# On-line estimation of a smooth regression function

Liptser, R.
jointly with L. Goldentyer
Tel Aviv University
Dept. of Electrical Engineering-Systems

December 19, 2002

## SETTING

We consider a tracking problem for smooth function

$$f = f(t),\ 0 \le t \le T,$$

under observation

$$X_{in} = f(t_{in}) + \sigma\xi_i,\ t_{in} = \frac{i}{n}\ \boxed{\text{n is large}};$$

- $(\xi_i)$ is i.i.d., $E\xi_i = 0$, $E\xi_i^2 = 1$;

- $\sigma^2$ is a positive constant.

Without additional assumptions on $f$, it is difficult to create an estimator even $n$ is large.

## Main assumption

$f$ is differentiable $k$-times and the oldest derivative is Lipschitz continuous.

**Filtering approach:** Bar-Shalom and Li

Simulate $f^{(k)}(t)$ with a help WHITE NOISE:

$$\frac{df^{(k)}(t)}{dt} = \text{“white noise”}.$$

it sounds as nonsense but works pretty good.

**Nonparametric Statistic Approach.**

$$f \in \Sigma(k, \alpha, L)$$

The Stone-Ibragimov-Khasminskii class containing $k$-times differentiable function with

$$\left| f^{(k)}(t'') - f^{(k)}(t') \right| \leq L|t'' - t'|^{\alpha}, \quad 0 < \alpha \leq 1.$$

## Task: to combine both approaches

Since a quality of estimating depends on $n$, any estimate of $f$ is marked by $n$, that is $\widehat{f_n^{(j)}}(t)$ are estimates of $f^{(j)}(t)$, $j = 0, 1, \ldots, k$ respectively.

It is known from Ibragimov and Khasminskii that for a wide class of loss

$$\sup_{f \in \Sigma(k,\alpha,L)} E\mathcal{L}\left(n^{\frac{k+\alpha-j}{2(k+\alpha)+1}} \|\widehat{f_n^{(j)}} - f^{(j)}\|_{L_p}\right) < C.$$

and

$$n^{\frac{k+\alpha-j}{2(k+\alpha)+1}}, \;\; j = 0, 1 \ldots, k$$

is the best rate, uniformly in the class, of estimating risk convergence to zero in

$$n \to \infty.$$

In particular, the risks

$$E\Big(\widehat{f_n^{(j)}}(t) - f^{(j)}(t)\Big)^2,\ \ j = 0, 1 \ldots, k$$

have the same rates in $n$:

$$\sup_{f\in\Sigma(k,\alpha,L)} \overline{\lim_n}\, E\left(n^{\frac{k+\alpha-j}{2(k+\alpha)+1}}|\widehat{f_n^{(j)}}(t) - f^{(j)}(t)|\right)^2 < C.$$

These rates cannot be exceeded uniformly on any nonempty open set from $(0, T)$.

Jointly with Khasminskii, we realize on-line filter guaranteeing the optimal rates in $n$.

Here $t_{in}$ and $\widehat{f_n^{(j)}}(t_{in})$ identify $t_i$ and $\widehat{f_n^{(j)}}(t_i)$.

For $j = 0, 1 \ldots, k-1$,

$$\begin{aligned}\widehat{f^{(j)}}(t_i) = \widehat{f^{(j)}}(t_{i-1}) + \\ + \frac{1}{n}\widehat{f^{(j+1)}}(t_{i-1}) \\ + \frac{q_j}{n^{\frac{(2(k+\alpha)-j)}{2(k+\alpha)+1}}}\Big(X_i - \widehat{f^{(0)}}(t_{i-1})\Big)\end{aligned}$$

and for $j = k$

$$\widehat{f^{(k)}}(t_i) = \widehat{f^{(k)}}(t_{i-1}) + \frac{q_k}{n^{\frac{(2(k+\alpha)-k)}{2(k+\alpha)+1}}}\Big(X_i - \widehat{f^{(0)}}(t_{i-1})\Big).$$

The vector $q$ with entries $q_0, \ldots, q_k$ has to be chosen such that all roots of characteristic polynomial

$$p^k(u,q) = u^{k+1} + q_0 u^k + q_1 u^{k-1} + \ldots + q_{k-1}u + q_k$$

are different and have negative real parts.

## Two problems

1. Choice of an appropriate initial conditions:

$$\widehat{f^{(0)}}(0), \widehat{f^{(1)}}(0), \ldots, \widehat{f^{(k)}}(0)$$

to minimize a boundary layer.

2. Choice of the vector $q$ such that the assumption about roots of the polynomial $p^k(u,q)$ remains valid and

$$C(q) \geq \sup_{f \in \Sigma(k,\alpha,L)} E\left( n^{\frac{k+\alpha-j}{2(k+\alpha)+1}} |\widehat{f_n^{(j)}}(t) - f^{(j)}(t)| \right)^2$$

is smallest as possible.

To manage these problems we need to restrict ourselves by

$$\alpha = 1.$$

## Boundary layer

The left side boundary layer

$$c(q)n^{-\frac{1}{2\beta+1}}\log n$$

where the optimal rates in $n$ might be lost is inevitable. This boundary layer is due to on-line limitations of the above tracking system.

One can readily suggest an off-line modification with the same recursion in the backward time subject to some boundary conditions independent of observation $X_i$'s. This modification obeys the right side boundary layer.

So, a combination of the forward and backward time tracking algorithms allows support the optimal rate in $n$ for $[0, T]$.

## Suitable choice of $\boxed{q}$

Vector $q$ should satisfy multiple requirements regarding

- $C(q)$ the upper bound for the normalized risk;

- $c(q)$ the parameter of the boundary layer;

- roots of polynomial $p^k(u, q)$.

These requirements might contradict each other.

## Example 1, $\boxed{\Sigma(0,1,L)}$

The worst $f(t) = f(0) \pm Lt$. Applying the Arzela-Ascoli theorem we find that

$$C(q) = \frac{\sigma q}{2} + \frac{L^2\sigma^2}{q^2}$$

and

$$q^\circ := \operatorname*{argmin}_{g>0} C(q) = (2L)^{2/3}\sigma^{1/3}.$$

Hence, a reasonable estimator is:

$$\widehat{f}(t_i) = \widehat{f}(t_{i-1}) + \left(\frac{2L}{n\sigma}\right)^{2/3}(X_i - \widehat{f}(t_{i-1})).$$

## General case, $\boxed{\Sigma(k > 0, 1, L)}$

With the worst $f(t)$ such that

$$f^{(k)}(t) = f^{(k)}(0) \pm Lt,$$

applying the Arzela-Ascoli theorem we find

$$C(q) = \text{trace}\Big(P(q) + M(g)M^*(q)\Big)$$

where

$$M(q) = L(a - qA)^{-1}b$$

and $P(q)$ solves the Lyapunov equation

$$(a - qA)P(q) + P(q)(a - qA)^* + \sigma^2 qq^* = 0.$$

Here,

$$a = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{(k+1)\times(k+1)},$$

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}_{1\times(k+1)}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}_{(k+1)\times 1}.$$

## Conditional minimization

A direct minimization of $C(q)$ is useless. A computer implementation is heavy enough. Even if

$$q^{\circ} = \operatorname*{argmin}_{g} C(q)$$

is found, the main requirement, expressed in term of eigenvalues the polynomial $p^k(u, q)$, might be not satisfied (numerical computations show that).

So, some kind of a conditional minimization procedure in vector $q$ is desirable.

The main tool for such minimization is

| adaptation to Kalman filter design. |
|---|

# Kalman filter design

In a frame of Bar Shalom idea, set

$$f^{(k)}(t_i) = f^{(k)}(t_{i-1}) + n^{-\frac{k+2}{2(k+1)+1}} \boxed{\gamma} \eta_i$$
$$X_i = f^{(0)}(t_{i-1}) + \sigma \xi_i$$

where $(\eta_i)$ is a white noise, independent of $(\xi_i)$, with $E\eta_1 = 0$, $E\eta_1 = 1$;

$\boxed{\gamma \text{ is free parameter.}}$

For any $\gamma \neq 0$, the Kalman filter possesses an asymptotic form in $n \to \infty$ and, being applied to the original function $f(t)$, guaranties the optimal rate in $n \to \infty$ for the estimation risk. In other words, that Kalman filter coincides with our proposed filter.

The remarkable fact is that $\boxed{q = q(\gamma)}$ and for any positive $\gamma$ roots of polynomial $p^k(u, q(\gamma))$ are different and have negative real parts.

Thus,

$$q(\gamma) = \frac{Q(\gamma)A^*}{\sigma^2}$$

with $Q(\gamma)$ being solution of the algebraic Riccati equation

$$aQ(\gamma) + Q(\gamma)a^* + \gamma^2 bb^* - \frac{Q(\gamma)A^*AQ(\gamma)}{\sigma^2} = 0.$$

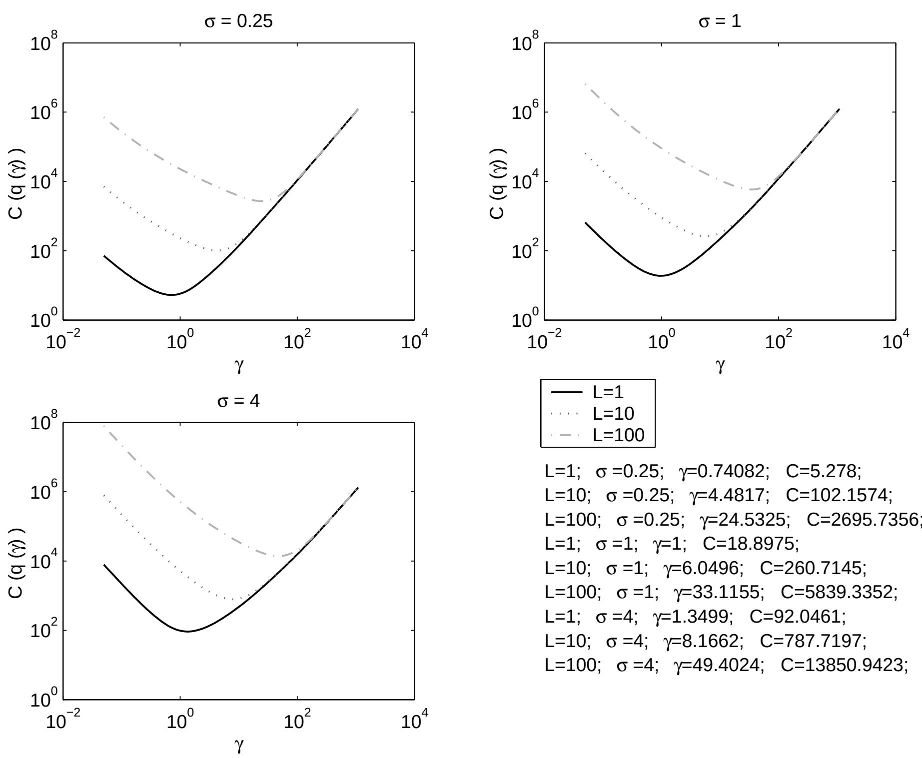
obeying the unique positive-definite solution since block-matrices

$$G_1 = \begin{pmatrix} A \\ Aa \\ \vdots \\ Aa^k \end{pmatrix} \quad \text{and} \quad G_2 = \begin{pmatrix} bb^* & abb^* & \dots & a^k bb^* \end{pmatrix}$$

are of full ranks (so called, observability and controllability conditions).

# $C(q(\gamma))$-minimization

We reduce the minimization of $C(q)$ with respect to vector $q$ to minimization of $C(q(\gamma)$ with respect to a positive parameter $\gamma$.




Here, $C\Big(q(\gamma)\Big)$ in logarithmic scale for $k = 2$ and various $L$ and $\sigma$.

## Explicit minimization procedure

Entries of $Q(\gamma)$ obey the following presentation

$$Q_{ij}(\gamma,\sigma) = U_{ij}\sigma^2\Big(\frac{\gamma}{\sigma}\Big)^{\frac{i+j+1}{k+1}},\ i,j = 0,1,\ldots,k,$$

where $U_{ij}$ are entries of the matrix $U$ also being solution of the algebraic Riccati equation free of $\sigma$ and $\gamma$:

$$aU + Ua^* + bb^* - UA^*AU = 0.$$

We have

$$q_0(\gamma) = U_{00}\Big(\frac{\gamma}{\sigma}\Big)^{1/k+1}$$
$$q_1(\gamma) = U_{01}\Big(\frac{\gamma}{\sigma}\Big)^{2/k+1}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$q_k(\gamma) = U_{0k}\Big(\frac{\gamma}{\sigma}\Big).$$

For $k = 0, \ldots, 4$

| k | $U_{00}$ | $U_{01}$ | $U_{02}$ | $U_{03}$ | $U_{04}$ |
|---|---|---|---|---|---|
| 0 | 1 | NA | NA | NA | NA |
| 1 | $\sqrt{2}$ | 1 | NA | NA | NA |
| 2 | 2 | 2 | 1 | NA | NA |
| 3 | $\sqrt{4+\sqrt{8}}$ | $2+\sqrt{2}$ | $\sqrt{4+\sqrt{8}}$ | 1 | NA |
| 4 | $1+\sqrt{5}$ | $3+\sqrt{5}$ | $3+\sqrt{5}$ | $1+\sqrt{5}$ | 1 |

## Roots of $p^k(u,q)$

$$k = 0: \quad -\left(\frac{\gamma}{\sigma}\right)$$

$$k = 1: \quad -\left(\frac{\gamma}{\sigma}\right)^{1/2}\left(\frac{1}{\sqrt{2}} \pm i\frac{1}{\sqrt{2}}\right)$$

$$k = 2: \quad -\left(\frac{\gamma}{\sigma}\right)^{1/3}\left(1;\ \frac{1}{2} \pm i\frac{\sqrt{3}}{2}\right)$$

$$k = 3: \quad -\left(\frac{\gamma}{\sigma}\right)^{1/4}\Big(0.924 \pm i0.383;\quad 0.383 \pm i0.924\Big)$$

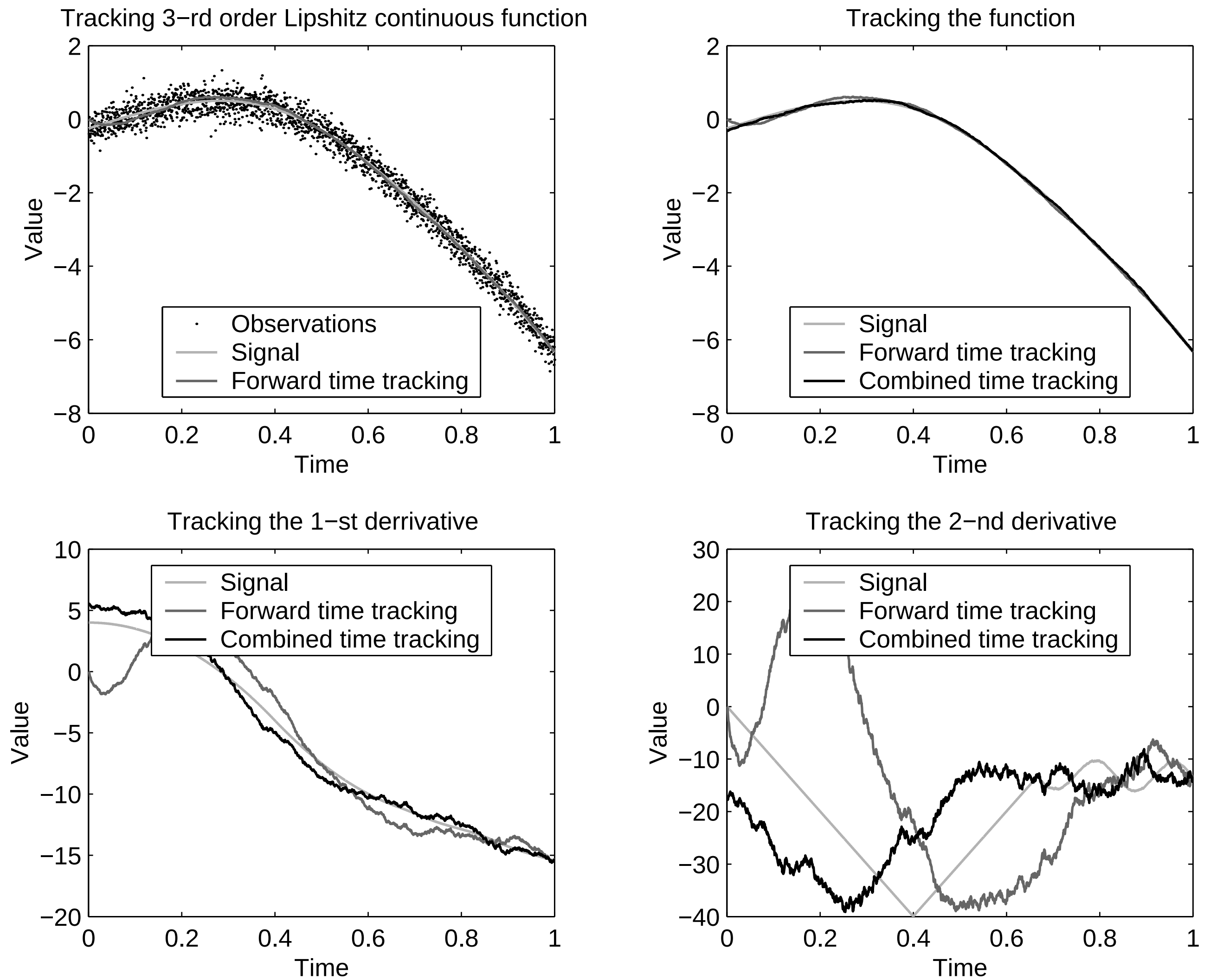
$$k = 4: \quad -\left(\frac{\gamma}{\sigma}\right)^{1/5}\Big(1;\ 0.809 \pm i0.588;\quad 0.309 \pm i0.951\Big).$$

# Example 2

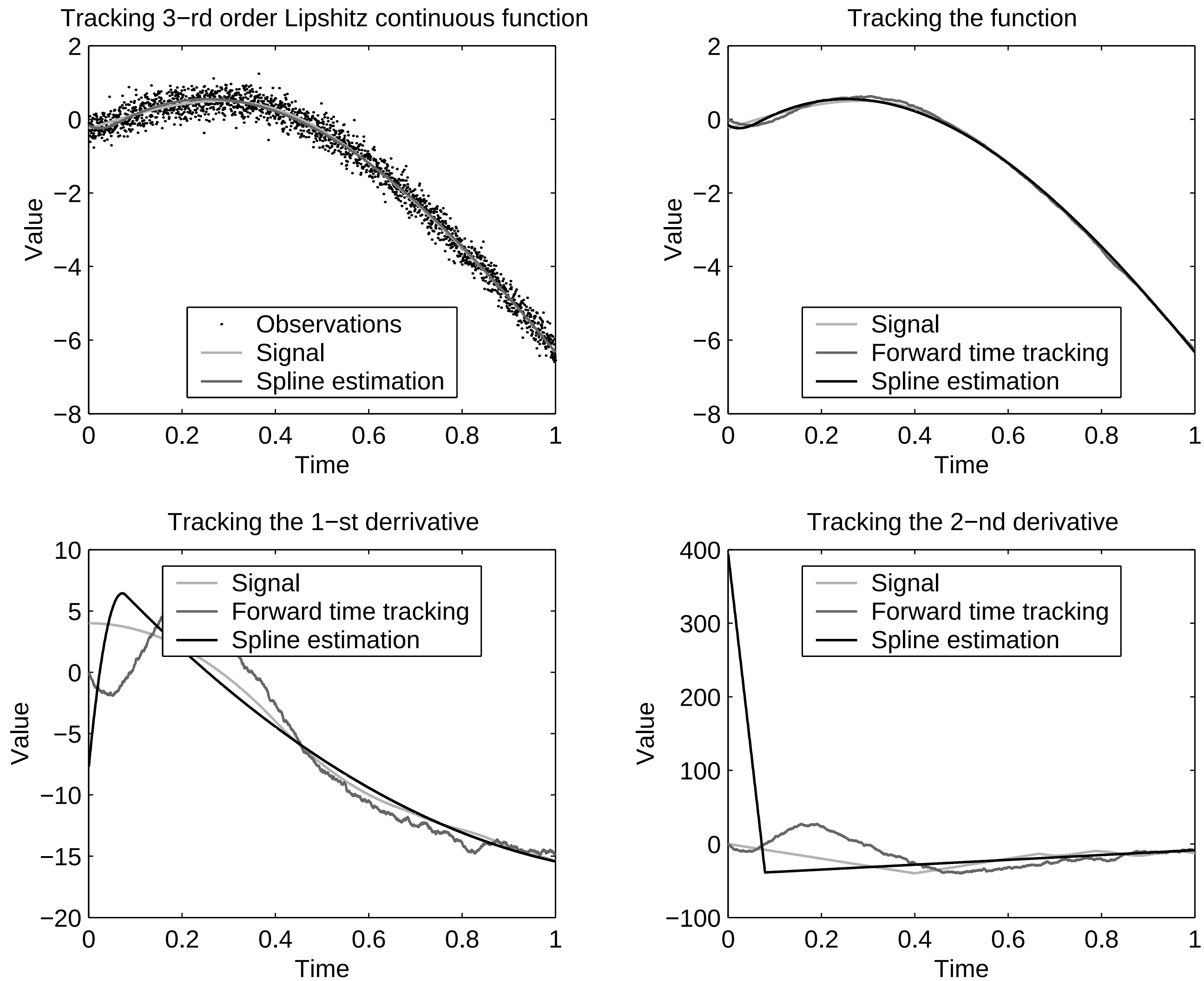
$k = 2$, $L = 100$, $\sigma = 0.25$.

$$\begin{aligned}
\widehat{f^{(0)}}(t_i) &= \widehat{f^{(0)}}(t_{i-1}) + \frac{1}{n}\widehat{f^{(1)}}(t_{i-1}) \\
&\quad + \frac{9.225}{n^{6/7}}\Big(X_i - \widehat{f^{(0)}}(t_{i-1})\Big) \\
\widehat{f^{(1)}}(t_i) &= \widehat{f^{(1)}}(t_{i-1}) + \frac{1}{n}\widehat{f^{(2)}}(t_{i-1}) \\
&\quad + \frac{42.550}{n^{5/7}}\Big(X_i - \widehat{f^{(0)}}(t_{i-1})\Big) \\
\widehat{f^{(2)}}(t_i) &= \widehat{f^{(2)}}(t_{i-1}) \\
&\quad + \frac{98.132}{n^{4/7}}\Big(X_i - \widehat{f^{(0)}}(t_{i-1})\Big).
\end{aligned}$$

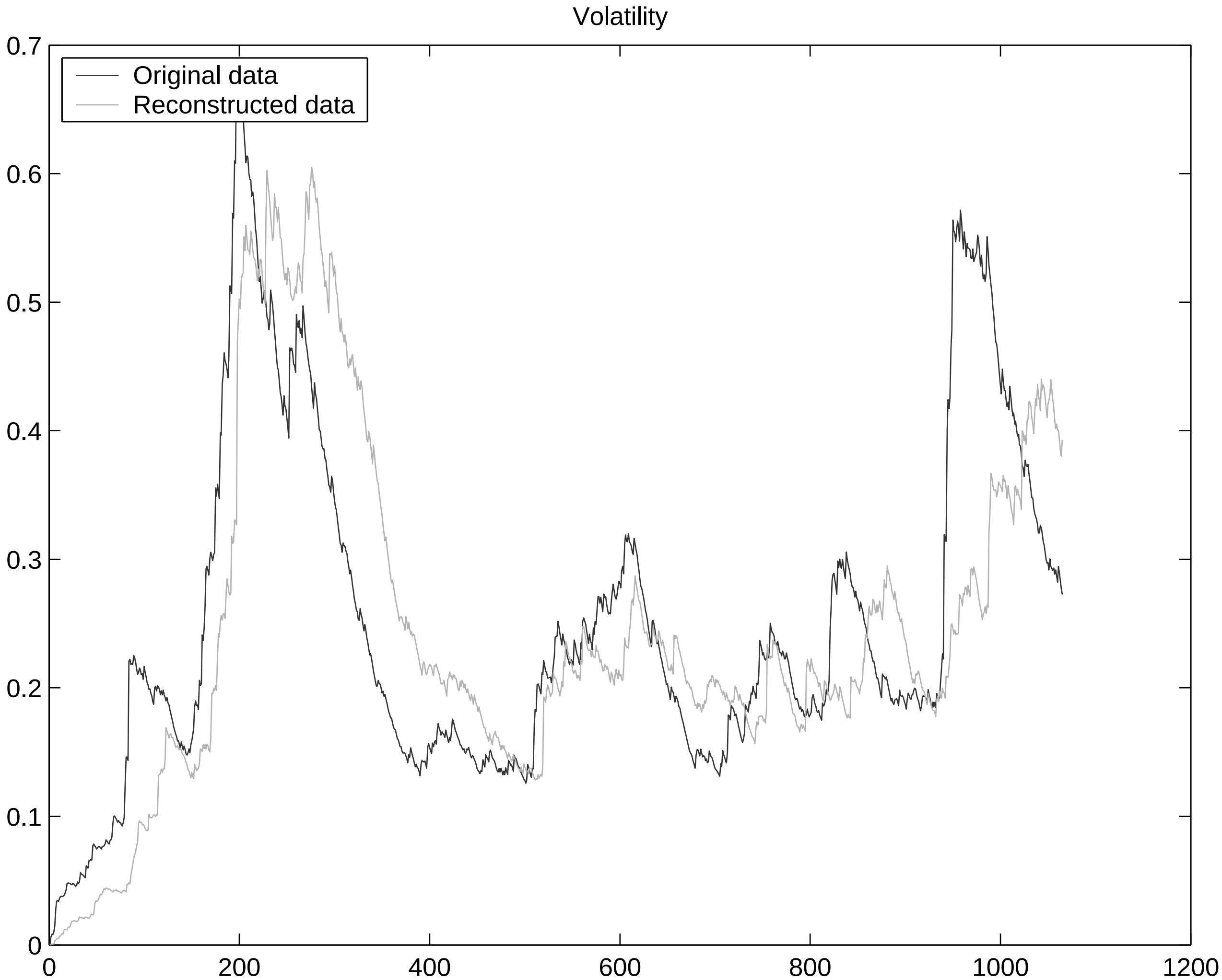
# Forward and backward time tracking with $n = 2000$



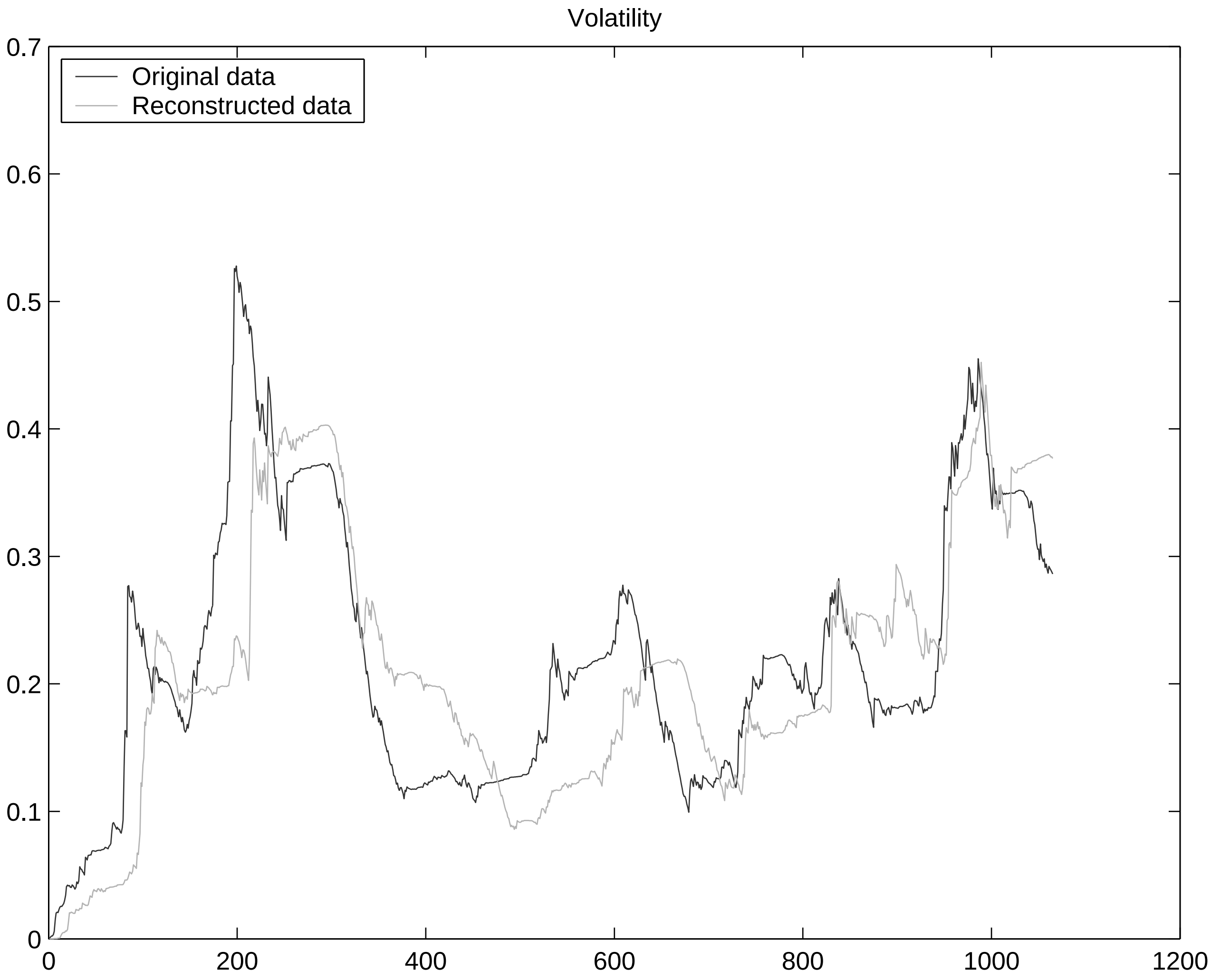
# Comparison with spline $n = 2000$



# Application to Finance




Volatility is assumed to be Lipschitz continuous functions.




Tracking with adaptation.