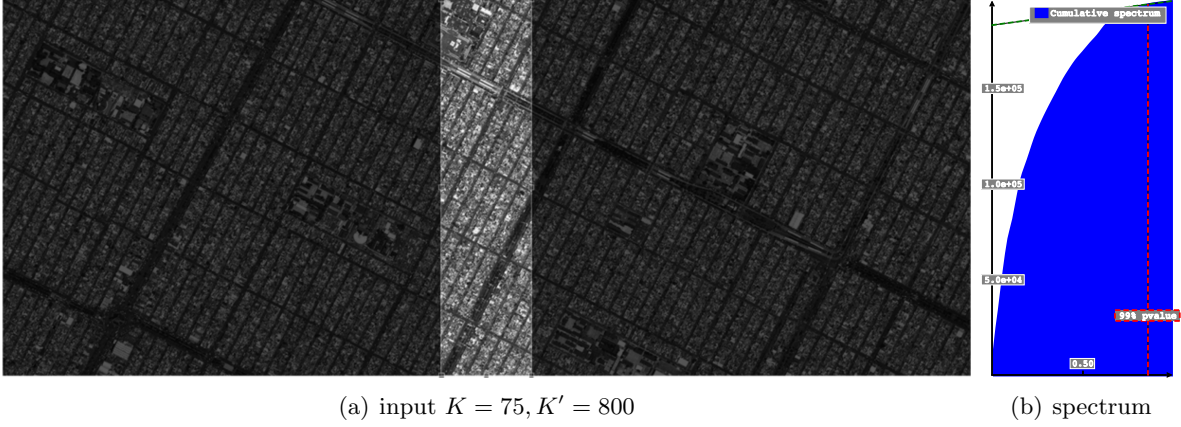
For estimation purpose, we will assume that the spectrum is absolutely continuous

$$d\Psi_X(\omega) = \psi_X(\omega)d\omega.$$

Then the Wiener-Khintchine theorem boils down to the following identity

$$\psi_X(\omega) = \mathbb{E}[|\mathcal{F}(X)(\omega)|^2].$$

Consistently to our approach to empirically estimate the truncation error, we continue to think of each row  $X_{i,t}$  as a realization of a unique RP  $X_t$ . Thanks to the FFT algorithm, for each row  $i$ ,  $|\mathcal{F}(X_i)|^2(\omega)$  can be estimated at uniformly sampled frequency loci  $\omega_l$  on  $[-\pi, \pi]$ ,



**Figure 7.2.** Image of a urban zone, input and cumulative spectrum.

more precisely  $\omega_l := \frac{2\pi l}{N}$  for  $|l| \leq K$ . Subsequently, we can empirically average these values to get  $\psi_X(\omega_l) \simeq \frac{1}{M} \sum_{i=1}^M |\mathcal{DFT}(X_i)|^2(\omega_l)$ . The previous spectrum estimator  $\psi_X(\omega_l)$  is known as a periodogram. It is long established that the periodogram is biased and inconsistent. More sophisticated techniques leading to a better trade-off between bias and variance are commonly preferred. In our case, we have used the multitaper approach [21].

From this point, the estimators for  $\sigma'_\alpha$  and  $\sigma_\alpha^2$  are straightforward

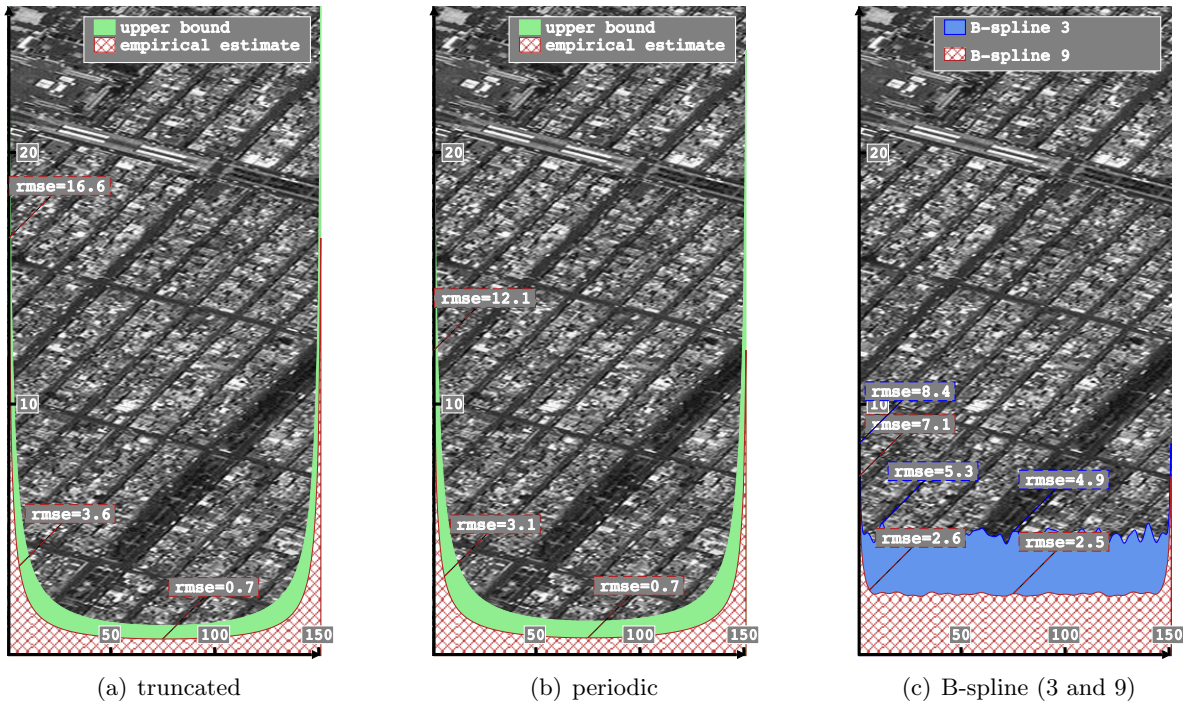
$$\sigma'_\alpha \simeq \max_{\omega_l \geq \alpha\pi} (\psi_X(\omega_l)) \text{ and } \sigma_\alpha^2 \simeq \frac{1}{N} \sum_{\omega_l} \frac{2}{1 + \cos(\omega_l)} \tilde{\psi}_\alpha(\omega_l),$$

with  $\tilde{\psi}_\alpha(\omega_l) = \max(0, \psi_\alpha(\omega_l) - \sigma'^2_\alpha)$ .

**Remark 7.1.** The parameter  $\alpha$  can be chosen arbitrarily in  $[0, 1]$ . This gives us a degree of freedom to optimize the tightness of our upper-bound. In practice, we evaluate the bound for several values of  $\alpha$  and keep the smallest result.

**7.3. Experimental observations.** This part is dedicated to the concrete evaluation of the RMSE estimator and the upper bound for both interpolation methods considered in this study. For each input image, we display in a first figure (such as Figure 7.2) the entire image and a mask highlighting the region that we have selected to test the interpolation accuracy. In the same figure, we display the cumulative spectrum of the image, estimated as explained in Section 7.2. This allows us to visually assess whether the image contains an important white-noise component or if it is more likely over-sampled. In the spectrum plots, the x axis corresponds to the frequency normalized between 0 and 1, that is to say to the parameter  $\alpha$ . Also to help assessing the degree of oversampling, the 99% p-value of the cumulative spectrum is displayed. Lower p-values indicate strong over-sampling.

For a given input image, the actual results are presented in a second figure (such as Figure 7.3), where the curve of the RMSE estimate is overlaid on top of the theoretical upper bound. Each time, two sub-figures are used, showing the results obtained with both interpolation methods. In fact, in order to make the figures more readable we have extracted the



**Figure 7.3.** RMSE estimator and upper bounds for the Urban image. The image interpolated with the considered method is shown in the background. The “naive” B-spline interpolation (3 and 9) shown on the right keeps the observed samples. A better global B-spline interpolate can nonetheless be obtained by a modification of these samples [22].

envelope of the errors, by considering only half-integer locations. For comparison purpose, a third sub-figure displays the RMSE estimator in the case of the 3<sup>rd</sup>-order and 9<sup>th</sup>-order B-spline. Beware that, although prevalent in image processing, the considered B-spline implementation is naive and a better global approximation can be obtained by using a corrective digital filter [22].

To illustrate the different behaviors predicted thanks to our upper bounds, we have chosen two images with different spectral content. The urban image corresponds to a strong texture similar to a white noise and the moor image to a lower one. More examples can be easily considered through our interactive IPOL demo<sup>4</sup>. As a first remark, we would like to point out that the experiments we have driven validate our upper bounds. Besides, the upper bounds provide a fair assessment of the actual error.

Beyond its validation purpose, the RMSE estimator allows us to draw several conclusions with respect to the accuracy of the considered image interpolation techniques. One striking fact concerns the comparison of the accuracy of both interpolation methods to the quantization error, which we recall to be equal to 0.29. In cases with large white noise-like textural component, the interpolation error is everywhere larger than the quantization error. Indeed,

<sup>4</sup>Submitted to Image Processing on Line, (www.ipol.im), workshop, [http://dev.ipol.im/~simonl/ipol\\_demo/loic\\_truncate/](http://dev.ipol.im/~simonl/ipol_demo/loic_truncate/)



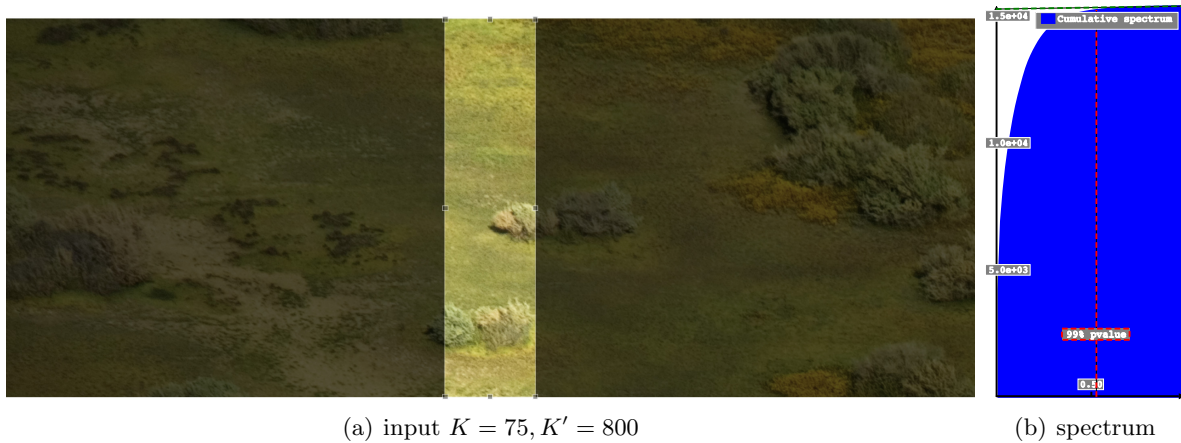


Figure 7.4. Image of a moorland, input and cumulative spectrum.

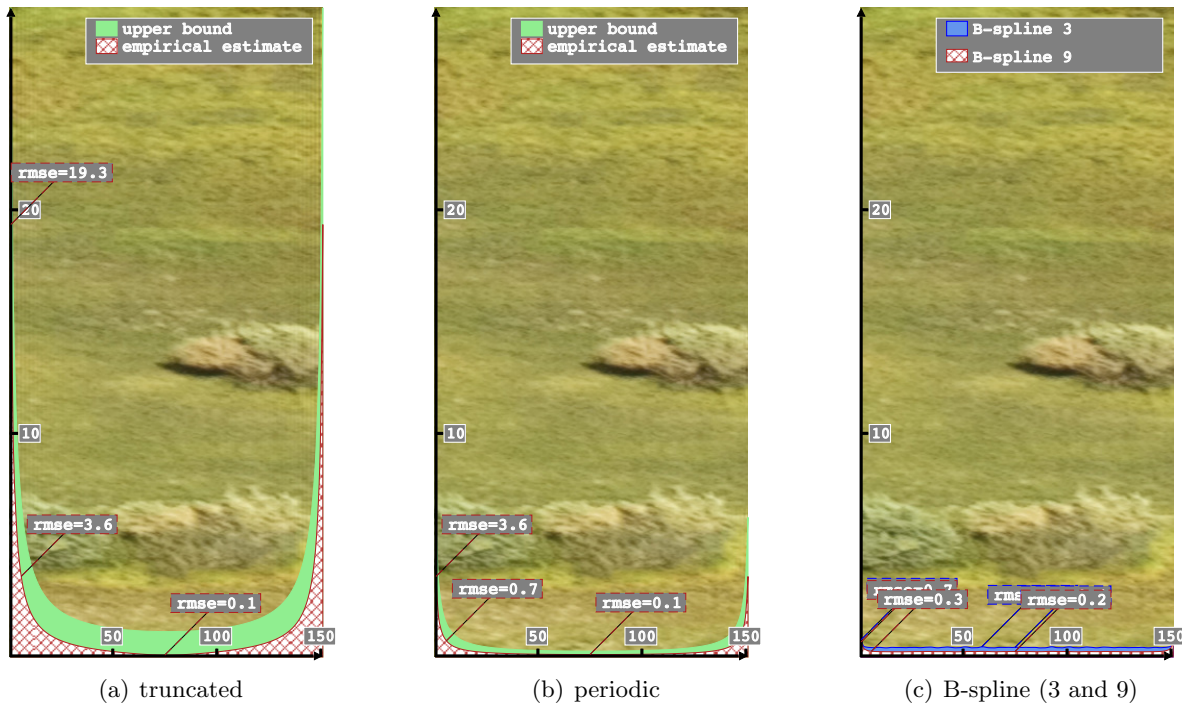


Figure 7.5. RMSE estimator and upper bounds for the Moor image. The image interpolated with the considered method is shown in the background.

at the center of the urban image (that is at a distance of 75 from the border), it is equal to 0.7 for both Shannon-Whittaker interpolations. For sure, the previous fact would not hold true for larger images. And since we have shown that the Shannon-Whittaker interpolation RMSE varies as the square root of the distance to the image border, we can extrapolate that these methods will yield errors negligible compared to quantization at places located further than

750 pixels from the border. This is already quite problematic but it remains manageable. On the contrary, if one desires the interpolation accuracy to compete with quantization in the context of high dynamic range images, the situation would become simply intractable. Indeed, one can interpret HDR images as mapping onto the range of values between 0 and 255 (that is like normal images) and having a quantization step smaller than 1. For instance, this step would be  $\frac{1}{256}$  for images encoded on 16 bits. As a result, the quantization RMSE would be divided by 256, and the necessary distance from the boundary to obtain this given degree of accuracy would be multiplied by  $256^2$ .

Considering the comparison between the Shannon-Whittaker methods and naive B-splines, the most noticeable difference concerns the shape of the RMSE curve. In the case of B-splines the error decreases much faster. This is an expected behavior since the B-spline interpolation kernel itself decreases exponentially. This fast decay leads to a region where the error is approximately constant. Unfortunately, this constant is much larger than the corresponding values for Shannon-Whittaker methods, and thus, larger than the quantization RMSE. The impact of the order of the B-spline is also illustrated in Figure 7.3, where B-spline orders 3 and 9 are considered. The higher order spline leads to a lower error. This observation is coherent with the well-known convergence properties of B-splines towards sinc as the order tends to  $\infty$ . It is however noteworthy that even an 9-order spline results in quite different error magnitudes compared to Shannon interpolation.

**8. Conclusion.** Stimulated by the prospect of high accuracy image processing, we have studied the interpolation errors due to spatial truncation. Sub-pixel image registration is a typical example where such information matters. Although similar studies have been published in the past, their knowledge does not seem widely spread among the aforementioned community. More importantly, their applicability to natural images is limited.

We now summarize the main contributions of the article. First, building upon a generic spectral representation of the interpolation RMSE, we have proposed to decompose the signal into an oversampled part and a white-noise. This simple scheme enabled us to consider the two types of signal separately and to undertake the derivation of dedicated upper bounds. Up to this point every result was expressed generically and could then be applied to any linear and shift-invariant interpolation method. On the contrary, closed-form upper-bounds were only obtained for the truncated Shannon-Whittaker interpolation and the DFT interpolation. Up to minor differences, both methods lead to similar upper bounds. On the contrary, the two types of components yield highly different behaviors of the error decay along the distance to the image boundaries. For over-sampled signals the RMSE decreases as the inverse of the distance, while for white-noise it decreases as its square root.

Besides, we have conducted a set of experiments on real images that permitted us to compare the upper-bounds to an empirical estimate of the RMSE. Not only did this validate the correctness of our theoretical developments, but it also completed the study by providing actual values of the RMSE. As a by-product, we have shown that naive B-splines approaches are superior to the DFT only near the border of the image. However, this conclusion is only valid under the band-limited assumption. On the contrary, we surmise that a local spline interpolation model is more adapted than the Shannon-Whittaker interpolation for aliased or noisy images, and it is systematically better adapted near the boundary. It is also viable when

the image is largely oversampled. Besides, B-splines present several computational advantages and are not prone to visual artefacts such as the Gibbs phenomenon.

In addition, success in high accuracy processing depends on the conditions under which an interpolation approach achieves a level of fidelity comparable to quantization. From this study, one can conclude that for 8-bit quantization the conditions are met at locations from several hundreds to a thousand pixels away from the border. Given the typical range of current image sizes, this constraint is certainly limiting but not prohibitive. However, perspectives are much less promising for HDR images since every additional quantization bit translates virtually into a four-fold rise of the necessary distance.

In future works, we plan to take advantage of the results developed inhere to provide a better analysis of various image processing techniques. On the long run we believe that it will enable us to improve the performance of these techniques. Among others, our priority goes to sub-pixel image registration and subsequently to stereo/multi-view 3D reconstruction.

**Acknowledgments:** Work partly founded by the European Research Council (advanced grant Twelve Labours n°246961), the Office of Naval Research (ONR grant N00014-14-1-0023), and ANR-DGA project ANR-12-ASTR-0035. We thank Thibaud Briand and the anonymous reviewers for valuable suggestions.

## REFERENCES

- [1] AKRAM ALDROUBI AND MICHAEL UNSER, *Sampling procedures in function spaces and asymptotic equivalence with shannon's sampling theory*, Numerical functional analysis and optimization, 15 (1994), pp. 1–21.
- [2] A. BALAKRISHNAN, *A note on the sampling principle for continuous signals*, Information Theory, IRE Transactions on, 3 (1957), pp. 143–146.
- [3] Y.K. BELYAEV, *Analytic random processes*, Theory of Probability & Its Applications, 4 (1959), pp. 402–409.
- [4] JEAN-PAUL BERRUT, *A formula for the error of finite sinc-interpolation over a finite interval*, Numerical Algorithms, 45 (2007), pp. 369–374.
- [5] F BEUTLER, *On the truncation error of the cardinal sampling expansion*, Information Theory, IEEE Transactions on, 22 (1976), pp. 568–573.
- [6] RAJENDRA BHATIA, *Positive definite matrices*, Princeton University Press, 2007.
- [7] THIERRY BLU AND MICHAEL UNSER, *Approximation error for quasi-interpolators and (multi-) wavelet expansions*, Applied and Computational Harmonic Analysis, 6 (1999), pp. 219–251.
- [8] ———, *Self-similarity: Part II—optimal estimation of fractal processes*, Signal Processing, IEEE Transactions on, 55 (2007), pp. 1364–1378.
- [9] J.L. BROWN, *Bounds for truncation error in sampling expansions of band-limited signals*, Information Theory, 15 (1969), pp. 440–444.
- [10] ———, *Truncation error for band-limited random processes*, Information Sciences, 1 (1969), pp. 261–271.
- [11] LL CAMPBELL, *Sampling theorem for the fourier transform of a distribution with bounded support*, SIAM Journal on Applied Mathematics, 16 (1968), pp. 626–636.
- [12] ULF GRENANDER, *Stochastic processes and statistical inference*, Arkiv för matematik, 1 (1950), pp. 195–277.
- [13] HD HELMS AND JB THOMAS, *Truncation error of sampling-theorem expansions*, Proceedings of the IRE, 50 (1962), pp. 179–184.

- [14] D. JAGERMAN, *Bounds for truncation error of the sampling expansion*, SIAM Journal on Applied Mathematics, 14 (1966), pp. 714–723.
- [15] A. J. JERRI, *The Shannon sampling theorem - its various extensions and applications; a tutorial review*, Proceedings of the IEEE, 65 (1977), pp. 1565–1596.
- [16] ERIK MEIJERING, *A chronology of interpolation: from ancient astronomy to modern signal and image processing*, Proceedings of the IEEE, 90 (2002), pp. 319–342.
- [17] ANDRIY YA OLENKO AND TIBOR K POGÁNY, *Universal truncation error upper bounds in sampling restoration*, Georgian Mathematical Journal, 17 (2010), pp. 765–786.
- [18] TSYBAKOV B. S. AND IAKOVLEV V. P., *On the accuracy of restoring a function with a finite number of terms of Kotel'nikov series*, Radio Engineering and Electronics, 4 (1959), pp. 274–275.
- [19] GERT TAMBERG, *Exact values of truncation errors for generalized sampling operators*, SAMPTA, (2011).
- [20] P. THÉVENAZ, T. BLU, AND M. UNSER, *Interpolation revisited*, Medical Imaging, IEEE Transactions on, 19 (2000), pp. 739–758.
- [21] DAVID J THOMSON, *Spectrum estimation and harmonic analysis*, Proceedings of the IEEE, 70 (1982), pp. 1055–1096.
- [22] M. UNSER, *Sampling-50 years after Shannon*, Proceedings of the IEEE, 88 (2000), pp. 569–587.
- [23] MICHAEL UNSER AND AKRAM ALDROUBI, *A general sampling theory for nonideal acquisition devices*, Signal Processing, IEEE Transactions on, 42 (1994), pp. 2915–2925.
- [24] N. WIENER, *Generalized harmonic analysis*, Acta Mathematica, 55 (1930), pp. 117–258.
- [25] Z. XU, B. HUANG, AND X. LI, *On fourier interpolation error for band-limited signals*, Signal Processing, IEEE Transactions on, 57 (2009), pp. 2412–2416.
- [26] K. YAO AND J. B. THOMAS, *On truncation error bounds for sampling representations of band-limited signals*, Aerospace and Electronic Systems, IEEE Transactions on, (1966), pp. 640–647.
- [27] PEI-XIN YE AND ZHAN-JIE SONG, *Truncation and aliasing errors for Whittaker-Kotelnikov-Shannon sampling expansion*, Applied Mathematics-A Journal of Chinese Universities, 27 (2012), pp. 412–418.