



**HAL**  
open science

# Preserving time structures while denoising a dynamical image

Yves Rozenholc, Markus Reiss

► **To cite this version:**

Yves Rozenholc, Markus Reiss. Preserving time structures while denoising a dynamical image. Florack, L. and Duits, R. and Jongbloed, G. and van Lieshout, M.-C. and Davies, L. *Mathematical Methods for Signal and Image Analysis and Representation*, Springer-Verlag, Berlin, pp.207-220, 2012. hal-00712351

**HAL Id: hal-00712351**

**<https://hal.science/hal-00712351v1>**

Submitted on 27 Jun 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Chapter 1

## Preserving time structures while denoising a dynamical image

Yves Rozenholc <sup>\*</sup>, Markus Reiß <sup>\*\*</sup>

### Other related Communications / Chapter

1. Structure enhancement / Line diffusion / DTI  
see also contributions of

- Remco Duits and Erik Franken,
- Bernhard Burgeth *et al*,
- Markus van Almsick,
- Yaniv Gur *et al*,
- Hanno Scharr and Kai Krajsek.

2. Local adaptation / filtering techniques  
see also contributions of

- Luc Florack,
- Michael Felsberg,
- Jörg Polzehl and Karsten Tabelow.

3. Local Markovian Models  
see also contributions of

- Michal Haindl.

4. Low X-Ray CT images  
see also contributions of

- Markus van Almsick.

### Abstract

---

<sup>\*</sup> University Paris Descartes, MAP5, UMR CNRS 8145.

<sup>\*\*</sup> Institute of Mathematics, Humboldt University Berlin.

In restoration or denoising of a movie, the classical procedures often do not take into account the full information provided by the movie. These procedures are either applied spatially "image per image" or locally on some neighborhood combining both closeness in one image and closeness in time. The local information is then combined homogeneously in order to realize the treatment. There is one type of movie where both approaches fail to provide a relevant treatment. Such a movie, called *dynamical image*, represents the same scene along the time with only variations, not in the positions of the objects in the scene but, in their gray levels, colors or contrasts. Hence, at each point of the image, one observes some regular temporal dynamics. This is the typical output using Dynamic Contrast Enhanced Computed Tomography (DCE-CT) or Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) where at each spatial location (called voxel) of the image a time series is recorded.

In such a case, in order to preserve the full temporal information in the image, a proper denoising procedure should be based on averaging over spatial neighborhoods, but using the full dynamics of all pixels (or voxels) within the neighborhood. It is thus necessary to search homogeneous spatial neighborhoods with respect to the full dynamical information.

We introduce a new algorithm which meets these requirements. It is based on two main tools : a multiple hypothesis testing for the zero mean of a vector and an adaptive construction of homogeneous neighborhoods. For both tools, results of mathematical statistics ensure the quality of the global procedure. Illustrations from medical image sequences show a very good practical performance.

**Key words:** Statistical Hypothesis Testing; Time Series; Denoising; Adaptive Estimation; DCE-CT; Microcirculation; Angiogenesis.

## 1.1 Introduction

Classically, the restoration or the denoising of a movie involve 2D techniques for smoothing [Lepski and Spokoiny, 1997, Buades et al., 2008], enhancing (see contributions on structure enhancement and use of line diffusion), filtering (see contribution of Michael Felsberg and references within), partitioning (see [Polzehl and Spokoiny, 2000] or contributions of J. Polzehl and K. Tabelow), *etc.* applied image per image. If some methods use instead of one image a small number of images to take into account the time domain, all methods ignore the full time structure. A natural mathematical reason, outside the computational cost, is that the information contained in one image is considered to be less and less connected to the next images as the time increases. In some sense, the time regular structures existing in a movie do not have a long range effect and disappear quickly as the time distance between two images

increases. One exception to this last rule is given by movies made from one *dynamical image* where at each 2D location (pixel or voxel) one observes not only a gray level or a color, but a full dynamical process represented by a time series or a vector. This is typically the case when the objects in the picture do not have changes in their position, but have changes only their brightness, contrast or color. Medical examples of interest are sequences like DCE-CT or DCE-MRI which register the variations of a contrast medium in the internal tissues of the body in response to a dynamical injection.

Our aim is to develop a new technique, adapted to these *dynamical images*, where each pixel presents a time series. In order to denoise by local averaging, we combine two approaches: statistical multiple hypothesis testing in the spirit of the works [Baraud et al., 2003, Durot and Rozenholc, 2006] to compare the time series between two pixels, together with a spatial growth of locally homogeneous neighborhoods closely related to [Lepskii, 1990, Lepski and Spokoiny, 1997]. The originality of our method is not only to use the full dynamics to compare pixels, but also the way of sequentially growing the neighborhood. While usual paradigms to grow such neighborhoods compare the estimates built on the nested neighborhoods as in [Lepski and Spokoiny, 1997], our method compares, at each step, independent estimates to stop or not the recursive growth. In the context of noise with heavy tails like Laplace noise for example, this comparison using independent estimates and robust statistics like the median offers a clear benefit in reducing the error.

Because the test procedure is based on the comparison to zero of the difference between two noisy vectors, i.e. two noisy time series, no modeling assumption is used. As we use an adaptive multiple hypothesis testing procedure, neither regularization in time is needed nor made. From a theoretical point of view, we only need the dynamics to be not too wild in the sense that the enhancement differences – viewed as a function of time – should remain in some Hölder regularity ball.

The article is organized as follows: In Section 1.2, we introduce the statistical framework of *dynamical images* and summarize our method. The two main statistical tools used to construct this method, multiple testing and neighborhood/ring growth, are respectively studied in 1.3 and in 1.4. In these sections, results from mathematical statistics are provided to guarantee the behavior of our method. Section 1.5 is devoted to the comparison of our estimation technique with estimates introduced in [Lepski and Spokoiny, 1997] and in [Polzehl and Spokoiny, 2000]. Finally in Section 1.6, we briefly describe an application to DCE-CT of renal metastasis in the liver, an extract from the specific work [Reiß et al., 2009a].

## 1.2 Denoising a dynamical image

We consider the following model called *dynamical image* indexed by the two quantities  $t$  and  $x$  representing time and spatial location, respectively:

$$I = \{I_x(t), x \in \mathcal{X}, t \in \{t_1, t_2, \dots, t_K\}\}.$$

Here  $I_x(\cdot)$  denotes a noisy time series at location  $x$ . The notation  $I_x$  will denote the vector

$$(I_x(t_1), \dots, I_x(t_K)),$$

of the discrete observation at times  $t_1, \dots, t_K$  of this time series. In our setting,  $\mathcal{X}$  denotes the finite grid of pixels.

We assume that the noisy observations  $I_x(t_k)$  may be written as

$$I_x(t_k) = i_x(t_k) + \sigma \varepsilon_k^x, \quad (1.1)$$

where  $i_x(t)$  denotes an unobservable true gray level at time  $t$ , where  $\varepsilon_k^x$  denotes a standardized noise with fixed distribution and where  $\sigma$  denotes the level of noise. We assume that the noise variables  $\varepsilon_k^x$  are independent and identically distributed with respect to both, space location  $x$  and (time index)  $k$ . It is not necessary to assume that  $\sigma$  and the distribution of  $\varepsilon^x$  are known, but we will illustrate our construction using the simple setting  $\sigma$  known and  $\varepsilon^x$  having a Gaussian distribution.

In order to compare the full time series, we use multiple hypothesis tests to decide whenever the vectors  $I_x$  and  $I_y$  at spatial locations  $x$  and  $y$  are statistically similar or not. For this purpose, we ask if the expectation of difference vector  $I_x - I_y$  is the zero vector or not at a given level  $\alpha$  which could be considered as a natural tuning parameter. Such a test is presented in Section 1.3.

Given a set  $V$  of pixels, let us define the estimated dynamics using  $V$  as the coordinate-wise empirical mean

$$\hat{I}_V = \frac{1}{|V|} \sum_{y \in V} I_y, \quad (1.2)$$

where  $|A|$  denotes the cardinality of a set  $A$ . Our aim is to construct at each spatial location  $x$ , a spatial neighborhood  $\mathcal{V}_x$  of  $x$  made from voxels  $y$  such that the difference  $I_x - I_y$  does not deviate significantly from the zero vector and such that the bias of the estimated dynamics obtained by replacing  $V$  in (1.2) by  $\mathcal{V}_x$  remains under control. Controlling the statistical bias means that we aim at using a procedure which ensures that the principal goal of nonparametric statistics, which is the trade-off between bias and variance, is achieved. This method is introduced in Section 1.4 and can be summarized as follows:

Let a geometrically increasing sequence  $(n_i)$  of positive integers be given with  $n_1 = 1$ . For a fixed location  $x \in \mathcal{X}$ , we consider the set  $\mathcal{W}_x$  of all pixels  $y$  such that, at a level  $\alpha \in (0, 1)$ ,  $I_x - I_y$  is accepted to have a zero mean by the previous test. We set  $i = 0$  as well as  $V_0 = \emptyset$  and  $W_0 = \{x\}$ . Our method repeats the following steps until one test, at least, rejects the null hypothesis:

- Increment  $i$ .
- Define  $V_i = V_{i-1} \cup W_{i-1}$  and build the set  $W_i$  of the  $n_i$  closest  $y$  in  $\mathcal{W}_x \setminus V_i$ .
- For  $j = 1, \dots, i$ , test the  $i - 1$  null hypotheses  $\mathbb{E}(\hat{I}_{W_i} - \hat{I}_{V_j}) = 0$  for  $j = 1, \dots, i - 1$  against a non-zero expectation.

The final estimate is  $\hat{I}_{V_{i-1}}$ . The algorithm realizes a pointwise adaptive selection of homogeneous dynamics. Because the last step involves a multiple hypothesis test, in order to ensure that its level is  $\alpha$ , each individual test has to be calibrated with respect to  $\alpha$  and  $i$  by a Bonferroni correction.

### 1.3 Multiple testing

Given two spatial locations  $x$  and  $y$ , we present the statistical test used to compare to the zero vector the expectation of the difference vector  $Z$  with components

$$Z_k := I_y(t_k) - I_x(t_k), \quad k = 1 \dots K.$$

Such a test of comparison to the zero vector derives from the theoretical works [Baraud et al., 2003, Durot and Rozenholc, 2006] which consider a more general framework where  $\sigma$  is unknown and/or where  $\varepsilon$  is not necessarily assumed to be Gaussian. The theoretical study relies on large  $K$  asymptotics, which in our case means that we observe the continuous time signal more and more frequently. For the sake of simplicity, we write  $Z = f + \sqrt{2}\sigma\varepsilon$  and we introduce these tests in their simplified version which is the Gaussian case with  $\sigma$  known. We aim to test that the mean vector  $f$  is zero or not and hence consider the hypotheses

$$\mathcal{H}_0 : "f = 0" \text{ versus } \mathcal{H}_1 : "f \neq 0".$$

To enlighten the notations, we suppose that  $K$ , the number of time indices, is of the form  $K = 2^d$  and we consider the regular dyadic decomposition of the observation times  $t_1 \dots t_{2^d}$ . For  $j = 0 \dots d - 1$ , we denote by  $T_1^j \dots T_{2^j}^j$  the  $2^j$  intervals of time indexes

$$T_l^j = \{t_k : k = 2^{d-j}(l-1) + (1 \dots 2^{d-j})\}.$$

The set  $T_l^j$  contains  $2^{d-j}$  time indices. Given  $j$  in  $0 \dots d - 1$ , let us introduce the projection of the observation  $Z$  onto the space spanned by the vectors with same coordinates on each time index  $T_l^j$ :

$$\Pi_j Z = \underbrace{(m_1^j, \dots, m_1^j)}_{2^{d-j} \text{ times}}, \dots, \underbrace{(m_{2^j}^j, \dots, m_{2^j}^j)}_{2^{d-j} \text{ times}}$$

with the local mean over  $T_l^j$

$$m_l^j = \frac{1}{2^{d-j}} \sum_{t \in T_l^j} Z_t.$$

The test is based on the size of the squared Euclidean norm  $\|\Pi_j Z\|_K^2$  which is:

$$2^{d-j} \sum_{l=1}^{2^j} (m_l^j)^2 = \frac{1}{2^{d-j}} \sum_{l=1}^{2^j} \left( \sum_{t \in T_l^j} Z_t \right)^2.$$

Under  $\mathcal{H}_0$ , the difference vector  $Z = I_y - I_x$  is a centered Gaussian vector with covariance matrix  $2\sigma^2 Id_K$ , where  $Id_K$  denotes the identity matrix in  $\mathbb{R}^K$ . Hence, under  $\mathcal{H}_0$ ,

$$\|\Pi_j Z\|_n^2 / 2\sigma^2 \sim \chi^2(2^j)$$

which defines a distribution-free test statistic. Let us denote by  $\Psi_D^{-1}$  the inverse cumulative distribution function of a  $\chi^2$  with  $D$  degrees of freedom, using a Bonferroni correction, our test procedure works as follows:

Reject  $\mathcal{H}_0$  at level  $\alpha$  if for any  $j = 0 \dots d-1$ ,  $\|\Pi_j Z\|_K^2 / 2\sigma^2 > \Psi_{2^j}^{-1}(\alpha/d)$ .

If  $\sigma$  is unknown, instead of a  $\chi^2$ -test based procedure, [Baraud et al., 2003] proposes a Fisher like construction and considers the following procedure:

Reject  $\mathcal{H}_0$  at level  $\alpha$  if for any  $j = 0 \dots d-1$ ,

$$\frac{\|\Pi_j Z\|_K^2 / 2^j}{\|Z - \Pi_j Z\|_K^2 / 2^{K-j}} > \mathcal{F}_{2^j, 2^{K-j}}^{-1}(\alpha/d),$$

where  $\mathcal{F}_{2^j, 2^{K-j}}^{-1}$  denotes the inverse cumulative distribution function of a Fisher distribution with  $2^j$  and  $2^{K-j}$  degrees of freedom. Moreover, if the distribution of the noise variables  $\varepsilon_k^x$  is unknown, as  $\varepsilon := (\varepsilon^x - \varepsilon^y) / \sqrt{2}$  is by construction standardized and symmetrical, under the weak condition that the common distribution of the  $\varepsilon$  does not put mass in 0, we can follow the construction proposed in [Durot and Rozenholc, 2006]:

Consider  $p$  Rademacher vectors  $R_1, \dots, R_p$  of length  $K$ . Rademacher vectors have independent coordinates  $R_i$  with  $P(R_i = +1) = P(R_i = -1) = 1/2$ . Introduce the vectors  $N_j = Z \otimes R_j$  where  $\otimes$  denotes the coordinate-wise product. The previous constructions can be generalized by considering the following procedure:

Reject  $\mathcal{H}_0$  at level  $\alpha$  if for any  $j = 0 \dots d - 1$ ,

$$\|\Pi_j Z\|_K^2 > \mathcal{Q}^{-1} [\|\Pi_j Z \otimes R_1\|_K^2, \dots, \|\Pi_j Z \otimes R_p\|_K^2] (\alpha/d),$$

where  $Q^{-1}[X_1, \dots, X_K]$  denotes the inverse cumulative distribution of the empirical distribution of the  $X_i$ ,  $i = 1, \dots, K$ .

Introducing the continuous-time signal  $F$  defined by  $F(t) := i_x(t) - i_y(t)$ , the difference vector  $Z = I_x - I_y$  follows the fixed-design regression model

$$Z_k = f_k + \sqrt{2}\varepsilon_k = F(t_k) + \sqrt{2}\sigma\varepsilon_k.$$

In this framework, the introduced multiple hypothesis tests are known to have a power which may be controlled on balls of the form  $\|f\|_K \geq \rho(K)$  where  $\rho(K)$  is a radius decreasing with  $K$ , the number of components of  $f$  (= number of observation times in the *dynamical image*). Moreover, these multiple hypothesis tests are adaptive with respect to the unknown Hölder regularity  $s$  of the unknown function  $F$  : at a fixed level  $\alpha$  and for a given fixed power  $1 - \beta$ , this test automatically achieves (up to logarithmic factors) the minimax rate of testing  $\rho_s(K)$  obtained in [Gayraud and Pouet, 2005] for all regularities  $s > 1/4$ . More precisely, we can derive following [Baraud et al., 2003, Th 1] and [Durot and Rozenholc, 2006, Th 4]) the following control of the power :

**Theorem 1 :**

If the distribution of  $\varepsilon_1$  does not put mass on 0, assuming that : (Bernstein's conditions)  $\max_{1 \leq i \leq n} E(\varepsilon_i^{2p}) \leq \gamma p! \mu^{p-2}$  holds for all integers  $p \geq 1$  and some positive  $\gamma$  and  $\mu$ , then there are  $A_1, \dots, A_4 > 0$  such that the multi-test procedure is of power  $1 - \beta$  outside the ball defined by  $\|f\|_K^2 \leq \inf_j \Delta(J_j, f, A_1, \dots, A_5)$  where

$$\begin{aligned} \Delta(J_j, f, A_1, \dots, A_5) &= A_1 \|f - \Pi_{J_j} f\|_K^2 + A_2 \sqrt{\frac{\gamma}{\beta}} 2^{(j-1)/2} + \\ &+ A_3 (\sqrt{\gamma} + \mu + \max_{k=1..K} (f_k^2)) \left[ 1 + \frac{j \log 2 + \log(1/\beta)}{2} \right] \log \left( \frac{2d}{\beta\alpha} \right) + \\ &+ A_4 (\sqrt{\gamma} + \mu + \max_{k=1..K} (f_k^2)) 2^{(j-1)/2} \sqrt{\left[ 1 + \frac{j \log 2 + \log(1/\beta)}{2} \right] \log \left( \frac{2d}{\beta\alpha} \right)} \end{aligned}$$

**Corollary 1** If  $f_k = F(t_k)$  then the test is adaptative w.r.t. the Hölder regularity  $s$  of  $F$  and achieves the optimal minimax rate for  $s > 1/4$ .

This theorem is the exact translation of [Durot and Rozenholc, 2006, Thm. 4], using our simple setting with  $K = 2^d$  and dyadic partitions. It ensures a good power of the test (up to logarithmic terms) if the true unknown function to



be tested is, at least for one dyadic partition, at a squared distance from 0 large enough to ensure that a noisy version has also its squared distance from 0 large enough. Let us recall that the squared distance of a noisy version is of the order of the sum of the variance term given by the number of pieces of the partition and the bias term  $\|f - \Pi_{J_j} f\|_K^2$  which measures how well  $f$  can be approximated by a piecewise constant vector on the index partition.

We have now at hand the tool to compare dynamics at two locations  $x$  and  $y$  with respect to the known noise level  $\sigma$ . We denote " $I_y \equiv_{\sigma^2}^\alpha I_x$ " when the enhancement difference average vector  $i_x - i_y$  is accepted to be the zero vector at level  $\alpha$  with respect to a noise variance  $2\sigma^2$  following the above construction.

## 1.4 Spatial neighborhood growth

We now present the construction of the spatial neighborhood  $\mathcal{V}_x$  for a fixed location  $x$ . In all that follows, the distance between spatial locations  $x$  in  $\mathcal{X}$  can be measured by any metric ; usually the Euclidean metric will be taken.

The main idea of the sequential neighborhood growth can be explained in terms of testing. Suppose that neighborhoods  $V_1 \subset V_2 \subset \dots \subset V_M$  of  $x$  are given. Then the stochastic fluctuations in the mean  $\hat{I}_{V_i}$  of the observed signal over  $V_i$ , as measured by the variances, decrease with  $i$ . On the other hand, the bias terms usually increase with  $i$  (we smooth with a larger bandwidth). We can assume that the bias in  $\hat{I}_{V_1}$  is negligible (or we just set  $V_1 = \{x\}$ ). Considering next the empirical mean  $\hat{I}_{V_2 \setminus V_1}$  over the observations in the set difference  $V_2 \setminus V_1$ , a test is constructed to decide on the hypothesis that  $\mathbb{E}(\hat{I}_{V_2 \setminus V_1} - \hat{I}_{V_1}) = 0$  (recall that  $\mathbb{E}Z$  denotes the coordinate-wise expectation of a vector  $Z$ ). If the test rejects, we conclude that the bias of the new observations is already too large and we take  $V_1$  as neighborhood, thus keeping  $\hat{I}_{V_1}$  as estimator. Otherwise, we accept  $V_2$  as new neighborhood and consider  $V_3 \setminus V_2$ . If  $\hat{I}_{V_3 \setminus V_2}$  does not deviate significantly from  $\hat{I}_{V_1}$  and from  $\hat{I}_{V_2}$ , we continue. Otherwise, we stop and select  $V_2$  as neighborhood,  $\hat{I}_{V_2}$  as estimator. This way we continue growing the neighborhoods until the first time the new observations differ statistically significantly from any of the estimators on the previously built neighborhoods or until we reach the last index  $M$ .

This simple idea is reminiscent of Lepski's method [Lepskiĭ, 1991, Lepski and Spokoiny, 1997] in mathematical statistics with the important difference that at the  $i$ th step we do not test whether the new estimator  $\hat{I}_{V_{i+1}}$  is statistically homogeneous with all previous  $\hat{I}_{V_j}$ ,  $j \leq i$ , but we compare only the average  $\hat{I}_{V_{i+1} \setminus V_i}$  over the new observations with the previous estimates. This has the clear advantage that the underlying test statistics are based on independent quantities and that in practice the number of too large grown neighborhoods is reduced because changes in the signal are detected earlier.

In order to correctly define our method, it is, of course, of great importance how to define precisely the tests and how to preselect the potential neighborhoods ( $V_i$ ). In the static case (i.e. we only observe a scalar, not a time series at each point) this is described in full detail in [Reiß et al., 2009b], here we describe the main points. Let us first indicate how two averages  $\hat{I}_V$  and  $\hat{I}_{V'}$  over disjoint sets  $V$  and  $V'$  are compared. The test statistic is always based on the size of the difference  $Z_{V,V'} = \hat{I}_V - \hat{I}_{V'}$ . In the static case and for Gaussian noise, we can standardize  $Z_{V,V'}^2$  by its expectation under the null, which is  $\rho(V, V') = |V|^{-1} + |V'|^{-1}$ , and compare it with some critical value  $z_{V,V'}$ . This critical value can be specified by a Bonferroni correction of the corresponding  $\chi^2$ -quantile, taking into account the multiple testing. A more refined and powerful, but also computationally more demanding calibration of the quantiles in this case is proposed in [Reiß et al., 2009b]. Together with the prescription of geometrically growing neighborhoods ( $V_i$ ), it is proved there that this results in estimators that optimally adapt to the local regularity of the signal (in the asymptotic minimax sense over Hölder balls). We thus achieve an unsupervised estimation procedure which is pointwise adaptive.

Here, the focus is on dynamic images and we need to compare the mean of time series over different spatial locations  $V$  and  $V'$ . For this we use exactly the adaptive multiple testing procedure presented in the preceding section. A second feature is that an a priori prescription of potentially good neighborhoods ( $V_i$ ) around each pixel or voxel  $x$  is not feasible. The potential neighborhoods should rather be preselected based on the data observed in order to adapt to possibly heterogeneous geometric structures like borders or bands where a priori defined balls or squares are obviously not adequate. In a first preprocessing step we therefore exclude pixels that carry with high probability a significantly different time series. On the pixels kept after preprocessing we apply our neighborhood growth, which eventually results in the following algorithm.

We consider the set

$$\mathcal{W}_x = \{y \in \mathcal{X} \text{ s.t. } y \neq x \text{ and } y \equiv_{\sigma^2}^{\alpha} x\}$$

of the spatial locations for which the time series are statistically similar to those of  $x$  with respect to the multi-test introduced in Section 1.3 above. Let us emphasize that  $\mathcal{W}_x$  is neither necessarily contiguous with  $\{x\}$  nor connected.

Our algorithm works as follows:

1. Start with  $i = 0$  and set  $V_0 = \{x\}$  and  $\hat{I}_0 = I_x$  as the first neighborhood and its associated denoised time series.
2. Find  $W_i$  the subset of the  $C_0 2^i$  closest points to  $V_i$  in  $\mathcal{W}_x \setminus V_i$ . We call this set the "ring" around  $V_i$ , alluding to the case  $\mathcal{W}_x = \mathcal{X}$  with the Euclidean distance.
3. Compute  $\hat{J}_i$  the estimated time series using locations in the ring  $W_i$ .

4. Test if  $\hat{J}_i$  is statistically not to be distinguished from all previously constructed estimates  $\hat{I}_j$  for  $j = 0 \dots i$ . See hereafter for precision on this test.
5. If equality is accepted, consider

$$V_{i+1} = V_i \cup W_i$$

and define the estimate  $\hat{I}_{i+1}$  using locations in  $V_{i+1}$ . Return to 2) with  $i = i + 1$ .

6. Else stop, define  $\mathcal{V}_x = V_i$  and use  $\hat{I}_x = \hat{I}_i$  as denoised time series in  $x$ .

To compare  $\hat{J}_i$  with the previously constructed estimates  $\hat{I}_0, \dots, \hat{I}_i$ , we test the hypothesis

$$\mathcal{H}_0 : \text{"}\mathbb{E}(\hat{J}_i - \hat{I}_j) = 0, \text{ for all } j = 0 \dots i\text{"},$$

against

$$\mathcal{H}_1 : \text{"}\mathbb{E}(\hat{J}_i - \hat{I}_j) \neq 0, \text{ for at least one } j\text{"},$$

using a generalization of our test introduced in Section 1.3. We accept  $\mathcal{H}_0$  if

$$\hat{J}_i \equiv \frac{\alpha/(i+1)}{\sigma^2 \rho(|V_j|, |W_i|)} \hat{I}_j.$$

The use of  $\alpha/(i+1)$  is a Bonferroni correction due to multiple hypothesis tests. It would be possible to use more clever corrections as proposed in [Benjamini and Hochberg, 1995] for example. The correction factor  $\rho$  in the noise level takes into account that the estimates come from independent samples with respective size  $|V_j|$  and  $|W_i|$ , which in the Gaussian case is just the variance  $\rho(|V_j|, |W_i|) = |V_j|^{-1} + |W_i|^{-1}$ .

## 1.5 Static toy examples

Not yet exploring the full scope of the method for treating dynamical images, we present first how the denoising is accomplished for one- and two-dimensional signals (images) without an attached time series structure.

### 1.5.1 1D example

We propose a simple 1D example to compare the use of rings combined with balls with the algorithm introduced in [Lepski and Spokoiny, 1997] based only on balls. Given an increasing sequence of neighborhoods  $V_j$ , indexed by  $j = 1, \dots, j_{\max}$ , the selected index in [Lepski and Spokoiny, 1997] is defined by

$$\hat{j} = \inf \left\{ j \geq 0 \text{ s.t. } \exists \ell \leq j, \frac{\|\hat{I}_{V_j} - \hat{I}_{V_\ell}\|^2}{\mathbb{E}[\|\hat{I}_{V_j} - \hat{I}_{V_\ell}\|^2]} > z_\ell^2 \right\} \wedge j_{\max}$$

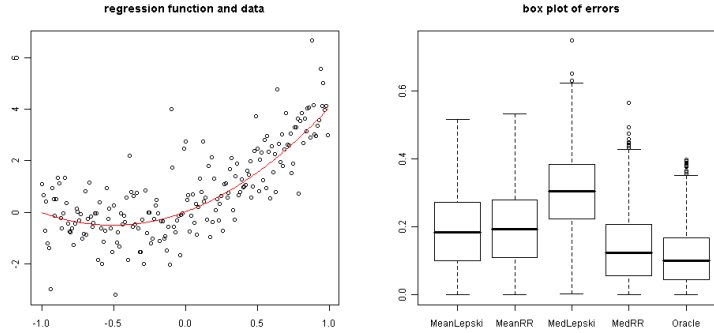
while for our method the index is selected as follow

$$\hat{j} = \inf \left\{ j \geq 0 \text{ s.t. } \exists \ell \leq j, \frac{\|\hat{I}_{V_j \setminus V_\ell} - \hat{I}_{V_\ell}\|^2}{\mathbb{E}[\|\hat{I}_{V_j \setminus V_\ell} - \hat{I}_{V_\ell}\|^2]} > z_\ell^2 \right\} \wedge j_{\max}$$

Both methods achieve the same rate of convergence for  $\alpha$ -Hölder classes of signals in  $\mathbb{R}^d$ :

$$\mathbb{E}[\|\hat{I}_{V_j} - i\|^2] \leq c \left( \frac{\log n}{n} \right)^{2\alpha/(2\alpha+d)},$$

where  $n$  is the size of the neighborhood. This rate is known to be optimal in a minimax sense over classes of Hölder functions, already in the classical mean regression model with known errors (see [Lepskiĭ, 1991]). The logarithmic term has to be paid in order to be adaptive. Note that the dimension  $d$ , driven by the dimension of the neighborhood, is two for classical images, the length  $K$  of the time series itself does not change the rate, but will certainly influence the finite sample performance: more data in time increases the signal to noise ratio.



**Fig. 1.1** A simple example with a parabolic signal. Estimation point is  $x = 0$  and noise is Laplace distributed. Errors from 1000 Monte Carlo replications

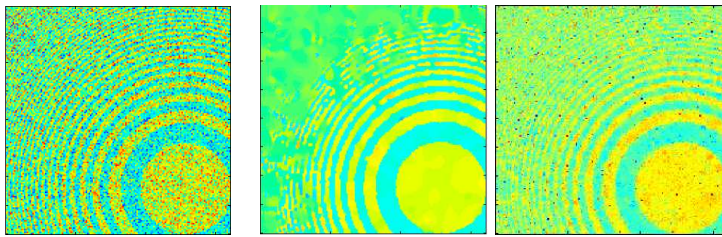
In the left part of Figure 1.1 the true signal is presented together with a typical sample of 200 data points obtained by adding iid Laplace distributed noise. The estimation point is at  $x = 0$  in the center. As possible balls, i.e. intervals, around  $x$  we take  $V_k := [-a_j, +a_j]$  with  $a_j = 4 \cdot (5/4)^j / 200$  with  $j = 1, 2, \dots$ . We compare the following methods : (1) balls as in Lepski [Lepski and Spokoiny, 1997] using the local mean; (2) balls/ring using the local mean; (3) balls as in Lepski [Lepski and Spokoiny, 1997] using the

local median; (4) balls/ring using the local median; all these methods are compared to the (5) oracle where the median is taken over the (a posteriori optimal) interval  $[-0.39, +0.39]$ , which is slightly larger than  $V_{10}$ . We have used the empirical median in addition to the empirical mean (1.2) since for Laplace distributed noise the median is more efficient and robust. The calibration of the methods follows always the same algorithm as recommended in [Reiß et al., 2009b]. The boxplots present the errors computed for the same 1000 samples with each method. The boxplots in Figure 1.1 show a comparable performance for our method in the mean case with the benchmark Lepski approach while in the median case a clear advantage for our ball/ring method can be seen. It is remarkable that the oracle error is almost achieved.

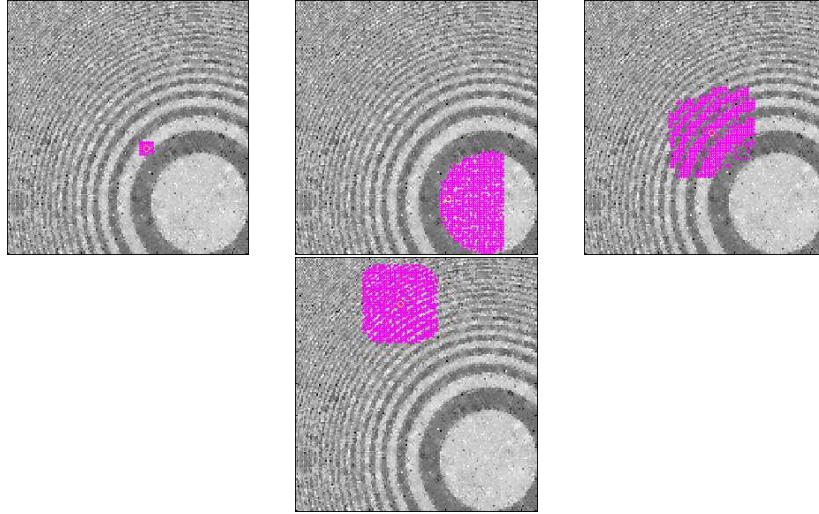
### 1.5.2 A 2D example

Figure 1.2 compares our method to AWS introduced in [Polzehl and Spokoiny, 2000]. The left subfigure shows a noisy image obtained using a pattern with two values and a Laplace noise with noise level equal to the difference of these two values. The middle subfigure shows the denoised version using AWS, a typical benchmark procedure for off-the-shelf spatially adaptive denoising. We have used the R package AWS version 1.3-3.1 developed by the authors of [Polzehl and Spokoiny, 2000] available on the web site of the CRAN using the default 2D settings to produce this subfigure. Finally, the right subfigure shows the denoised image obtained by our method. While, at the used default setting, AWS offers a very clean picture in the right lower corner, all details are lost in the upper left corner. Using our method, the signal is less smoothed out, but allows to recover details even in the last part of the upper corner. Clearly more details can be found in this last image even if one could complain about the remaining noise due the noise level used.

In Figure 1.3, four examples of selected neighborhoods are presented. One can remark the special effect of the first selection of the set  $\mathcal{W}_x$  in the presented examples: the neighborhood follows the borders and can even be not connected, allowing for rich geometrical structures in the signal.



**Fig. 1.2** (left) Noisy data - (center) AWS - (right) Balls/Rings



**Fig. 1.3** Examples of neighborhood (pink crosses) found for 4 selected points of interest (red circle)

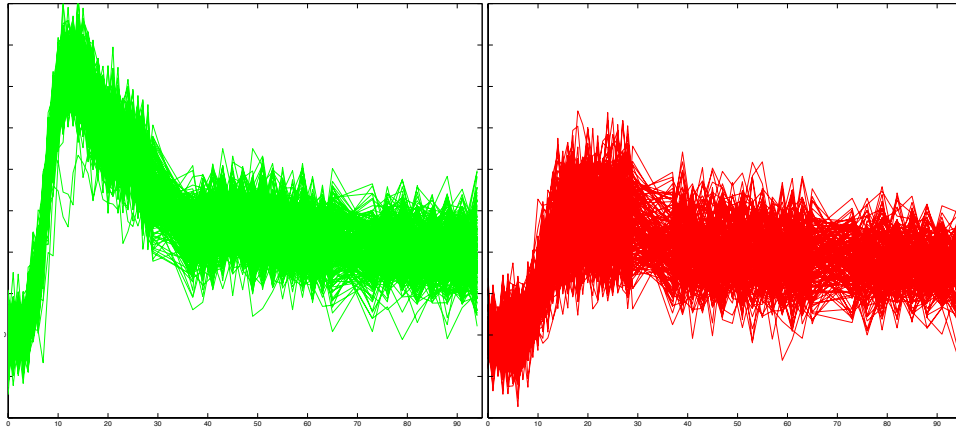
## 1.6 Practical results

A complete study, using our technique in the framework of DCE-CT, has been conducted with radiologists at the European George Pompidou Hospital (HEGP), Paris, and the complete results are reported in [Reiß et al., 2009a]. In order to provide a flavor of how our method works on real data, we have extracted from this study a few illustrations. The precise setting of the medical experiments as well as the way the tuning parameters like  $\sigma$  and  $\alpha$  are fixed is left to that work.

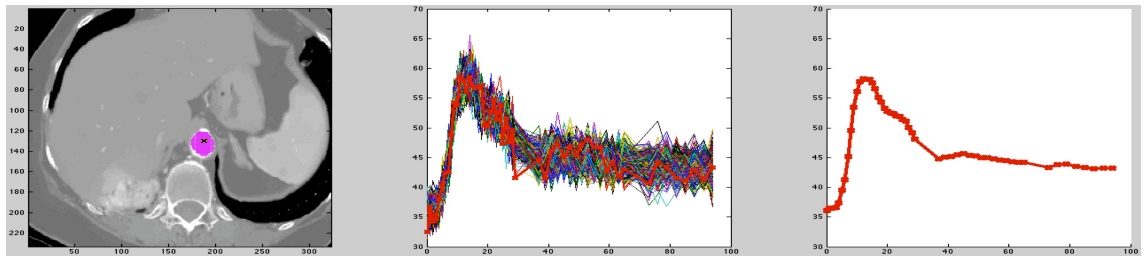
Due to the limited irradiation dose used during the sequential acquisition, the *dynamical images* in this study suffer from a poor signal-to-noise ratio. Figure 1.4 shows typical time series obtained from voxels in manually selected Regions of Interest (of size  $\simeq 100$  voxels) within the Aorta (left curves) and a tumor (right curves).

The Figures 1.5 and 1.6 present the result of our method applied at two different voxels, one inside the aorta and the other inside the tumor.

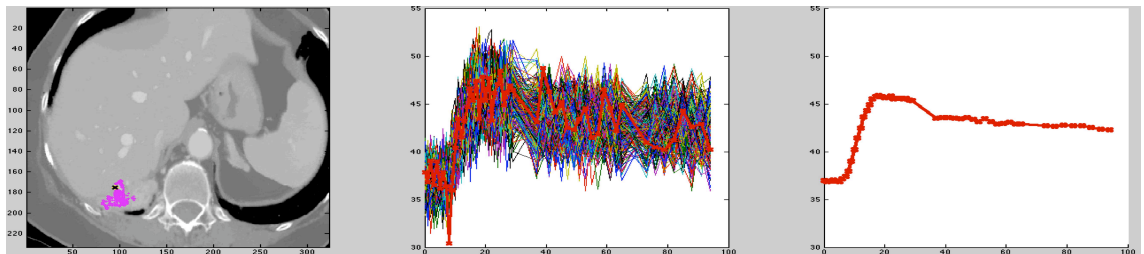
Each figure is divided into 3 sub-figures. Left: the denoised slice at a specific time, with the selected voxel  $x$  (black cross) and the voxels from its neighborhood  $\mathcal{V}_x$  (pink dots). Center: the original enhancement vector  $I_x$  (red thick curve) and the associated enhancement vectors  $I_y$  for  $y \in \mathcal{V}_x$  (background curves). Right: the estimated enhancement using a generalized median as a robust center of the selected neighborhood (see Section 1.4).



**Fig. 1.4** Typical curves (time, enhancement) of individual voxels in ROI manually drawn within the Aorta (left) and the tumor (right)



**Fig. 1.5** Voxel within the aorta



**Fig. 1.6** Voxel within the tumor

## 1.7 Conclusion

Using a two-step procedure, each step based on multiple hypothesis testing, we have introduced a novel algorithm to denoise *dynamical images* where each point of the picture exhibits a complete time series. Based on a comparison of the time series dynamics, this algorithm preserves the full structure of

the time series by averaging spatially over adaptively grown neighborhoods. The efficiency of our algorithm is shown on artificial examples and on real medical images. The quality of the denoised dynamical image is shown by the remarkable reconstruction of the details in these spatially heterogeneous images.

## Acknowledgment

This work was supported in 2008 by a grant "Bonus Qualité Recherche" from the University Paris Descartes for the project "Cancer Angiogénèse et Outils Mathématiques".

## References

- [Baraud et al., 2003] Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251.
- [Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.
- [Buades et al., 2008] Buades, A., Coll, B., and Morel, J.-M. (2008). Nonlocal image and movie denoising. *Int. J. Comput. Vision*, 76(2):123–139.
- [Durot and Rozenholc, 2006] Durot, C. and Rozenholc, Y. (2006). An adaptive test for zero mean. *Math. Methods Statist.*, 15(1):26–60.
- [Gayraud and Pouet, 2005] Gayraud, G. and Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields*, 133(4):531–558.
- [Lepski and Spokoiny, 1997] Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6):2512–2546.
- [Lepskiĭ, 1990] Lepskiĭ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35(3):454–466.
- [Lepskiĭ, 1991] Lepskiĭ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, 36(4):682–697.
- [Polzehl and Spokoiny, 2000] Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(2):335–354.
- [Reiß et al., 2009a] Reiß, M., Rozenholc, Y., Balvay, D., and Cuenod, C.-A. (2009a). Growing time homogeneous neighborhoods for denoising and clustering dynamic contrast enhanced-ct sequences. <http://www.math-info.univ-paris5.fr/~rozen/RRBC-preprint/last-version.pdf>, Submitted.
- [Reiß et al., 2009b] Reiß, M., Rozenholc, Y., and Cuenod, C.-A. (2009b). Pointwise adaptive estimation for robust and quantile regression. *preprint arXiv:0904.0543v1*, Submitted.