AN ALGORITHM FOR TRACKING A RANDOM WALK WITH UNKNOWN DRIFT

Jean Luc LE CALVEZ †, *Bernard DELYON* ‡, *Anatoli JUDITSKY* *

† IRISA, ‡ IRISA/INRIA
Campus de Beaulieu, 35042 Rennes Cedex, FRANCE
∗ INRIA Rhône-Alpes
655 avenue de l'Europe, 38330 MONTBONNOT SAINT MARTIN, FRANCE
{jlecalve,delyon}@irisa.fr, Anatoli.Iouditski@inrialpes.fr

ABSTRACT

In this paper we study the problem of tracking a random walk observed with noise when the variance of the walk increment is unknown. We describe a sequence of estimators of the random walk and we design an algorithm to choose the best estimator among all the sequence. We give also a bound for the mean square error of this estimator. Finally some simulations are presented and we compare our algorithm with the Kalman filter when the variance of the walk increment is estimated.

1. INTRODUCTION

Tracking the possibly time-varying dynamic of a stochastic system is a major objective in signal processing and system identification. A lot of work about this subject has given numerous articles. Nevertheless, we can cite [6] for a survey in system identification, [1] about the design of adaptive algorithms, [3] about asymptotical study and [5] about stability of algorithms.

In this paper, we want to track a d-dimensional random walk with noisy observations of this walk

$$\theta_{t+1} = \theta_t + \gamma w_t \quad (1)$$

$$y_t = \theta_t + e_t \quad (2)$$

where γ is an scalar, (w_t) and (e_t) are stationary white noise processes, mutually independent, with covariance matrices respectively equal to I and $\sigma_e^2 I$. The parameter γ is assumed to be unknown.

This model is also well suited when the observations are a noisy linear regression

$$y_t = \varphi_t^T \theta_t + e_t.$$

In this case, we add a step of recursive estimation with a large gain μ

$$z_t = z_{t-1} + \mu(y_t - \varphi_t^T z_{t-1})\varphi_t$$

and then we consider that (z_t) are the noisy observations in the model (1)-(2).

Under Gaussian assumptions on noises, the solution is obtained from the Kalman filter (See, e.g., [4]). But we must note that the Kalman filter requires the knowledge of the covariance matrices $\gamma^2 I$ and $\sigma_e^2 I$.

As the parameter γ is unknown, we can't apply the Kalman filter. A natural idea is then to estimate γ with an online algorithm which minimizes $E(y_t - \hat{\theta}_t)^2$ (see section 4.4 of [2]).

We design in this paper a tracking algorithm for the model (1)-(2) when the parameter γ is unknown and is not estimated.

2. BIAS VS NOISE

At each time t, we consider the sequence of estimators defined on a window of length $m_i = 2^i m_0$, $(m_0 \geq 1)$ by

$$\hat{\theta}_t^{(i)} = \frac{1}{m_i} \sum_{k=t-m_i+1}^{t} y_k, \quad i = 0, \cdots, [\log_2 \frac{t}{m_0}]. \quad (3)$$

Then we want to find the estimator which minimizes the Mean Square Error

$$E\left\|\hat{\theta}_t^{(i)} - \theta_t\right\|_2^2$$

where $\left\|x\right\|_2$ is the Euclidean norm of x. That is to say we want to find the length of the optimal window. Let us remark that the estimators (3) can be written as

$$\hat{\theta}_t^{(i)} = \theta_t + \xi_t^{(i)} + \zeta_t^{(i)}$$

where

$$\xi_t^{(i)} = \frac{1}{m_i} \sum_{k=t-m_i+1}^{t} e_k$$

$$\zeta_t^{(i)} = \frac{1}{m_i} \sum_{k=t-m_i+1}^{t} (\theta_k - \theta_t).$$

As (w_t) and (e_t) are mutually independent, $(\xi_t^{(i)})$ and $(\zeta_t^{(i)})$ are also mutually independent for all $i = 0, \cdots, [\log_2 \frac{t}{m_0}]$. Then

$$E\|\hat{\theta}_t^{(i)} - \theta_t\|_2^2 = E\|\xi_t^{(i)}\|_2^2 + E\|\zeta_t^{(i)}\|_2^2.$$

As (e_t) is assumed to be a white noise with covariance matrix $\sigma_e^2 I$, we have for the first term,

$$\sigma_i^2 \triangleq E\|\xi_t^{(i)}\|_2^2 = \frac{\sigma_e^2 d}{m_i}.$$

For the second term, we note from (1) that

$$\theta_t = \theta_k + \gamma \sum_{j=0}^{t-k-1} w_{t-j},$$

hence

$$\begin{aligned} b_i^2 \triangleq E\|\zeta_t^{(i)}\|_2^2 &= \frac{\gamma^2}{m_i^2} E\| \sum_{k=t-m_i+1}^{t-1} \sum_{j=0}^{t-k-1} w_{t-j}\|_2^2 \\ &= \frac{\gamma^2}{m_i^2} E\| \sum_{j=0}^{m_i-1} (m_i - 1 - j) w_{t-j}\|_2^2 \\ &= \frac{\gamma^2 d}{6} \frac{(m_i-1)(2m_i-1)}{m_i}. \end{aligned}$$

As the mapping $i \to m_i$ is increasing and $m_i \geq 1$ for all $i = 0, \cdots, [\log_2 \frac{t}{m_0}]$, σ_i^2 is a decreasing function of m_i and b_i^2 an increasing function. We call them respectively the "noise" and the "bias" of the estimator.

Thus the optimal window is the window which balances σ_i and b_i. We obtain that the optimal length m^* verifies the equivalence

$$m^* \sim \max(1, \frac{\sigma_e}{\gamma}). \tag{4}$$

In other words, if we had no noise on observations (i.e. $\sigma_e = 0$), the optimal estimator would be the current observation. On the contrary, if the parameter θ_t is time invariant (i.e. $\gamma = 0$ and $\theta_t = \theta$), the optimal window would be an infinite window.

As the parameter γ is unknown, a direct and simple computation of m^* from (4) is not available. We describe in the next section an algorithm which computes this optimal length without a priori knowledge on the value of the parameter γ.

3. THE ALGORITHM

Consider the confidence region of $\hat{\theta}_t^{(r)}$ with area equal to $2K\sigma_r$ and K a constant suitably chosen. We explain now the heuristic of the algorithm. We remind that the variance of the the estimator is decreasing, and the bias increasing, with the length of the window.

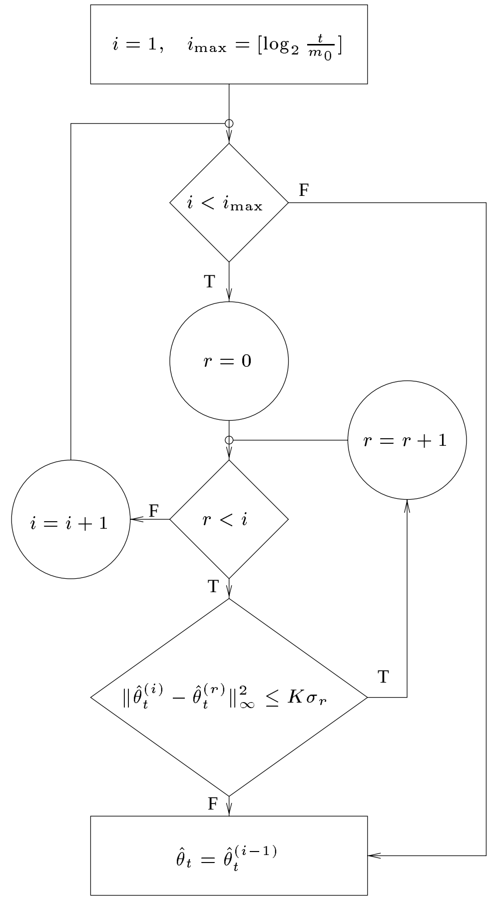

Figure 1: Computation of the estimator $\hat{\theta}_t$. The figure must be read from top to bottom. Diamond boxes represent a boolean test with one input and two outputs (True or False). Circles represent affectation.

An estimator $\hat{\theta}_t^{(r)}$, defined on a small window (that is to say r small), has a small "bias". Hence its region is expected to contain the parameter θ_t.

On the other hand, if $\hat{\theta}_t^{(r)}$ is defined on a large window, its confidence region will be small but the bias will be large and maybe too large. Then the confidence region may not contain θ_t.

Having taken this remarks into consideration, if the estimators $\hat{\theta}_t^{(r)}$, from 0 to $i-1$, have been accepted then $\hat{\theta}_t^{(i)}$ will be accepted if it belongs to all the confidence regions from 0 to $i-1$. Otherwise $\hat{\theta}_t^{(i)}$ will be rejected. This algorithm is represented in the figure 1.

4. A BOUND FOR THE MEAN SQUARE ERROR

If $u = (u_1, \cdots, u_d)$ is a d-dimensional vector, we note the infinite norm of u,

$$\|u\|_\infty = \max_{1 \le i \le d} |u_i|.$$

We make the following assumptions

- $\sigma_0 \ge b_0$,
- $P\big(\|e_t\|_\infty \ge \lambda\big) \le P\big(|\mathcal{N}(0, \sigma_e^2)| \ge \lambda\big)$,
- $P\big(\|w_t\|_\infty \ge \lambda\big) \le P\big(|\mathcal{N}(0, 1)| \ge \lambda\big)$,
- $\|e_t\|_\infty$ and $\|w_t\|_\infty$ have symmetric density probability function.

We define the integer i^* by

$$i^* = \max\{0 \le i \le i_{\max} \,:\, b_i \le \sigma_i\}$$

Remark: As the σ_i's are decreasing and the b_i's are increasing, the first assumption is necessary for the definition of i^*, but it is not too restrictive in practice. A simple calculus gives that the parameter γ must verify the condition

$$\gamma^2 \le \frac{6\sigma_e^2}{(2m_0 - 1)(4m_0 - 1)}.$$

If this condition is not verified, it means that the variance of (e_t) is smaller than the variance of (w_t). In this case, the noise can be omitted from the model what simplifies the tracking problem.

Under these assumptions we can prove the

Theorem

There exists a constant C such that

$$\big(E\|\hat\theta_t - \theta_t\|_2^2\big)^{1/2} \le \big(E\|\hat\theta_t^{(i^*)} - \theta_t\|_2^2\big)^{1/2} + Kd\sigma_{i^*} + C\sigma_0\big(Kde^{-\frac{K^2d^2}{16}} i^*\big)^{1/2}$$

Sketch of the proof:

We note C_j the cube of length $2K\sigma_j$, centered in $\hat\theta_t^{(j)}$, and we note $I_i = \cap_{j=1}^{i} C_j$ the intersection of the cubes from 0 to i.

Hence $\hat\theta_t = \hat\theta_t^{(\hat\imath)}$ where

$$\hat\imath = \max\{i \,:\, \hat\theta_t^{(i)} \in I_i\}.$$

Then we divide the mean square error into two parts

$$E\|\hat\theta_t - \theta_t\|_2^2 = E1_{\hat\imath \ge i^*}\|\hat\theta_t - \theta_t\|_2^2 + E1_{\hat\imath < i^*}\|\hat\theta_t - \theta_t\|_2^2$$

We remark that for any d-dimensional vector u

$$\|u\|_\infty^2 \le \|u\|_2^2 \le d\|u\|_\infty^2.$$

To bound the first term we apply the

Lemma 1

If $\hat\imath \ge i^*$ *then* $\|\hat\theta_t - \theta_t\|_2 \le Kd\sigma_{i^*} + \|\hat\theta_t^{(i^*)} - \theta_t\|_2$

To prove this lemma, we just have to remark that $I_{\hat\imath} \subseteq I_{i^*}$ and then $\hat\theta_t \in C_{i^*}$. Hence we obtain that

$$\big(E1_{\hat\imath > i^*}\|\hat\theta_t - \theta_t\|_2^2\big)^{1/2} \le Kd\sigma_{i^*} + \big(E\|\hat\theta_t^{(i^*)} - \theta_t\|_2^2\big)^{1/2}. \quad (5)$$

For the second term, we use the following lemmas

Lemma 2 *If* $\hat\imath < i^*$ *then* $\sup_{i \le i^*} \|\xi_t^{(i)} + \zeta_t^{(i)}\|_\infty > \frac{K}{2}\sigma_i$,

The core of the prove is that there exists a integer j such that $\hat\theta_t^{(i)}$ does not belong to C_j.

Lemma 3

If (e_t) and (w_t) are two sequences, mutually independent, of independent random vectors, such that

- $\|e_t\|_\infty$ *and* $\|w_t\|_\infty$ *have symmetric density probability function and for all* $\lambda \ge 0$
- $P\big(\|e_k\|_\infty \ge \sigma_e\lambda\big) \le P\big(|\mathcal{N}| \ge \lambda\big)$,
- $P\big(\|w_k\|_\infty \ge \lambda\big) \le P\big(|\mathcal{N}| \ge \lambda\big)$

then for all $i \ge 0$

$$P\big(\|\xi_t^{(i)} + \zeta_t^{(i)}\|_\infty \ge \lambda\big) \le P\big(|\mathcal{N}| \ge \frac{\sqrt{d}}{\sqrt{2}\sigma_i}\lambda\big)$$

with $\mathcal{N} \sim \mathcal{N}(0,1)$.

With the independence of (e_t) and (w_t), we prove that

$$P\big(\|\xi_t^{(i)} + \zeta_t^{(i)}\|_\infty \ge \lambda\big) \le P\big(|\mathcal{N}| \ge \frac{\sqrt{d}}{\sqrt{b_i^2 + \sigma_i^2}}\lambda\big).$$

To conclude, we just remind that for all $i \le i^*$, $b_i^2 < \sigma_i^2$.

With these lemmas we can find a constant C such that

$$\big(E1_{\hat\imath \le i^*}\|\hat\theta_t - \theta_t\|_2^2\big)^{1/2} \le C\sigma_0\big(Kde^{-\frac{K^2d^2}{16}} i^*\big)^{1/2}. \quad (6)$$

By combining (5) and (6), we finish the prove of the theorem.

5. SIMULATIONS

In the simulations, (w_t) and (e_t) are 1-dimensional centered Gaussian white noise. The random walk is tracked on a sample of length 5000. In both figures, the dashed line is the tracking parameter. The last simulation, figure 5, is a comparison between our algorithm and the Kalman filter when γ is estimated with $\hat\gamma \ne \gamma$.

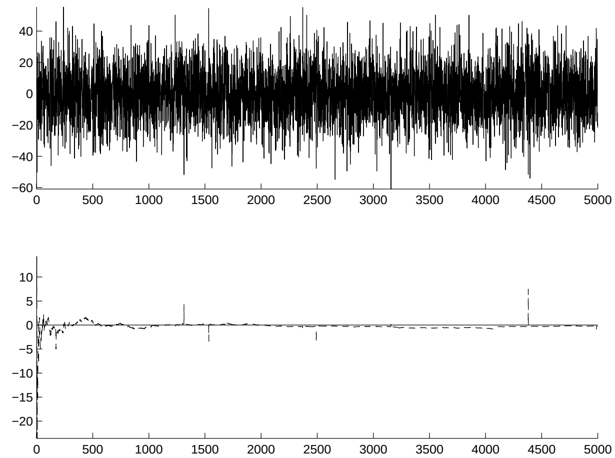

Figure 2: $\gamma = 0,\ \sigma_e = 16,\ K = \sqrt{\log(4 i_{\max} 2^{i_{max}})}$. The first plot is the random walk in the observations. The second plot is the random walk and the tracking parameter.

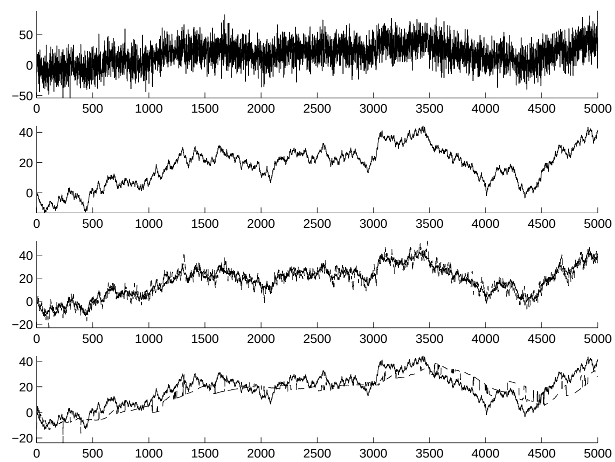

Figure 3: $\gamma = 0.7,\ \sigma_e = 16,\ K = \sqrt{\log(4 i_{\max} 2^{i_{max}})}$. The first plot is the random walk in noise. The second plot is the random walk. The third is the result of the Kalman filter with γ is estimated by $\hat{\gamma} = 2$. The fourth plot is the random walk and our tracking parameter.

6. CONCLUSION

In this paper we have designed an algorithm for tracking a d-dimensional random walk with unknown drift. The estimator is the mean of the noisy observations on a window of some length. Our algorithm approximate the optimal window which minimizes the mean square error. We have given also a bound for the mean square error which is close to the optimal bound. We have shown that our algorithm can be an alternative to the Kalman filter when γ is unknown.

7. REFERENCES

[1] A. Benveniste, “Design of Adaptive Algorithms for the Tracking of Time-varying Systems”, International Journal of Adaptive Control and Signal Processing, vol. 1, pp. 3-29, 1987.

[2] A. Benveniste, M. Métivier, P.Priouret, “Adaptive Algorithms and Stochastic Approximations”, Springer-Verlag, 1990.

[3] B. Delyon, A. Juditsky, “Asymptotical Study of Parameter Tracking Algorithms”, SIAM J. Control and Optimization, vol. 33, No 1, pp. 323-345, 1995.

[4] G.C. Goodwin, K.S. Sin, “Adaptive Filtering Prediction and Control”, Prentice-Hall, 1984.

[5] L. Guo, “Stability of Recursive Stochastic Tracking Algorithms” SIAM J. Control and Optimization, vol. 32, No 5, pp. 1195-1225, 1994.

[6] L. Ljung, S. Gunnarsson, “Adaptation and Tracking in System Identification-A Survey” Automatica, vol. 26, No 1, pp. 7-21, 1990.