

ADAPTIVE SIGNAL PROCESSING BY PARTICLE FILTERS AND DISCOUNTING OF OLD MEASUREMENTS

Petar M. Djurić[†], Jayesh Kotecha[†], Jean-Yves Tournet^{}, Stéphane Lesage^{*}*

[†]Department of Electrical and Computer Engineering, SUNY at Stony Brook, NY 11794, USA

^{*} ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France

ABSTRACT

In adaptive signal processing the principle of exponentially weighted recursive least-squares plays a major role in developing various estimation algorithms. It is based on the concept of discounting of old measurements and allows for better performance in problems with time-varying signals and signals in nonstationary noise. In this paper we show how this concept can be combined with the Bayesian methodology. We propose that the discounting of old measurements within the Bayesian framework be implemented by employing particle filters. The main idea is presented by way of a simple example. The methodology is very attractive and can be used in a very wide range of scenarios including ones that involve highly nonlinear models and non-Gaussian noise.

1. INTRODUCTION

Adaptive signal processing is a very important part of statistical signal processing and has applications in diverse areas including communications, controls, radar, sonar, and biomedical engineering. A wide variety of signal processing problems involve nonstationary signals or time-varying models, and the standard approach to resolving them involves application of adaptive filters. The number of applications where they have been successfully employed is very large, and examples of it such as linear prediction, channel equalization, beamforming, interference cancellation, and system identification abound in many standard textbooks [1], [2], [3].

A big class of adaptive filtering methods is based on the principle of recursive least-squares (RLS). In applications where the signals have time-varying parameters or the noise is nonstationary, and the time-varying nature of the unknowns is not known, the RLS algorithm is modified to exponentially weighted RLS. The objective of the exponential weighting is to give more weight to more recent than to older observations. Here we show how this idea can be used within the framework of Bayesian methodology. It is shown that if the equation that represents the model of the observations is extended with a state equation in the form of a random walk model, we obtain the effect of discounting of old measurements.

The objective of this work is twofold. First and foremost, we want to extend the idea of discounted measure-

ments to Bayesian methods. Second, the intention is to exploit this approach for solving as wide range of problems as possible. This can be achieved if we implement the adaptive processing scheme with discounted measurements by particle filters [4], [5]. Particle filters are based on tracking posterior densities of interest by propagating samples (particles) that are drawn from these densities. If particles from the posteriors are available at any instant of time, all kinds of estimators can be constructed for extracting desired information. Expectations of various functions can be estimated, MMSE estimates can be easily obtained, and uncertainties about the estimates can readily be quantified.

The paper is organized as follows. In Section 2 we present the main idea by working on a very simple example. Then, in Section 3, we discuss the implementation of the proposed method by particle filters. A simulation example is provided in Section 4, which shows that our expectations of the proposed method are met. Section 5 concludes the paper with some brief remarks.

2. DISCOUNTING OF OLD MEASUREMENTS AND BAYESIAN SIGNAL PROCESSING

Here we proceed by way of a very simple example. Suppose that the data y_t , $t = 1, 2, \dots$ are observed, and that

$$y_t = \theta + u_t \quad (1)$$

where θ is an unknown parameter, the noise samples u_t are independent and identically distributed, and $u_t \sim \mathcal{N}(0, \sigma_u^2)$, with σ_u^2 being known. After receiving one sample, all the information about θ is in its posterior, which takes the form

$$f(\theta|y_1) \propto \exp\left(-\frac{(\theta - y_1)^2}{2\sigma_u^2}\right) \quad (2)$$

where the prior of θ is assumed $f(\theta) \propto c$, with c being a constant. When the next sample y_2 is received, the posterior $f(\theta|y_1)$ is updated to $f(\theta|y_2, y_1)$ according to

$$f(\theta|y_2, y_1) \propto f(y_2|\theta)f(\theta|y_1) \quad (3)$$

which results in

$$f(\theta|y_2, y_1) \propto \exp\left(-\frac{(\theta - \hat{\theta}_2)^2}{\sigma_u^2}\right) \quad (4)$$

where

$$\hat{\theta}_2 = \frac{y_1 + y_2}{2}. \quad (5)$$

This work was supported by the National Science Foundation under Awards CCR-9903120 and CCR-0082607.

Note that in (3), the posterior from (2) acts as a prior for θ when it is estimated from y_2 , and it is multiplied with the likelihood $f(y_2|\theta)$. As new data are collected, this process is repeated. When y_t is received, we have

$$f(\theta|\mathbf{y}_{1:t}) \propto f(y_t|\theta)f(\theta|\mathbf{y}_{1:t-1}) \quad (6)$$

where $\mathbf{y}_{1:t} = [y_1 \ y_2 \ \dots \ y_t]$, and the posterior becomes

$$f(\theta|\mathbf{y}_{1:t}) \propto \exp\left(-\frac{(\theta - \hat{\theta}_t)^2}{\frac{2\sigma_u^2}{t}}\right) \quad (7)$$

where

$$\hat{\theta}_t = \frac{1}{t} \sum_{n=1}^t y_n. \quad (8)$$

It is obvious that the MAP estimate of θ after t samples is $\hat{\theta}_t$ and that it can be obtained by

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{t} (y_t - \hat{\theta}_{t-1}). \quad (9)$$

At this juncture, it is important to note that the RLS estimate of θ is found by minimizing the criterion

$$J_t = \sum_{n=1}^t (y_n - \theta)^2 \quad (10)$$

and that the resulting estimator is identical to (9). The difference between the estimators is that the Bayesian estimator tracks the full posterior of θ , whereas the RLS estimator provides only a point estimate. The RLS estimator, however, does not make distributional assumptions about the noise u_t .

The parameter θ however may change with time, and then its value at time instant t is denoted by θ_t . A standard approach to tracking its changes is to use the exponentially weighted RLS method that minimizes the criterion

$$J_t = \sum_{n=1}^t \lambda_n^{t-n} (y_n - \theta_n)^2 \quad (11)$$

where the λ_n 's are constants, and $0 < \lambda_n \leq 1$. In that case the RLS estimator is modified to [2]

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \gamma_t (y_t - \hat{\theta}_{t-1}) \quad (12)$$

where

$$\gamma_t = \frac{\gamma_{t-1}}{\lambda_t} \left(1 - \frac{\gamma_{t-1}}{\lambda_t + \gamma_{t-1}}\right). \quad (13)$$

The idea behind the use of the criterion (12) is to discount old measurements and allow newer measurements to affect the estimate of θ_t more than older measurements.

In absence of a function that models the dynamic nature of θ with time, is there an equivalent Bayesian approach to estimating θ_t , which in the case of Gaussian noise yields MAP estimate identical to the one given by (12)? The answer is yes. The following shows how we can find this estimator.

After receiving the first measurement, the posterior of θ_1 is given by (2). When y_2 is obtained, we can decrease the

effect of y_1 on θ_2 by using a different prior for θ_2 than the original one, which was $f(\theta_2|y_1) = f(\theta_1|y_1)$. The modified prior has the form

$$f(\theta_2|y_1, \lambda_1) \propto \exp\left(-\frac{(\theta_2 - y_1)^2}{\frac{2\sigma_u^2}{\lambda_1}}\right). \quad (14)$$

When this prior is combined with the likelihood $f(y_2|\theta_2)$, we obtain

$$f(\theta_2|y_2, y_1, \lambda_1) \propto \exp\left(-\frac{(\theta_2 - \frac{1}{1+\lambda_1}(y_2 + \lambda_1 y_1))^2}{\frac{2\sigma_u^2}{1+\lambda_1}}\right). \quad (15)$$

Next we receive y_3 , and the posterior (15) before becoming prior of θ_3 is spread out to reflect discounting of the measurements y_1 and y_2 , or

$$f(\theta_3|y_2, y_1, \lambda_2, \lambda_1) \propto \exp\left(-\frac{(\theta_3 - \frac{1}{1+\lambda_1}(y_2 + \lambda_1 y_1))^2}{\frac{2\sigma_u^2}{\lambda_2(1+\lambda_1)}}\right). \quad (16)$$

The new posterior then becomes

$$f(\theta_3|y_3, y_2, y_1, \lambda_2, \lambda_1) \propto \exp\left(-\frac{(\theta_3 - \hat{\theta}_3)^2}{\frac{2\sigma_u^2}{1+\lambda_2+\lambda_2\lambda_1}}\right) \quad (17)$$

where

$$\hat{\theta}_3 = \frac{1}{1+\lambda_2+\lambda_2\lambda_1}(y_3 + \lambda_2 y_2 + \lambda_2 \lambda_1 y_1). \quad (18)$$

The derivation of the following posteriors is analogous, and it is not difficult to show that the general expression for it is

$$f(\theta_t|\mathbf{y}_{1:t}, \boldsymbol{\lambda}_{1:t-1}) \propto \exp\left(-\frac{(\theta_t - \hat{\theta}_t)^2}{2\sigma_t^2}\right) \quad (19)$$

where

$$\hat{\theta}_t = \frac{y_t + \sum_{n=1}^{t-1} y_n \prod_{k=n}^{t-1} \lambda_k}{1 + \sum_{n=1}^{t-1} \prod_{k=n}^{t-1} \lambda_k} \quad (20)$$

and

$$\sigma_t^2 = \frac{\sigma_u^2}{1 + \sum_{n=1}^{t-1} \prod_{k=n}^{t-1} \lambda_k}. \quad (21)$$

It can be readily proved that the MAP estimate $\hat{\theta}_t$ given by (20) is identical to the RLS estimate (13). So, in summary, we conclude that we can invoke the concept of discounting measurements with the Bayesian methodology. This is done by appropriately modifying the posteriors, which serve as priors of the parameters that model the measurements yet to be taken.

The above example is for a very simple model and for Gaussian noise. It is not difficult to show that it can be straightforwardly replicated for more complex models, which not only can be nonlinear, but can also be models that involve non-Gaussian noise. Here we continue the work on our example, where the noise now is a Gaussian mixture with p components, i.e.,

$$u_t \sim \sum_{i=1}^p w_i \mathcal{N}(0, \sigma_{u_i}^2) \quad (22)$$

where the coefficients of the mixands are known, $\sum_{i=1}^p w_i = 1$, and $\sigma_{u_i}^2 \neq \sigma_{u_j}^2$ for $i \neq j$.

It can be shown that the posterior of θ_t given y_t, y_{t-1}, \dots, y_1 can be written as

$$f(\theta_t | \mathbf{y}_{1:t}, \boldsymbol{\lambda}_{1:t-1}) \propto \sum_{k=1}^{t^p} \tilde{w}_k f_{t_k}(\theta_t) \quad (23)$$

where the coefficients \tilde{w}_k and the mixands $f_{t_k}(\theta_t)$ can easily be determined. It is obvious, however, that the number of terms in the posterior grows exponentially as new measurements arrive, which makes the whole method of discounting old measurements in this case impossible to implement.

The tedious process of evaluating the posteriors when we have complicated posterior functions can be avoided if we adopt the concept of particle filters. Instead of tracking the analytical functions that represent the posterior, we follow a set of particles that come from the posterior and thereby approximate it. Any estimate that is of interest can then be easily obtained by using these particles. But how do we implement the discounting of old measurements using particle filters? An answer to this question is given in the next section.

3. IMPLEMENTATION BY PARTICLE FILTERS

First, we briefly explain the concept of particle filters. Suppose that an observed phenomenon is described by the equations

$$\theta_t = h_t(\theta_{t-1}, u_t) \quad (24)$$

$$y_t = g_t(\theta_t, v_t) \quad (25)$$

where $h_t(\cdot)$ and $g_t(\cdot)$ are some known functions, and u_t and v_t are noise samples from known distributions. The process θ_t is not observed, that is, it is hidden, and the objective is to track it sequentially using the samples y_t as soon as they become available.

We reiterate that all the information about θ_t is in its posterior density $f(\theta_t | \mathbf{y}_{1:t})$, so the best we can do is if a method is developed to track $f(\theta_t | \mathbf{y}_{1:t})$. Obviously, the nature of the method must be recursive, and therefore once y_{t+1} is received, the main idea is to modify $f(\theta_t | \mathbf{y}_{1:t})$ to $f(\theta_{t+1} | \mathbf{y}_{1:t+1})$. The recursive formula for the updating is

$$f(\theta_{t+1} | \mathbf{y}_{1:t+1}) = \frac{f(y_{t+1} | \theta_{t+1}) f(\theta_{t+1} | \mathbf{y}_{1:t})}{f(y_{t+1} | \mathbf{y}_{1:t})} \quad (26)$$

where

$$f(\theta_{t+1} | \mathbf{y}_{1:t}) = \int f(\theta_{t+1} | \theta_t) f(\theta_t | \mathbf{y}_{1:t}) d\theta_t. \quad (27)$$

When the functions in (24) and (25) are linear and the noises are Gaussian, the posteriors in (26) are also Gaussian, and as a result, it is sufficient to track only the first two moments of the posterior. In fact, the solution can then be obtained analytically, and the result is the Kalman filter. Deviations from linearity and Gaussianity lead to approximate solutions of which perhaps the most popular is the extended Kalman filter.

An interesting alternative to the standard solutions can be sought by employing particle filters, which are based on the concept of sequential importance sampling (SIS) [4], [5]. The main idea behind SIS is to approximate the posterior densities by samples (*particles*). Suppose that $\theta_t^{(m)}$, $m = 1, 2, \dots, M$ are particles from the density $f(\theta_t | \mathbf{y}_{1:t})$, each with probability mass $w_t^{(m)}$, where $\sum_{m=1}^M w_t^{(m)} = 1$. The particles with their probability masses represent an approximation of the posterior density from which they are drawn, i.e.,

$$\hat{f}(\theta_t | \mathbf{y}_{1:t}) = \sum_{m=1}^M w_t^{(m)} \delta(\theta_t - \theta_t^{(m)}) \quad (28)$$

where $\delta(\cdot)$ is the Dirac's delta function. As new data become available, the main idea is to propagate the particles and modify their weights so that the new set of particles and weights approximate $f(\theta_{t+1} | \mathbf{y}_{1:t+1})$.

Note that we can modify the posterior $f(\theta_{1:t} | \mathbf{y}_{1:t})$, with the arrival of y_{t+1} , according to

$$f(\theta_{1:t+1} | \mathbf{y}_{1:t+1}) = \frac{f(y_{t+1} | \theta_{t+1}) f(\theta_{t+1} | \theta_t)}{f(y_{t+1} | \mathbf{y}_{1:t})} f(\theta_{1:t} | \mathbf{y}_{1:t}). \quad (29)$$

The value of this expression is in that using the concept of particles it can be implemented recursively. If at time t , we have a set of particles and their weights from $f(\theta_{1:t} | \mathbf{y}_{1:t})$, they can be updated to particles with associated weights from $f(\theta_{1:t+1} | \mathbf{y}_{1:t+1})$ by applying the following sequential importance sampling procedure [4]:

1. Draw particles $\theta_{t+1}^{(m)}$, $m = 1, 2, \dots, M$, from a proposal density, known as *importance* function, $q(\theta_{t+1} | \theta_t, \mathbf{y}_{1:t+1})$.
2. Compute the weights of the particles by

$$w_{t+1}^{*(m)} = \frac{w_t^{(m)} f(y_{t+1} | \theta_{t+1}^{(m)}) f(\theta_{t+1}^{(m)} | \theta_t^{(m)})}{q(\theta_{t+1}^{(m)} | \theta_t^{(m)}, \mathbf{y}_{1:t+1})} \quad (30)$$

and

$$w_{t+1}^{(m)} = \frac{w_{t+1}^{*(m)}}{\sum_{k=1}^M w_{t+1}^{*(k)}}. \quad (31)$$

For details on the application of the procedure, see for example [4]. It is important to note that the accuracy of the method and the algorithms depends on the used importance function. For some choices of importance functions, consult [4], [5], [6], [7].

We now get back to our original topic of discussion, the enforcing of discounting of measurements with particle filters. It is worth noting two points: (a) the particle filters have no difficulties in handling hard problems like the one with Gaussian mixture presented above, and (b) in our problem statement we do not assume any model for the changes of the signal parameters. Now, the discounting of old measurements can easily be imposed if we use the particle filtering scheme on the following model:

$$\theta_t = \theta_{t-1} + v_t \quad (32)$$

$$y_t = \theta_t + w_t \quad (33)$$

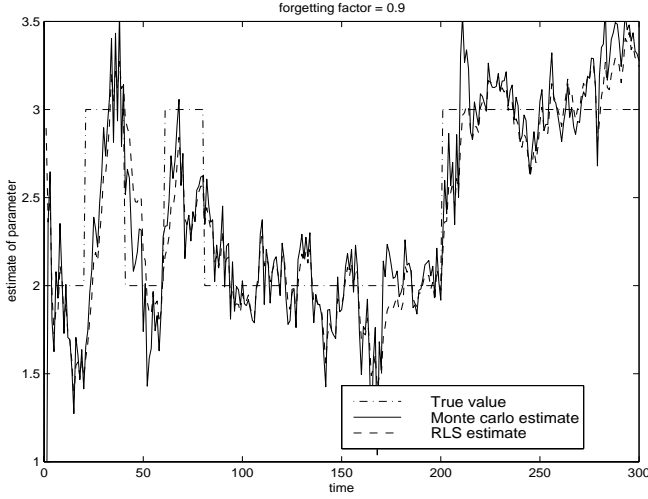


Figure 1: Performance comparison of the exponentially weighted RLS and proposed Monte Carlo sampling methods for parameter estimation. The noise variance $\sigma_u^2 = 1$.

where v_t is a zero mean noise sample with a known distribution, say Gaussian. It is obvious that the samples θ_t will have a wider distribution than the samples of θ_{t-1} because their distribution is convolved with the distribution of v_t . The value of the variance of v_t is easily deduced from (32), and it should be

$$\sigma_{v_t}^2 = \sigma_{\theta_{t-1}}^2 \left(\frac{1}{\lambda_t} - 1 \right). \quad (34)$$

In summary, the discounting of measurements is imposed by adding a system equation to the data model that represents a random walk. The resulting system is simple, and the tracking of the posterior density of θ_t by using particles is straightforward.

4. SIMULATION RESULTS

To illustrate the performance of the proposed methodology, we proceed with our simple example and compare its performance to that of the conventional RLS method. The model of the observations y_1 is given by (1), θ_t varies with time in an unknown way, and u_t is Gaussian with zero mean. To observe the behavior of the proposed method, θ_t was varied between the values 2 and 3, as shown in Figures 1 and 2. The simulation was run for $t = 300$ time samples, and σ_u^2 was 1 for the simulations in Figure 1 and 0.25 for Figure 2. The forgetting factor λ used for both the particle filter and RLS method was $\lambda_t = 0.9$, $t = 1, 2, \dots, 300$.

In both figures we observe that the particle filter algorithm tracks θ_t as well as the RLS method. As expected, due to the low value of the forgetting factor, there is considerable alertness to the dynamic nature of θ_t . However, the price is that the estimates are jittery and not very accurate. Note that the tracking performance of the particle filter is quite faithful to that of the conventional RLS.

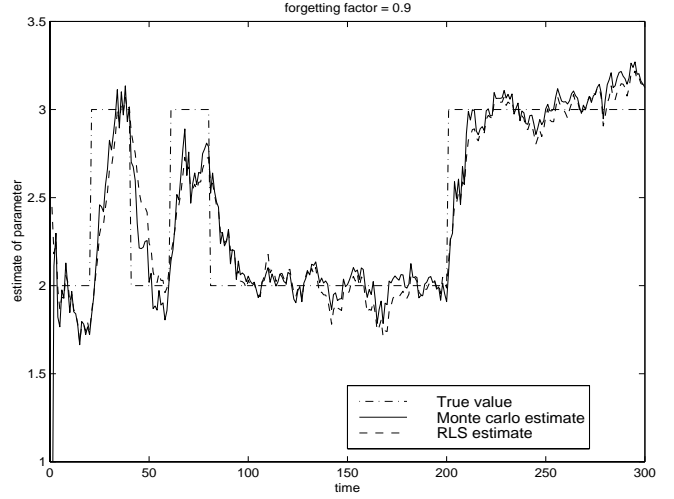


Figure 2: Performance comparison of the exponentially weighted RLS and proposed Monte Carlo sampling methods for parameter estimation. The noise variance $\sigma_u^2 = 0.25$.

5. CONCLUSIONS

We have proposed a Bayesian procedure for adaptive signal processing that employs discounting of old measurements. The discounting is implemented by convolving the most recent posterior of the tracked parameters with another density. The resulting density serves as a prior in processing the next observation. It is also proposed that the scheme is carried out by particle filters. A powerful feature of the method is that it can be readily applied to highly nonlinear problems that involve non-Gaussian noise.

6. REFERENCES

- [1] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.
- [2] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, The MIT Press, Cambridge, MA, 1983.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, third edition, 1996.
- [4] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [5] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.
- [6] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.
- [7] M. Pitt and N. Shepard, "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.