Discretized Wavelet Density Estimators for Continuous Time Stochastic Processes

Frédérique Leblanc[1]

[1] CREST, ENSAE, Timbre J340, 3 av. P. Larousse, 92241 Malakoff Cedex, France
[2] LSTA, Université Paris VI, pl. Jussieu, 75005 Paris, France

Abstract

We consider a strictly stationary continuous time stochastic process (X_t) such that the law of X_t has a bounded density f. Under certain assumptions which are satisfied for rather general diffusion processes, the L_2 error of the linear wavelet estimator of f constructed from the observation $(X_t, 0 \le t \le T)$ converges with the rate $\sqrt{1/T}$ when $f \in B^s_{p,q}$. In this work we study two discretized versions of this estimator, constructed from the dicrete observations $X_{t_1}, ..., X_{t_n}$, $0 \le t_i \le T$. We show that there is a minimal sampling size n that we should take to preserve the upper bound $\sqrt{1/T}$ for the L_2-error.

1 Introduction

In this work we consider a real strictly stationary process $(X_t, t \in I\!R^+)$, such that its marginal law has a density with respect to Lebesgue measure. We estimate the restriction of this density f to a compact set $[-K, K]$. In [Le1] the following estimator of the restriction of f to $[-K, K]$ is constructed from the observations $(X_t, 0 \le t \le T)$:

$$\hat{f}_T(x) = \sum_{k \in K_{j_0}} \hat{\alpha}_{j_0,k}\varphi_{j_0,k}(x), \qquad \hat{\alpha}_{j_0,k} = \frac{1}{T}\int_0^T \varphi_{j_0,k}(X_t)dt \qquad \text{with} \quad 2^{j_0} = \left[T^{\frac{1}{2s'}}\right], \tag{1}$$

where $s' = s - 1/p + 1/2$, $\varphi_{j_0,k}(x) = 2^{j/2}(2^j x - k)$ and φ is in the space $C^{r+1}, r \in I\!N$ has a compact support and is the scale function of a multiresolution analysis on the interval $]-K, K[$. Note that the cardinality of the set K_{j_0} is of order 2^{j_0}.

For some processes, the L_2-error of the estimator (1) converges with the rate $\sqrt{1/T}$ when f is in a ball of the Besov space $B^s_{p,q}$, $\quad s > 1/p, \quad 1 \le p \le 2, \quad$ and $\quad q \ge 1$ (see Appendix for a definition of Besov spaces).

Castellana and Leadbetter [CL] have shown that the rate $1/T$ (they called it "full rate") is attained for the quadratic error of δ-family estimator when the density regularity is $s = 2$ under condition of uniform integrability of the joint density $f_t(x, y)$ of (X_0, X_t), over $]0, +\infty[$. They have verified this assumption for certain gaussian processes.

Note that the hypothesis of integrability around zero is intimately related to the local regularity of trajectories and the integrability at $+\infty$ can be shown when the process (X_t) is strong mixing. For a wide class of diffusions processes sufficient assumptions to achieve the rate $\sqrt{1/T}$ for the L_2-error, are established in [Le1]. Moreover, the "parametric" lower bound $\sqrt{1/T}$ for the L_2-error can be obtained, if for instance, (X_t) is a solution of a linear stochastic differential equation (see section 3.3 in [Ku]).

In order to implement the estimation algorithm (1) we have to compute the integral in (1). To do this the trajectories of the process (X_t) should be discretized. In what follows we consider two types of discretization:

- **deterministic design**: when n observations are taken at the deterministic sampling instants t_i, $1 \le i \le n$.

A particular case of this type of sampling is the regular design obtained when $t_i = Ti/n$, $1 \le i \le n$. We refer to this case as to the regular deterministic design;

- **randomized design**: when t_i, $1 \le i \le n$ are random instants (independent of $(X_t, 0 \le t \le T)$).

The particular case where the random variables t_i, $1 \le i \le n$ are independent and identically distributed according to the uniform distribution over $[0,T]$, is called here the uniform random design.

First we consider the estimator obtained with regular deterministic design $t_1 = \delta_n, t_2 = 2\delta_n, ..., t_n = n\delta_n$ where $\lim_{n\to\infty} \delta_n = 0$ and $n\delta_n = T$ (we refer to T as the time of observation):

$$\bar{f}_{n,T}(x) = \sum_{k\in K_{j_0}} \bar{\alpha}_{j_0,k}\varphi_{j_0,k}(x) \quad \text{with} \quad \bar{\alpha}_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{j_0,k}(X_{i\delta_n}) \quad \text{and} \quad 2^{j_0} = \left[T^{\frac{1}{2s'}}\right]. \tag{2}$$

Next we study the discretized estimator constructed with the uniform random design (using n observations):

$$\tilde{f}_{n,T}(x) = \sum_{k\in K_{j_0}} \tilde{\alpha}_{j_0,k}\varphi_{j_0,k}(x) \quad \text{with} \quad \tilde{\alpha}_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{j_0,k}(X_{t_i}) \quad \text{and} \quad 2^{j_0} = \left[T^{\frac{1}{2s'}}\right]. \tag{3}$$

Let us denote $(t_{(1)}, ..., t_{(n)})$ the ordered sequence of the sampling instants $t_1, ..., t_n$. Then we can also write

$$\tilde{\alpha}_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{j_0,k}(X_{t_{(i)}}).$$

To implement the estimator (3) we generate the discretization sample $(t_1, ..., t_n)$ and take the observations of the process (X_t) at the sampling instants $t_{(1)}, ..., t_{(n)}$.

Our goal is to find, for both estimators (2) and (3), the minimal sampling size n that we should take to preserve the rate of convergence $1/T$ for the mean integrated square error.

Our result is that in both cases the minimal sampling size (up to a constant) is $n = \left[T^{(2s'+1)/(2s')}\right]$ when the density belongs to $B^s_{p,q}$, $s' = s - 1/p + 1/2$.

Kernel discretized estimators of the density have been already studied in the literature for particular randomized or deterministic designs (c.f. [Ma], [PR]). For example, in [Ma] such estimators are studied when it is supposed that $\min_{1\le i\le n} |t_{i+1} - t_i| \ge \nu > 0$. In this case (under the above additional constraint on the discretization scheme) the quadratic error is $O(n^{-2s/(2s+1)})$ where s is the regularity of the estimated density. Since for any $1 \le i \le n$, $|t_{i+1} - t_i| \ge \nu$, it is obvious that it is then not possible to consider more than T/ν observations of (X_t) on the time interval $[0, T]$. Hence, the quadratic error is of order $T^{(-2s)/(2s+1)}$ (up to a constant) with respect to the "observation time" T. When substituting brutally T for N (N is considered as the time of observation in the discrete case), we find again the optimal rate obtained for discrete time processes (see [Le2]). The explanation is that when supposing that the distance in time between two observations can not be smaller than $\nu > 0$, the situation becomes analogous to that of the discrete time case.

In this work we do not suppose that $\min_{1\le i\le n} |t_{i+1} - t_i| \ge \nu > 0$ but, on the contrary, that this quantity tends to zero when n (or T) tends to infinity. Due to the use of the local properties of the process (X_t), translated in condition **C1** or **C2** (see Section 2), we recover the rate $1/T$ (w.r.t. the time of observation) for the mean integrated square error.

Note that for kernel density and regression estimators the problem with the regular deterministic design is also considered in [Bo] and in [CP]. When the density has the regularity $s = 2$ (thus adapted to quadratic error problem) and the process satisfies a slightly stronger assumption than **C2**, the "minimal" size of the discretization sample which permits to preserve the "full rate" is of order $n^{5/4}$.

The paper is organized as follows: in Sections 2 and 3 we describe our assumptions and a class of diffusion processes which satisfy it. In Section 4 we present the main results (their proof are put in Section 5). Finally in Section 6 we present some simulation results.

2 Assumptions

We consider a stationary, measurable, locally integrable stochastic process (X_t) defined on the probability space $(\Omega, \mathcal{F}, P)$ which takes its values on $\mathbb{R}$. We suppose that the law of X_t has a bounded density f with respect to Lebesgue measure. Furthermore, we assume that the density, with respect to the Lebesgue measure on $\mathbb{R}^2$, f_u of (X_0, X_u) does exist for each $u > 0$ and we set $g_u = f_u - f \otimes f,\ \ u > 0$. Now the basic assumptions are

C1: for any Q and for any $(x, y) \in \Delta^2 = [-Q, Q]^2$

$$\int_0^\infty |g_u(x, y)| < \infty.$$

This assumption is used in [Le1] to obtain the rate $\sqrt{1/T}$ and is slightly weaker than hypothesis used in [CL].

C2: There exists a function π such that for any $\Delta = [-Q, Q]$,
- $\sup_{(y,z)\in\Delta^2} |g_u(y, z)| \le \pi(u),\ \ u > 0$.
- $\pi \in L_1(]0, +\infty[)$ and π decreases over $]0, \infty[$.

Note that **C2** implies **C1**.

3 Examples

Consider the diffusion process defined by the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{4}$$

and the initial condition

$$X_0 \text{ distributed according to the invariant law,} \tag{5}$$

where W_t is the standard brownian motion and where the drift and the diffusion coefficient satisfy the followings assumptions:

D1 There exist constants B, σ_1 and σ_0 such that for all x:

$$\begin{aligned} |b(x)| &\leq B(1+|x|), \\ \sigma_1 &\geq \sigma^2(x) \geq \sigma_0 > 0. \end{aligned}$$

There exists a ball $B_r = \{x, |x| < r\}$ such that outside B_r:

$$xb(x) \leq -\gamma|x| \text{ for some } \gamma > 0.$$

D2 There exist constants B', C, C', $0 < a \leq 1$ and $0 < a' \leq 1$ such that for all $(x,y) \in I\!R^2$,

$$\begin{aligned} \left|\frac{db(x)}{dx}\right| &\leq B', \\ \left|\frac{db(x)}{dx} - \frac{db(y)}{dy}\right| &\leq C|x-y|^a, \\ \left|\frac{d^2\sigma^2(x)}{dx^2} - \frac{d^2\sigma^2(y)}{dy^2}\right| &\leq C'|x-y|^{a'}. \end{aligned}$$

Under the above assumptions **D1** and **D2** the condition **C1** holds (see [Le1] p. 17-18). Furthermore, turning back to the proof of this result we can also establish **C2**.

To obtain the integrability of g_u when u tends to ∞ we use a result in [Ve2] which garanties the geometrical mixing of the process (X_t) under assumption **D1**. The integrability around zero is proved, when using a result in [BGJ] which gives the behavior of the transition probability density (from X_0 to X_t) for small values of t, under assumption **D2**. For a detailed proof we refer to [Le1].

When the diffusion coefficient $\sigma(\cdot)$ in (4) is a constant, the regularity conditions on the drift $b(\cdot)$ (in **D2**) can be relaxed. Indeed, the process solution of (4) with initial condition (5) in the particular case $\sigma(x) = 1$ and $b(x) = -\theta\text{sign}(x)$ satisfies assumption **C2** (see [Le1]). Note that, in this case the invariant density is the double exponential of parameter θ.

4 Main Results

$I\!E$ denote the expectation according to the law of the process (X_t), $F_{s,p,q} = \{f \in B^s_{p,q} | \ \|f\|_{B^s_{p,q}} \leq \infty\}$ (see Appendix for the definition of $\|\cdot\|_{B^s_{p,q}} \leq \infty\}$) and $\|\cdot\|_{2,K}$ the L_2-norm on the interval $]-K,K[$. When applying Theorem 3 of [Le1] to the particular case $p'=2$ we have for the estimator (1):

Theorem 1 *Let $f \in F_{s,p,q}$, $s > \frac{1}{p}$, $1 \leq p \leq 2$, $1 \leq q$,. We suppose that assumption* **C1** *is satisfied. Then for any K there exist a constant C such that for all T*

$$I\!E\|\hat{f}_T - f\|^2_{2,K} \leq \frac{C}{T} \tag{6}$$

where $s' = s - 1/p + 1/2$ and j_0 is such that $2^{j_0} = \left[T^{1/(2s')}\right]$.

We obtain for the estimator (2):

Theorem 2 *Let $f \in F_{s,p,q}$, $s > \frac{1}{p}$, $1 \leq p \leq 2$, $1 \leq q$. We suppose that* **C2** *is satisfied. If $n = \left[T^{\frac{2s'+1}{2s'}}\right]$ where $s' = s - 1/p + 1/2$, then for any K there exists a constant $\overline{C}$ such that for all T,*

$$I\!E\|\bar{f}_{n,T} - f\|^2_{2,K} \leq \frac{\overline{C}}{T},$$

where j_0 is such that $2^{j_0} = [T^{1/(2s')}]$.

For the estimator (3) we have an analoguous result:

Theorem 3 *Let $f \in F_{s,p,q}$, $s > \frac{1}{p}$, $1 \leq p \leq 2$, $1 \leq q$. We suppose that* **C1** *is satisfied. Then if $n = \left[T^{\frac{2s'+1}{2s'}}\right]$ where $s' = s - 1/p + 1/2$, for any K there exists a constant $\tilde{C}$ such that for all T,*

$$I\!E E_h\|\tilde{f}_{n,T} - f\|^2_{2,K} \leq \frac{\tilde{C}}{T}.$$

where h is the density of the uniform distribution on $[0,T]$, E_h the mathematical expectation with respect to h and j_0 is such that $2^{j_0} = [T^{1/(2s')}]$.

Remark: Some random linear approximations are proposed in [BC] and [IL] for stochastic integral of the form

$$\int_0^1 g(X_t)dt$$

and for some particular processes (X_t) (for example, gaussian processes are considered in [IL]). In both above works it is shown that the size n and the deterministic sampling design $(t_i, 1 \leq i \leq n)$ have to be adapted to the quadratic regularity of the trajectories $(X_t, 0 \leq t \leq 1)$. When applying these results for the particular case of the Ornstein-Uhlenbeck process, to discretize the integral in (1), we obtain that the regular deterministic sampling where $n = O(T^{\frac{2s'+1}{2s'}})$ give the best discretization.

In the case of diffusion processes (which is the case of interest for us) the quadratic regularity of the trajectories seems to be straightly linked to the behavior of the density transition probabilities $p_t(x,y)$ when t is near zero. Due to the result obtained for the Ornstein-Uhlenbeck process when using results of [IL], we can think that irregular deterministic design would not give a better result.

5 Proofs

In the following C, $\overline{C}$, $\tilde{C}$, c, c', $\overline{c}$ and $\tilde{c}$ are constants which do not depend on n and T. Theorem (2) (resp. Theorem (3)) is an immediat consequence of the following Proposition 1 (resp. Proposition 2).

Proposition 1 *Let Assumption* **C2** *holds and suppose that* $f \in F_{s,p,q}$, $s > \frac{1}{p}$, $1 \le p \le 2$, $1 \le q$. *Then for any* K *there exists a constant* $\overline{C}$ *such that for any* T *and* n,

$$\mathbb{E}\|\bar{f}_{n,T} - f\|_{2,K}^2 \le \overline{C}\left(\frac{1}{T} + \frac{T^{\frac{1}{2s'}}}{n}\right) \tag{7}$$

where $s' = s - 1/p + 1/2$.

Proposition 2 *Let Assumption* **C1** *holds and suppose that* $f \in F_{s,p,q}$, $s > \frac{1}{p}$, $1 \le p \le 2$, $1 \le q$. *Then for any* K, *there exists a constant* $\tilde{C}$ *such that for any* T *and* n,

$$\mathbb{E}E_h\|\tilde{f}_{n,T} - f\|_2^2 \le \tilde{C}\left(\frac{1}{T} + \frac{T^{\frac{1}{2s'}}}{n}\right) \tag{8}$$

■ **Proofs of Theorem 2 and 3:**
It is sufficient to take $n = [T^{(2s'+1)/(2s')}]$ in the inequality (7) (resp. (8)) to obtain Theorem 2 (resp. Theorem 3). Note that if we take n such that $n = o(T^{(2s'+1)/(2s')})$ we loose the "full rate" $1/T$ and if we take n such that $T^{(2s'+1)/(2s')} = o(n)$ then the rate $1/T$ is attained. We call the sample size $n = [T^{(2s'+1)/(2s')}]$ optimal (up to a constant) in the sense that it is the minimal value of n that we should take in the upper bounds of (7) and (8) to keep the rate $1/T$. ■

■ **Proofs of Propositions 1 and 2:** We denote h the density of the uniform distribution on $[0,T]$ and E_h the mathematical expectation with respect to h. To light the notations we denote $\|\cdot\|_2$ for $\|\cdot\|_{2,K}$. We have

$$\mathbb{E}\|\bar{f}_{n,T} - f\|_2^2 = \mathbb{E}\|\bar{f}_{n,T} - \hat{f}_T\|_2^2 + \mathbb{E}\|\hat{f}_T - f\|_2^2, \tag{9}$$

$$\mathbb{E}E_h\|\tilde{f}_{n,T} - f\|_2^2 = \mathbb{E}E_h\|\tilde{f}_{n,T} - \hat{f}_T\|_2^2 + \mathbb{E}\|\hat{f}_T - f\|_2^2. \tag{10}$$

In both above equalities the upper bound for the second terms of the right hand side is given by Theorem 1, that is

$$\mathbb{E}\|\hat{f}_T - f\|_2^2 \le \frac{C}{T}. \tag{11}$$

To bound the first term of the right hand side of (9) (resp. (10)) we use

$$\mathbb{E}\|\bar{f}_{n,T} - \hat{f}_T\|_2^2 = \sum_{k \in K_{j_0}} \mathbb{E}(\bar{\alpha}_{j_0,k} - \hat{\alpha}_{j_0,k})^2, \tag{12}$$

$$\mathbb{E}E_h\|\tilde{f}_{n,T} - \hat{f}_T\|_2^2 = \sum_{k \in K_{j_0}} \mathbb{E}E_h(\tilde{\alpha}_{j_0,k} - \hat{\alpha}_{j_0,k})^2. \tag{13}$$

Next we establish the two following lemmas:

Lemma 1

$$\mathbb{E}(\bar{\alpha}_{j_0,k} - \hat{\alpha}_{j_0,k})^2 \leq \overline{c}\left(\frac{2^{-j_0}}{n\delta_n} + \frac{1}{n}\right)$$

Lemma 2

$$\mathbb{E}E_h(\tilde{\alpha}_{j_0,k} - \hat{\alpha}_{j_0,k})^2 \leq \frac{\tilde{c}}{n}$$

■ **Proof of Lemma 1:**

$$\mathbb{E}(\bar{\alpha}_{j_0,k} - \hat{\alpha}_{j_0,k})^2 \leq 2\mathbb{E}(\hat{\alpha}_{j_0,k} - \alpha_{j_0,k})^2 + 2\mathbb{E}(\bar{\alpha}_{j_0,k} - \alpha_{j_0,k})^2. \tag{14}$$

The first term of the right hand side of this inequality has been studied in [Le1]. Indeed, since **C1** holds we have by Proposition 6, of [Le1]:

$$\mathbb{E}(\hat{\alpha}_{j_0,k} - \alpha_{j_0,k})^2 = \frac{1}{T^2}\mathbb{E}\left(\int_0^T(\varphi_{j_0,k}(X_s) - \mathbb{E}\varphi_{j_0,k}(X_s))ds\right)^2 \leq c\frac{2^{-j_0}}{T}. \tag{15}$$

Now we study the second term in the right hand side of (14). Since $\mathbb{E}\bar{\alpha}_{j_0,k} = \alpha_{j_0,k}$ we have

$$\begin{aligned}
\mathbb{E}|\bar{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2 &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\varphi_{j_0,k}(X_{i\delta_n}) - \mathbb{E}\varphi_{j_0,k}(X_{i\delta_n}))\right)^2 \\
&= \frac{1}{n^2}\left(\underbrace{\sum_{i=1}^{n}\mathbb{E}(\varphi_{j_0,k}(X_{i\delta_n}) - \mathbb{E}\varphi_{j_0,k}(X_{i\delta_n}))^2}_{v_n}\right. \\
&+ \left.\underbrace{2\sum_{i<i'}cov(\varphi_{j_0,k}(X_{i\delta_n}), \varphi_{j_0,k}(X_{i'\delta_n}))}_{c_n}\right)
\end{aligned} \tag{16}$$

Next by wavelet properties we have

$$\mathbb{E}(\varphi_{j_0,k}(X_{i\delta_n}))^2 \leq \int_{\mathbb{R}}\varphi^2_{j_0,k}(x)f(x)dx \leq \|f\|_\infty 2^{j_0}\int_{\mathbb{R}}\varphi^2(2^{j_0}x - k)dx \leq \|f\|_\infty.$$

Hence,

$$v_n \leq \|f\|_\infty n \tag{17}$$

To bound the term depending on the covariance we express c_n as

$$\begin{aligned}
c_n &= 2\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} cov((\varphi_{j_0,k}(X_{i\delta_n}), \varphi_{j_0,k}(X_{i'\delta_n})) \\
&= 2\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} cov((\varphi_{j_0,k}(X_0), \varphi_{j_0,k}(X_{(i'-i)\delta_n})) \text{ by strict stationarity} \\
&= 2\sum_{i=1}^{n-1}\sum_{i'=1}^{n-1} cov((\varphi_{j_0,k}(X_0), \varphi_{j_0,k}(X_{i'\delta_n})) \\
&= 2\sum_{i=1}^{n-1}(n-i) cov((\varphi_{j_0,k}(X_0), \varphi_{j_0,k}(X_{i\delta_n})). \qquad (18)
\end{aligned}$$

But,

$$cov(\varphi_{j_0,k}(X_0), \varphi_{j_0,k}(X_{i\delta_n})) = \int_{(y,z)\in[-K,K]^2} \varphi_{j_0,k}(y)\varphi_{j_0,k}(z) g_{i\delta_n}(y,z)\,dydz,$$

thus

$$\begin{aligned}
|c_n| &= 2\left|\sum_{i=1}^{n-1}(n-i)\left\{\int_{(y,z)\in[-K,K]^2} \varphi_{j_0,k}(y)\varphi_{j_0,k}(z) g_{i\delta_n}(y,z)\right\}\right| \\
&\leq 2\delta_n^{-1} n \int_{(y,z)\in[-K,K]^2} |\varphi_{j_0,k}(y)|\,|\varphi_{j_0,k}(z)| \left\{\sum_{i=1}^{n-1} \delta_n |g_{i\delta_n}(y,z)|\right\} dydz \\
&\leq 2\delta_n^{-1} n \int_{(y,z)\in[-K,K]^2} |\varphi_{j_0,k}(y)|\,|\varphi_{j_0,k}(z)| \left\{\sum_{i=1}^{n-1} \delta_n \pi(i\delta_n)\right\} dydz \\
\text{by } \mathbf{C2} \quad &\leq 2\delta_n^{-1} n \int_{(y,z)\in[-K,K]^2} |\varphi_{j_0,k}(y)|\,|\varphi_{j_0,k}(z)| \left\{\sum_{i=1}^{n-1} \left(\int_{(i-1)\delta_n}^{i\delta_n} \pi(u)du\right)\right\} dydz \\
&\leq 2\delta_n^{-1} n \left(\int |\varphi_{j_0,k}(y)| dy\right)^2 \|\pi\|_1 \\
&\leq 2\|\pi\|_1 \|\varphi\|_1^2 \delta_n^{-1} n 2^{-j_0} \qquad (19)
\end{aligned}$$

Now by substituting the bounds (19) and (17) in (16) we have

$$I\!\!E|\bar{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2 \leq c'\left(\frac{1}{n} + \frac{2^{-j_0}}{n\delta_n}\right). \qquad (20)$$

Finally from (14), (15) and (20) permit to conclude. ∎

Proof of Lemma 2:

Let us fixe the trajectory $(X_t, 0 \leq t \leq T)$ and estimate $E\{(\hat{\alpha}_{j_0,k} - \tilde{\alpha}_{j_0,k})^2|(X_t, 0 \leq t \leq T)\}$. Since the random variables are supposed to be independent of $(X_t, 0 \leq t \leq T)$ we have

$$E\{(\hat{\alpha}_{j_0,k} - \tilde{\alpha}_{j_0,k})^2|(X_t, 0 \leq t \leq T)\} = E_h(\hat{\alpha}_{j_0,k} - \tilde{\alpha}_{j_0,k})^2.$$

$(t_1, ..., t_n)$ are i.i.d. uniform random variables, thus we get

$$\begin{aligned}
E_h\tilde{\alpha}_{j_0,k} &= \frac{1}{n}\sum_{i=1}^{n} E_h(\varphi_{j_0,k}(X_{t_i})) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\int_0^T \varphi_{j_0,k}(X_t)dt\right) = \hat{\alpha}_{j_0,k}. \qquad (21)
\end{aligned}$$

Then

$$\begin{aligned} E_h(\hat{\alpha}_{j_0,k}-\tilde{\alpha}_{j_0,k})^2 &= E_h(\tilde{\alpha}_{j_0,k}-E_h(\tilde{\alpha}_{j_0,k}))^2 \\ &= E_h\left|\frac{1}{n}\sum_{i=1}^{n}(\varphi_{j_0,k}(X_{t_i})-E_h(\varphi_{j_0,k}(X_{t_i}))\right|^2. \end{aligned}$$

Next $E_h(\varphi_{j_0,k}(X_{t_i})\varphi_{j_0,k}(X_{t_j}))=0$ for $i\neq j$ because the variables $(t_i, 1\leq i\leq n)$ are i.i.d. random variables independent of $(X_t, 0\leq t\leq T)$, hence

$$\begin{aligned} E_h(\hat{\alpha}_{j_0,k}-\tilde{\alpha}_{j_0,k})^2 &= \frac{1}{n^2}\sum_{i=1}^{n}E_h(\varphi_{j_0,k}(X_{t_i})-E_h\varphi_{j_0,k}(X_{t_i}))^2 \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}E_h(\varphi_{j_0,k}(X_{t_i}))^2 \\ &\leq \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{T}\left(\int_0^T\varphi_{j_0,k}^2(X_t)dt\right). \end{aligned}$$

Now by taking the expectation with respect to the density f of X_t we obtain

$$I\!E(E_h(\tilde{\alpha}_{j_0,k}-\hat{\alpha}_{j_0,k})^2) \;\leq\; \frac{1}{n^2}\sum_{i=1}^{n}I\!E\left(\frac{1}{T}\int_0^T\varphi_{j_0,k}^2(X_t)dt\right). \tag{22}$$

But we have

$$\begin{aligned} I\!E\left(\int_0^T\varphi_{j_0,k}^2(X_t)dt\right) &= \int_0^T I\!E(\varphi_{j_0,k}^2(X_t))dt \\ &\leq \int_0^T\left(\int\varphi_{j_0,k}^2(x)f(x)dx\right)dt \\ &\leq \tilde{c}\,T. \end{aligned} \tag{23}$$

Thus

$$I\!E E_h(\tilde{\alpha}_{j_0,k}-\hat{\alpha}_{j_0,k})^2\leq\frac{\tilde{c}}{n},$$

hence the inequality presented in Lemma 2. ∎

To finish the proofs of Proposition 1 and 2, we recall that $card(K_{j_0})=O(2^{j_0})$, and substituting the upper bounds which are gived by Lemmas 1 and Lemma 2 into inequalities (12) and (13) we obtain

$$I\!E\|\bar{f}_{n,T}-\hat{f}_T\|_2^2 \;\leq\; \overline{C}\left(\frac{2^{j_0}}{n}+\frac{1}{\delta_n n}\right)\leq\overline{C}\left(\frac{T^{\frac{1}{2s'}}}{n}+\frac{1}{T}\right), \tag{24}$$

because $2^{j_0}=\left[T^{\frac{1}{2s'}}\right]$ and $n\delta_n=T$.

And we have for Proposition 2:

$$I\!E E_h\|\tilde{f}_{n,T}-\hat{f}_T\|_2^2 \;\leq\; \tilde{C}\left(\frac{2^{j_0}}{n}\right)\leq\tilde{C}\left(\frac{T^{\frac{1}{2s'}}}{n}\right). \tag{25}$$

Next the bounds (24) (resp. (25)) and (11) inserted in (9) (resp. (10)) give Proposition 1 (resp. Proposition 2). ∎

6 Simulations

We present here some simulations results for the estimator (2). This work have been done in collaboration with A. Juditsky and a detailed study can be found in [JL]. To compute the estimator we implement a fast algorithm (see [DJ]). The whole effort to compute the estimator is of order $(n \ln n)$ for the sample size n. To implement estimator (2) we use the `TeachWave` Matlab-toolbox developped by D. Donoho and I. Johnstone [DoJo]. This toolbox contains various programs which implement, among other things, fast wavelet transform and a convenient graphic interface.

6.1 Modelling Diffusion Processes

We turn now to the modelling of continuous time diffusion process $(X_t, 0 \leq t \leq T)$, which satisfies the stochastic differential equation (SDE)

$$X_t = b(X_t)dt + dW_t.$$

Let δ be the step of discretisation, and X_t^δ the discrete approximation of X_t. Note that, obviously δ has to be much smaller than δ_n which denotes the step of discretization of the interval $[0, T]$ and determines the observations used in the algorithm (2).

We simulate the following processes:

(i) *Gaussian:* (X_t) is the Ornstein-Uhlenbeck process, i.e. the solution of the linear SDE:

$$dX_t = -1/2 X_t dt + dW_t, \quad X_0 = 0.$$

The invariant density for this equation is

$$f(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2/2).$$

(ii) *Chameau:* (X_t) is a solution of the linear SDE:

$$dX_t = b(X_t)dt + dW_t, \quad X_0 = 0,$$

where

$$b(x) = \begin{cases} -x-2 & \text{for} \quad x < -1, \\ x & \text{for} \quad -1 \leq x < 1, \\ -x+2 & \text{for} \quad x \geq 1. \end{cases}$$

The invariant density of X_t is

$$f(x) = C_0 exp(g(x)),$$

with

$$g(x) = -(x+2)^2 \cdot 1_{\{x<-1\}} - (x-2)^2 \cdot 1_{\{x \leq 1\}} + (x^2 - 2) \cdot 1_{\{-1 \leq x < 1\}};$$

and $C_0 \approx 0.2731$;

(iii) *Laplace:* (X_t) is a solution of the SDE:

$$dX_t = -\mathrm{sign}(X_t)dt + dW_t, \quad X_0 = 0.$$

The invariant density f in this case is

$$f(x) = \exp(-2|x|);$$

(iv) *Trident:* (X_t) is the solution of the equation

$$dX_t = b(X_t)dt + dW_t, \quad X_0 = 0,$$

where

$$b(x) = \begin{cases} 1 & \text{for} \quad x < -2,\ -1 \leq x < 0,\ 1 \leq x < 2, \\ -1 & \text{for} \quad -2 \leq x < -1,\ 0 \leq x < 1,\ x \geq 2; \end{cases}$$

with the invariant density

$$f(x) = C_1 exp(g(x)),$$

with

$$g(x) = 2x \cdot 1_{\{x<-2\}} - 2x \cdot 1_{\{x \geq 2\}} + (-8 - 2x) \cdot 1_{\{-2 \leq x < -1\}}$$
$$+ (-4 + 2x) \cdot 1_{\{-1 \leq x < 0\}} + (-4 - 2x) \cdot 1_{\{0 \leq x < 1\}} + (-8 + 2x) \cdot 1_{\{1 \leq x < 2\}}$$

and $C_1 \approx 20.0042$;

We present the "normalized" plots of the above densities on figure 1.

6.2 Comments

- Dependence on n: figures 2 and 3 illustrate very well the theoretical result that there is an "optimal" number of observations to use to achieve the rate $\sqrt{1/T}$. Indeed, when the discretization sample size is large enough the error stabilizes because the estimation error associated to the continuous time estimator (1) is already attained.

- Dependence on T and n: In the experiment on figure 4 the couple $(T, n(T))$ used to compute the estimor (2) is such that $n(T)$ is at least as large as the theoretical optimal $n = [T^{(2s'+1)/2s'}]$. This graph is the result of a mean of ten sample paths. The rate of convergence given by the slopes for the "Gaussian" and "Laplace" densities is nearly $\sqrt{1/T}$. For other two densities the result is less impressive due to a slower mixing rate.

7 Appendix: Besov spaces

We do not recall the definition of a multiresolution analysis on $I\!R$ which can be found in [Da]. We denote φ for the scale funcion and ψ the wavelet. The wavelet decomposition can also be obtained on any interval. For instance, one can construct an orthonormal basis of $L_2((-K,K))$ which satisfies the same properties than a wavelet orthonormal basis of $L_2(I\!R)$. It is sufficient to correct the wavelets which step on the boundary of the interval $[-K,K]$ (see [AHJP] and [CDJV]).

There are several equivalent definitions of Besov spaces (see [Tr]). We use the following one: Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$,

for $0 < s < 1$ let denote,

$$\gamma^s_{p,q}(f) = \left(\int_{I\!R} \left(\frac{\|\tau_h f - f\|_p}{|h|^s} \right)^q \frac{dh}{|h|} \right)^{\frac{1}{q}}, \text{ where } \tau_h f(x) = f(x-h)$$

and for $1 = s$ let denote,

$$\gamma^1_{p,q}(f) = \left(\int_{I\!R} \left(\frac{\|\tau_h f + \tau_{-h}(f) - 2f\|_p}{|h|} \right)^q \frac{dh}{|h|} \right)^{\frac{1}{q}}$$

(with the usual modification for $q = \infty$).

Definition 1 *For $0 < s \leq 1, 1 \leq p \leq \infty, 1 \leq q \leq \infty$, we define $B^s_{p,q}$ as :*

$$B^s_{p,q} = \{f \in L_p | \gamma^s_{p,q} < \infty\}, \text{ equipped with the norm } \|f\|_{B^s_{p,q}} = \|f\|_{L_p} + \gamma^s_{p,q}(f).$$

For $s > 1$ consider the decomposition $s = n + s_0$ with $n \in I\!N$ and $0 < s_0 \leq 1$.
We define $B^s_{p,q}$ as :

$$B^s_{p,q} = \{f | f^{(m)} \in B^{s_0}_{p,q}, \forall m \leq n\} : \text{equipped with the norm } \|f\|_{B^s_{p,q}} = \|f\|_p + \sum_{m \leq n} \gamma^{s_0}_{p,q}(f^{(m)}).$$

We can also consider Besov spaces on the interval $]-K,K[$.

References

[AHJP] Anderson, L., Hall, N., Jawerth, B., Peters, G., Wavelets on Closed Subsets of the Real Line, Manuscript (1993).

[BC] Benhenni, K., Cambanis, S., Sampling designs for estimating integrals of stochastic processes, Ann. Statist. 20 (1993), 161-194.

[BGJ] Bichteler, K., Gravereaux, J.B., Jacod, J., Malliavin Calculus for Processes with Jumps, Stochastics Monographs 2, ed. Mark Davis (1987).

[Bo] Bosq, D., Optimal sampling for density estimation, Personal communication (1994).

[CL] Castellana, J.V., Leadbetter, M.R., On smoothed probability density estimation for stationary processes, Stoch. Proc. Appl. 21 (1986), 179-193.

[Ca] Castellana, J.V., Integrated consistency of smoothed density estimators, Stoch. Proc. Appl. 33 (1989), 335-346.

[CP] Cheze-Payaud, N., Régression, Prédiction et Discrétisation des Processus à Temps Continu, Thèse de doctorat de l'Université Paris 6 (1994).

[CDJV] Cohen, A., Daubechies, I., Jawerth, B., Vial, P., Multiresolution analysis, wavelets and fast algorithms on an interval, C.R. Acad. Sci. Paris 316 (1992), 417-421.

[Da] Daubechies, I., Orthogonal bases of compactly supported wavelets, Comm. Pure Appl.Math. 41 (1988), 909-996.

[DJ] Delyon, B., Juditsky, A. : On the Computation of Wavelet Coefficients IRISA Technical Report 856, (1994).

[DoJo] Donoho, D., Johnstone, I., TeachWave V 0.550, `wavelab@playfair.stanford.edu`, (1993).

[IL] Istas, J., Laredo, C., Estimating Functionals of a Stochastic Process, Manuscript (1994).

[JL] Juditsky, A., Leblanc, F., Computing Wavelet Density Estimators for Stochastic Processes, IRISA Technical Report 895 (1994).

[Ku] Kutoyants, Yu. A., Parameter Estimation for Stochastic Processes, R& E Series 6, Heldermann Verlag, Berlin (1984).

[Le1] Leblanc, F., Estimation of the Marginal Density of a Continuous Time Stochastic Process by Wavelets and an Application to Diffusion Processes, CREST Document de travail 9456 (1994).

[Le2] Leblanc, F., Wavelet linear density estimator for a discrete time stochastic process: L_p-losses, (Submitted for publication).

[Ma] Masry, E., Probability density estimation from sampled data, IEEE Trans. Inform. Theory, 29 (1983), 696-709.

[PR] Prakasa Rao, B. L. S., Nonparametric density estimation for stochastic processes from sampled data, Publ. Inst. Stat. Univ. Paris, V XXXV, 3, p. 51-83 (1990).

[Tr] Triebel, H., Theory of Function Spaces II, Birkhauser Verlag, Berlin, (1992).

[Ve1] Veretennikov, A., Bounds for the mixing rate in the theory of stochastic equations, Theor. Probab. and App. 32, 2 (1988), 273-281.

[Ve2] Veretennikov, A., On hypoellipticity conditions and estimates of the mixing rate for stochastic differential equations, Soviet Math. Dokl. 40 (1990), 94-97.

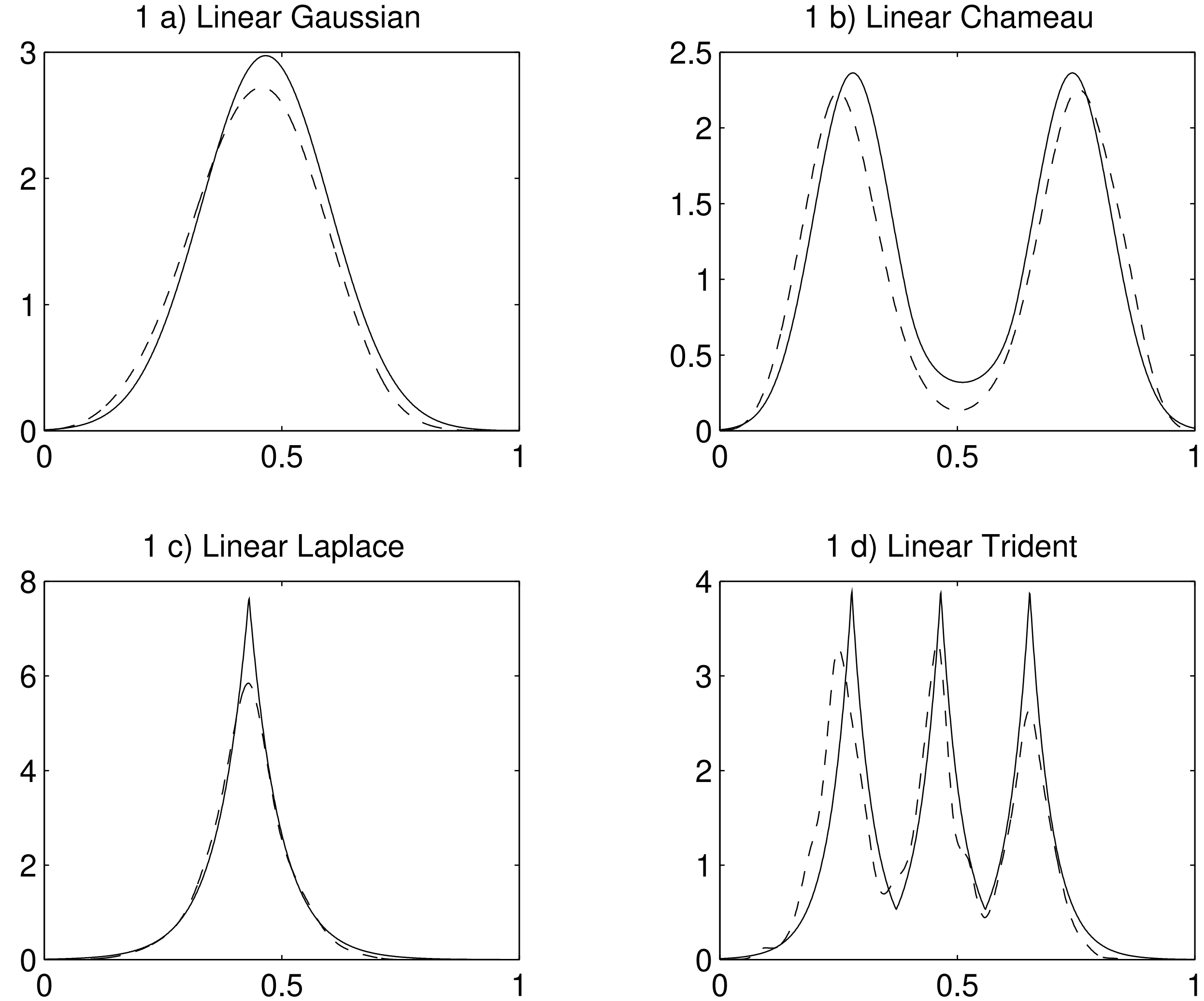

Figure 1: Wild thresholding adaptive algorithm. Continuous time. $T = 2000$, $\delta = 0.02$ and $n = 5000$.

Density type	L_2-error
Gaussian	0.15
Chameau	0.24
Laplace	0.23
Trident	0.40

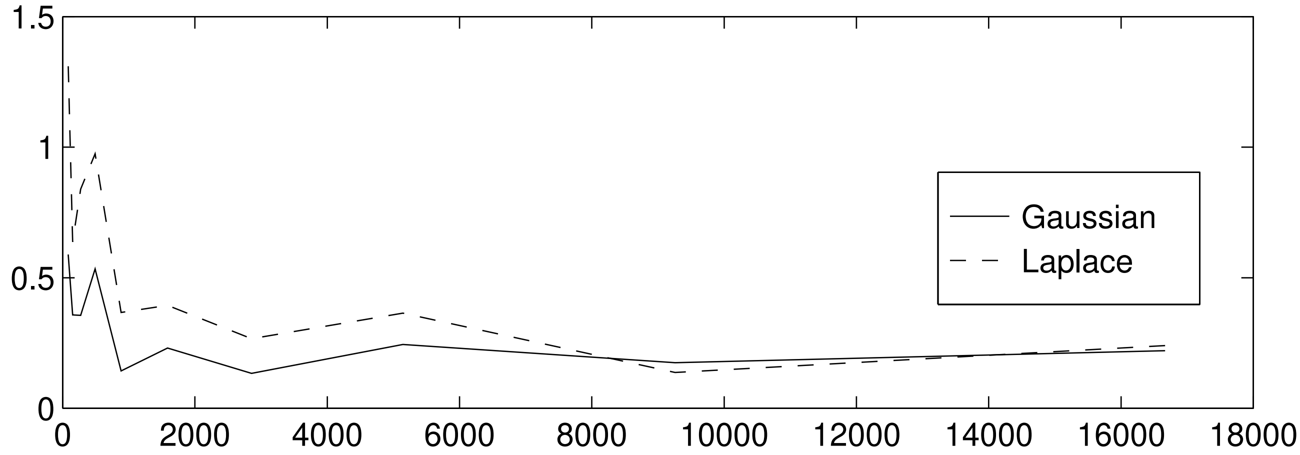

Figure 2: Wild thresholding ("linear") adaptive algorithm. L_2-error as function of n. $\delta = 0.01$, $T = 300$. The smallest $n = 1500$, the largest $n = 30000$.

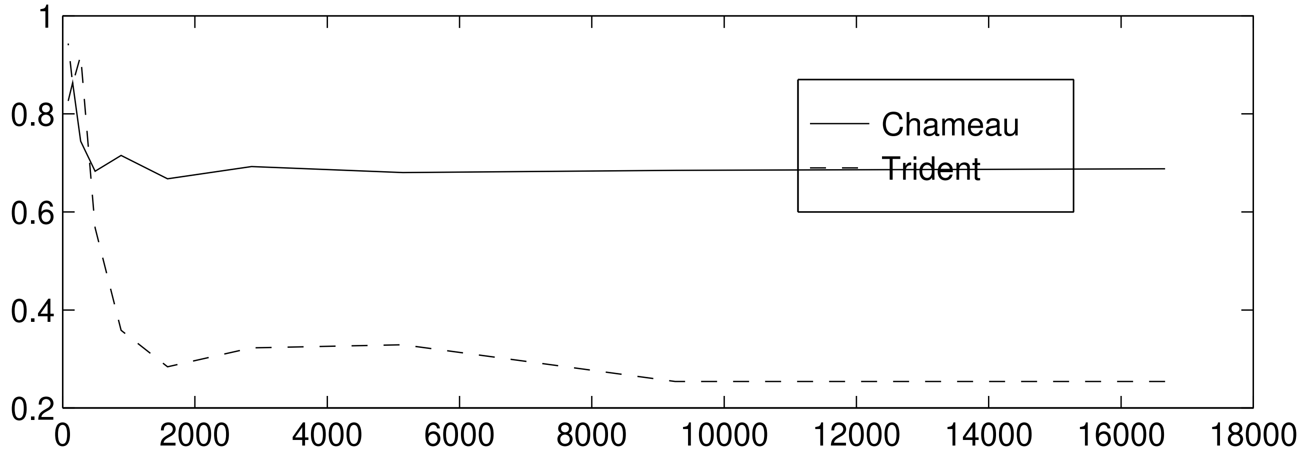

Figure 3: Wild thresholding ("linear") adaptive algorithm. L_2-error as function of N. $\delta = 0.01$, $T = 300$. The smallest $n = 1500$, the largest $n = 30000$.

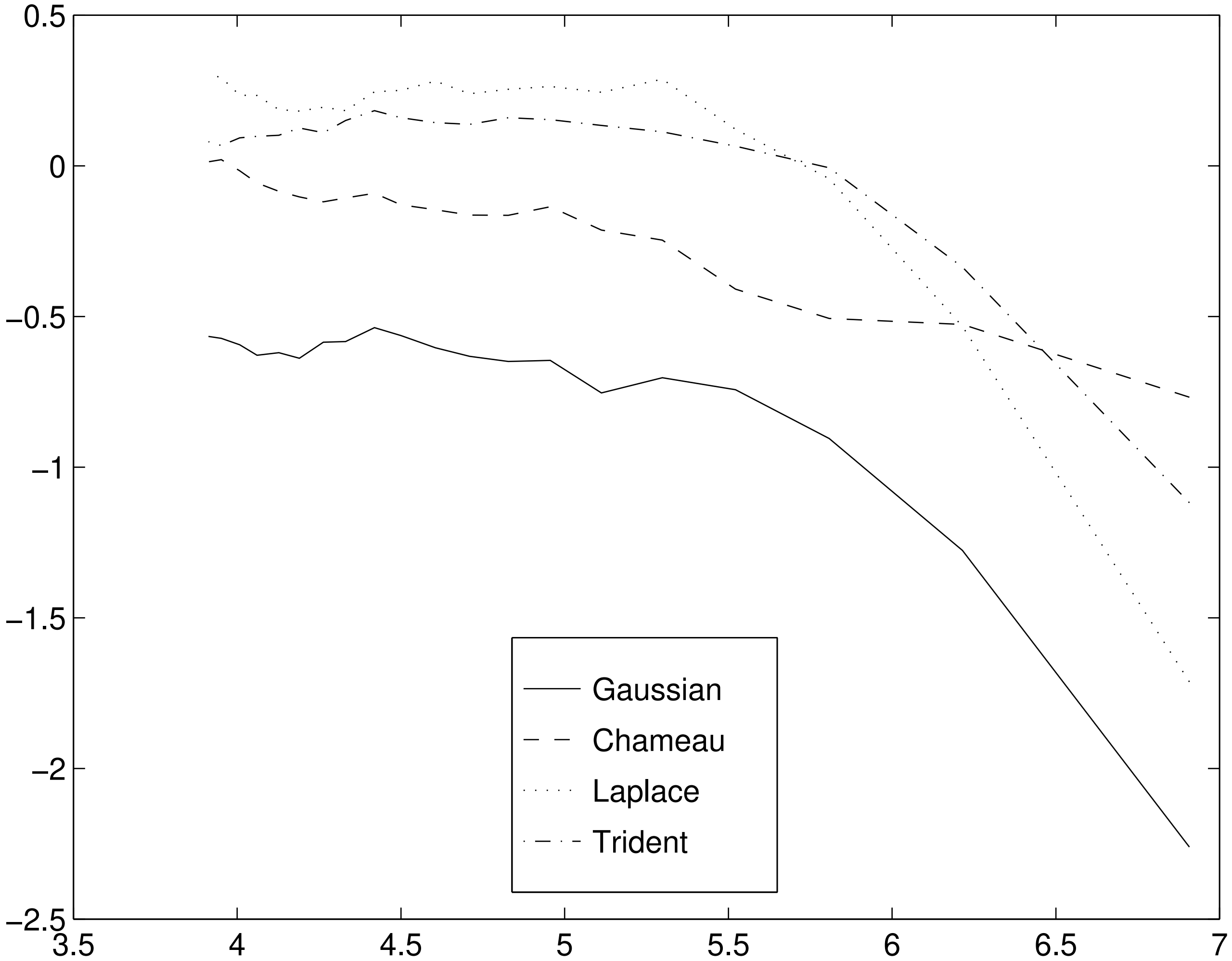

Figure 4: Wild thresholding ("linear") adaptive algorithm. L_2-error as function of T. $T = 50...1000$, $\delta = 0.01$. Averaged over 10 realizations. *Logarithmic axes.*