
ELSEVIER

Computer Physics Communications 110 (1998) 80–86

Computer Physics Communications

On the treatment of energy loss in track fitting

R. Frühwirth [a,1], S. Frühwirth-Schnatter [b]

[a] *Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften, A-1050 Wien, Nikolsdorfer Gasse 18, Austria*
[b] *Institut für Statistik, Wirtschaftsuniversität Wien, A-1090 Wien, Augasse 2–6, Austria*

Abstract

As the energy loss of electrons in matter is described by a strongly non-Gaussian distribution, the Kalman filter is not necessarily the optimal procedure for the fitting of electron tracks. We show that the momentum resolution can be improved by using nonlinear methods, such as the Gaussian-sum filter and the Metropolis–Hastings algorithm. We report results of a simple simulation experiment and comment on the respective merits of these methods. © 1998 Elsevier Science B.V.

Keywords: Energy loss; Electron tracks; Gaussian-sum filter; Metropolis–Hastings algorithm

1. Introduction

In track reconstruction with the Kalman filter, both measurement and process noise are implicitly assumed to be Gaussian. This assumption is justified in many cases. As far as the measurement noise is concerned, most position detectors show an approximately Gaussian behaviour. For minimum ionizing particles the process noise consists mainly of multiple Coulomb scattering. It is well known that the distribution of the projected scattering angle of multiple Coulomb scattering can be adequately modeled by a Gaussian. The most important case where the assumption of normality breaks down is the energy loss of electrons in matter, which is dominated by bremsstrahlung above a certain energy. This requires a stochastic model with a strongly non-Gaussian distribution [1]. In this case the Kalman filter is not necessarily the optimal filter, even if the model is perfectly linear. It is therefore worthwhile to investigate whether nonlinear estimators yield estimates with smaller variance.

In this note we examine two nonlinear estimators by means of a simulation experiment: the Gaussian-sum filter [2,3] and the Metropolis–Hastings algorithm [4]. The Gaussian-sum filter is a computationally inexpensive generalization of the Kalman filter. It requires, however, the approximation of non-Gaussian distributions by Gaussian mixtures; hence the performance of the filter depends on the quality of the approximation. In order to assess the performance of the Gaussian-sum filter we compare it to the Bayes estimator. We propose to compute the Bayes estimator via the Metropolis–Hastings algorithm, which is very simple to implement although it is not particularly efficient with respect to the computing time. Using a simplified model of the momentum

[1] E-mail: fruhwirth@hephy.oeaw.ac.at

PII S0010-4655(97)00157-4

measurement in the detector and of the process noise due to energy loss, we show that the estimation of the inverse momentum (or of the curvature) of electron tracks can actually be improved by taking into account the actual shape of the energy-loss distribution via a nonlinear estimator.

2. The simulation model

The model used in the simulation experiment is confined to the estimation of the inverse momentum of an electron track in front of and behind a layer of material with a thickness equivalent to t radiation lengths. The true momenta in front of and behind the material are denoted by p_0 and p_1, respectively. Energy loss is simulated according to the formula $p_1 = p_0 \cdot z$, where $-\ln z$ is Γ-distributed. The probability density function and the first two moments of z are given by

$$f(z) = \frac{(-\ln z)^{t/\ln 2 - 1}}{\Gamma(t/\ln 2)}, \quad 0 < z \leq 1; \qquad E_z = e^{-t}, \qquad V_z = e^{-t\ln 3/\ln 2} - e^{-2t}.$$

Fig. 1 shows the probability density function of z for a few values of t. Obviously the distribution of z is confined to the interval $(0, 1]$ and therefore has moments of all orders.

In a typical detector the estimate of the inverse momentum $q = 1/p$ is unbiased and approximately Gaussian. We assume that estimates $\hat{q}_0$ and $\hat{q}_1$ are available from track segments both in front of and behind the material. We consider these estimates as our observations. In the model they are drawn from Gaussian distributions with mean $q_i = 1/p_i$ and variance $V_i = q_i^2 \tau_i^2$, assuming a constant relative standard error $\tau_i = \sigma(\hat{q}_i)/q_i$.

3. Two nonlinear estimators

The baseline of our simulation experiment is the Kalman filter [5]. We are mainly interested in improving the estimate of q_0 on the front side of the material, which is closer to the interaction vertex. Therefore the filter runs in the direction opposite to the particle. The distribution of z is replaced by a Gaussian with mean E_z and variance V_z, and the system equation $q_0 = q_1 \cdot z$ is approximated by a linear one,

$$q_0 = q_1 + \delta, \qquad \delta = -q_1(1 - z), \qquad E_\delta \approx -\hat{q}_1(1 - E_z), \qquad V_\delta \approx \hat{q}_1^2 V_z.$$

The Kalman filter estimate $\tilde{q}_0$ can be written as a weighted mean of the observations $\hat{q}_0$ and $\hat{q}_1$,

$$\tilde{q}_0 = \hat{q}_0 \frac{V_1 + V_\delta}{V_0 + V_1 + V_\delta} + (\hat{q}_1 + E_\delta) \frac{V_0}{V_0 + V_1 + V_\delta}.$$

The Gaussian-sum filter (GSF) is a generalization of the standard Kalman filter [2,3]. The distribution of the process noise is now approximated by a mixture of Gaussians. The resulting filter is a weighted sum of Kalman filters, the weights depending on the observations. It is therefore nonlinear. The associated smoother is described in [3].

We have approximated the distribution of $z = p_1/p_0$ by a mixture of six Gaussians, using the general purpose minimization program MINUIT [6]. The objective function was the L_2-distance between the true density function and the mixture, plus two penalty terms forcing the sum of the mixture weights to 1 and the mean of the mixture to the true mean. We have tabulated the mixture parameters for $t = 0.05, 0.06, \ldots, 0.25$. The quality of the approximation has been checked by comparing the first four moments around 0. These are given by the following expressions:

$$\mu_1 = E(z) = e^{-t},$$
$$\mu_2 = E(z^2) = e^{-t\ln 3/\ln 2} = e^{-1.585t},$$

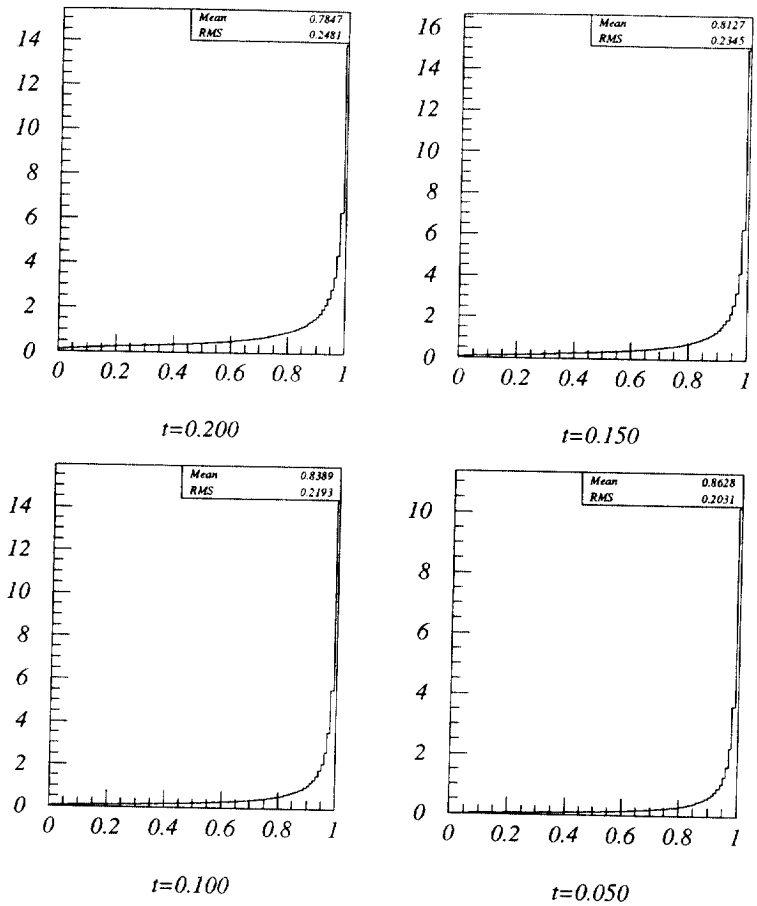

Fig. 1. The probability density function of $z = p_1/p_0$.

$$\mu_3 = E(z^3) = e^{-t\ln 4/\ln 2} = e^{-2t},$$

$$\mu_4 = E(z^4) = e^{-t\ln 5/\ln 2} = e^{-2.322t}.$$

Table 1 shows the first four moments of the actual distribution and of the approximating Gaussian mixture, for a few values of the thickness t. The first two moments match almost perfectly, while the deviation of the third and fourth moments is not more than 2–4%. Fig. 2 gives a visual impression of the quality of the approximation.

The posterior density of the GSF estimate is only an approximation to the true posterior. A fully Bayesian analysis can be implemented via the Metropolis–Hastings (M–H) algorithm [4], thus avoiding the cumbersome numerical evaluation of otherwise intractable integrals. Rather than computing the posterior distribution of the estimate, a Markov chain is constructed which has the required posterior (the "target" density) as its equilibrium distribution. The moments of the posterior can then be computed by sampling from the equilibrium distribution. In our case the target density is the joint density of q_0 and q_1 conditional on the observations $\hat{q}_0$ and $\hat{q}_1$. By means of Bayes' theorem, we have

$$f(q_0, q_1 | \hat{q}_0, \hat{q}_1) \propto f(\hat{q}_0 | q_0) \cdot f(\hat{q}_1 | q_1) \cdot f(q_0 | q_1).$$

Candidates for the Markov chain are drawn from a proposal density $g(q_0, q_1)$, which we have chosen as

$$g(q_0, q_1) = f(\hat{q}_1 | q_1) \cdot f(q_0 | q_1).$$

Table 1
Moments of true and approximating distributions

	μ_1		μ_2		μ_3		μ_4	
	true	appr.	true	appr.	true	appr.	true	appr.
$t = 0.05$	0.951	0.951	0.924	0.916	0.905	0.887	0.890	0.857
$t = 0.10$	0.905	0.904	0.853	0.849	0.819	0.808	0.793	0.765
$t = 0.15$	0.861	0.860	0.788	0.787	0.741	0.737	0.706	0.684
$t = 0.20$	0.819	0.819	0.728	0.729	0.670	0.671	0.629	0.609

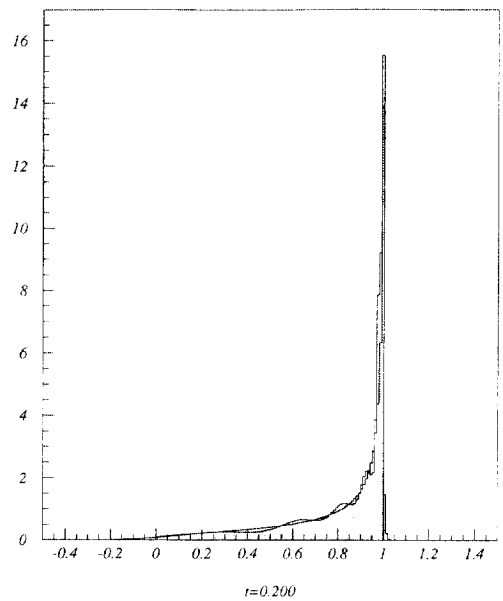

Fig. 2. The true probability density function of z (shaded) and the approximating Gaussian mixture (line).

If the most recent element of the Markov chain is denoted by $(q_0, q_1)^{(m)}$, the candidate $(q_0, q_1)^*$ is accepted as $(q_0, q_1)^{(m+1)}$ with the following probability:

$$\alpha = \min\left(\frac{w[(q_0, q_1)^*]}{w[(q_0, q_1)^{(m)}]}, 1\right),$$

where w is defined by

$$w(q_0, q_1) = \frac{f(q_0, q_1 | \hat{q}_0, \hat{q}_1)}{g(q_0, q_1)}.$$

In our case, we have

$$w(q_0, q_1) \propto \exp[-(\hat{q}_0 - q_0)^2/(2V_0)].$$

This leads to a simple expression for the acceptance probability and to reasonable acceptance rates. Also w is bounded, so that the entire support of the target density can be explored. Note that if the candidate $(q_0, q_1)^*$ is rejected, we set $(q_0, q_1)^{(m+1)} = (q_0, q_1)^{(m)}$.

Table 2
Efficiency of M–H and GSF relative to the Kalman filter

Metropolis–Hastings					Gaussian-sum filter				
$t =$	0.05	0.10	0.15	0.20	$t =$	0.05	0.10	0.15	0.20
$\rho = 0.2$	1.98	1.69	1.47	1.37	$\rho = 0.2$	1.98	1.68	1.47	1.37
$\rho = 0.4$	1.69	1.49	1.36	1.27	$\rho = 0.4$	1.69	1.48	1.38	1.27
$\rho = 0.6$	1.33	1.36	1.25	1.24	$\rho = 0.6$	1.35	1.36	1.26	1.24
$\rho = 0.8$	1.18	1.22	1.20	1.18	$\rho = 0.8$	1.22	1.27	1.22	1.18
$\rho = 1.0$	1.12	1.16	1.14	1.10	$\rho = 1.0$	1.14	1.18	1.15	1.11

Sampling from the proposal density g can be realized by first drawing q_1 from a Gaussian with mean $\hat{q}_1$ and variance V_1, and then multiplying with z, where $-\ln z$ is drawn from a Γ-distribution with parameter $c = t/\ln 2$. Sampling from the chain starts only after it has passed a "burn-in" phase and can be supposed to be in the equilibrium. We have worked with a burn-in phase of length 1000 and a sampling phase of length 5000. In particular, the mean value of the sample is the M–H estimate of (q_0, q_1). As the Markov chain is not an independent random sample from the distribution of (q_0, q_1), the computation of higher moments, in particular of the variance of the estimate, is not entirely straightforward. This is clearly a drawback of this approach. Another one is the relatively long computing time required to generate a Markov chain of sufficient length.

4. Results of the simulation experiment

It turns out that the improvement of the momentum resolution with respect to the Kalman filter depends critically on two model parameters. The first one is the number t of radiation lengths which are traversed by the electron; the second one is the ratio of the relative standard errors $\rho = \tau_1/\tau_0$ of the curvature measurements behind and in front of the material. In order to study the influence of these parameters we have run a batch of 5000 events for various combinations of t and ρ. The quality of the estimate $\tilde{q}_0$ can be characterized by its mean squared deviation (MSD) from the true value q_0. Since our baseline is the Kalman filter we measure the performance of the estimator by its relative efficiency η, which is the MSD of the Kalman filter divided by the MSD of the estimator.

The results are summarized in Table 2 which shows η as a function of t and ρ, for both the GSF and the M–H algorithm. We observe that the results are almost identical. It follows that the GSF operates at the Bayes limit, which is gratifying even though somewhat surprising in view of the various approximations which are used in setting up the mixture model of the process noise. A closer inspection of the results, however, reveals that the GSF estimator has a small bias which is not shown by the M–H estimator. There is an appreciable gain in efficiency when ρ is small, especially for relatively thin layers, in which case the distribution of z is much more skew than for thicker layers. In addition, Fig. 3 shows the frequency distribution of the residuals $\tilde{q}_0 - q_0$ of the estimated inverse momentum with respect to the true value, for both the GSF and the Kalman filter. It is clear that the GSF gives estimates which are nearly always closer to the true value; however, their distribution is obviously less similar to a true Gaussian than the linear estimates of the Kalman filter.

5. Discussion and outlook

We have shown that at least in our simple model the momentum resolution of electrons can be improved by using a nonlinear estimator. The computational load of the GSF is only slightly larger than the one of the Kalman filter; the M–H algorithm, however, is much more expensive in terms of computing time. In addition,

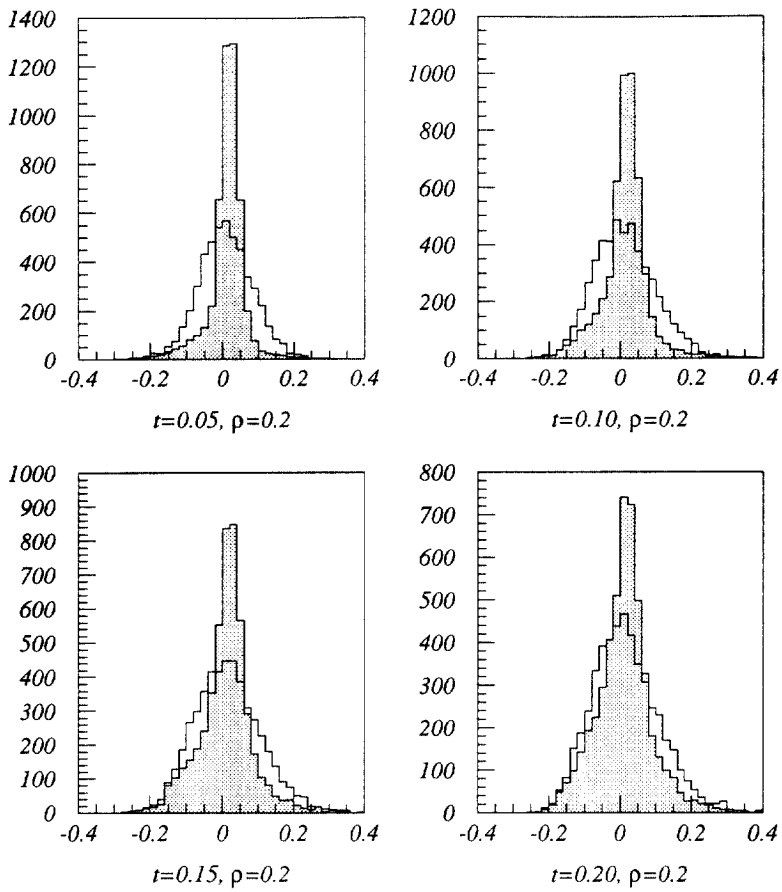

Fig. 3. Frequency distributions of the residuals of the estimates inverse momentum for the GSF (shaded) and the KF (line).

the variance of the estimate is obtained very easily from the GSF; it is simply the variance of the posterior mixture and can be computed analytically from the first two moments of the components. If the posterior mixture is written as

$$f(q_0) = \sum_{i=1}^{n} w_i \varphi(q_0; \tilde{q}_{0i}, \sigma_i^2) ,$$

the estimate and its variance are given by

$$\tilde{q}_0 = \sum_{i=1}^{n} w_i \tilde{q}_{0i} , \qquad \sigma^2 = \sum_{i=1}^{n} w_i \sigma_i^2 + \sum_{i=1}^{n} w_i (\tilde{q}_{0i}^2 - \tilde{q}_0^2) .$$

In the case of the M–H algorithm, the variance has to be computed from the Markov chain. Since the Markov chain is not an independent random sample, autocorrelations of higher orders have to be taken into account. Given the fact that the GSF operates virtually at the Bayes limit we find the arguments in favour of the GSF overwhelming. We therefore intend to use it in further studies of electron reconstruction. The next step is the extension of the simple model to a full five-dimensional state space model of the track and to a realistic model of the detector and of the process noise, including multiple scattering and correlations of $q = 1/p$ with the

other track parameters. If the conclusions drawn here are validated, the method could be put to its final test on real data.

References

[1] D. Stampfer, M. Regler, R. Frühwirth, Comput. Phys. Commun. 79 (1994) 157.
[2] G. Kitagawa, Ann. Inst. Statist. Math. 46 (1994) 605.
[3] R. Frühwirth, Comput. Phys. Commun. 100 (1997) 1.
[4] S. Chib, E. Greenberg, The American Statistician 49 (1995) 327.
[5] R. Frühwirth, Nucl. Instr. Meth. A 262 (1987) 444.
[6] F. James, MINUIT - Function Minimization and Error Analysis, CERN Program Library Long Writeup D506 (CERN, Geneva, 1994).