On the Treatment of Energy Loss in Track Fitting

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

[a] *Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften, A-1050 Wien, Nikolsdorfer Gasse 18*

[b] *Institut für Statistik, Wirtschaftsuniversität Wien, A-1090 Wien, Augasse 2–6*

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 non-linear 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.

Key words: Energy loss; electron tracks; Gaussian-sum filter; Metropolis–Hastings algorithm

1 The simulation model

In track reconstruction with the Kalman filter both measurement and process noise are implicitely assumed to be Gaussian. This assumption is justified in many cases. The most important exception 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 note we show that the estimation of the inverse momentum (or of the curvature) of electron tracks can be improved by taking into account the actual shape of the distribution. Using a highly simplified model of the momentum measurement in the detector and of the process noise due to energy loss, we examine two non-linear estimators by means of a simulation experiment: the Gaussian-sum filter [2,3] and the Metropolis–Hastings algorithm [4].

The model used in the simulation experiment is confined to the estimation of the inverse momentum of the track in front of and behind a layer of material with a thickness equivalent to d radiation lengths. The true momenta in front

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

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

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

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 $\tau_i = \sigma(\hat{q}_i)/q_i$.

2 Two non-linear estimators

The baseline of our simulation experiment is the Kalman filter. 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:

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

The estimate $\tilde{q}_0$ is a weighted mean of the observations $\hat{q}_0$ and $\hat{q}_1$.

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 filter is a weighted sum of Kalman filters, the weights depending on the observations. 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 [5]. 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 $d = 0.05, 0.06, \ldots, 0.25$. In order to check the quality of the approximation beyond a visual inspection we have compared the first four moments around 0. The largest difference is about 2 percent.

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 $\hat{q}_1$. Candidates for the Markov chain are drawn from a proposal density $g(q_0, q_1)$, which we have chosen as $g(q_0, q_1) = f(\hat{q}_1 | q_1) \cdot f(q_0 | q_1)$. This choice leads to a simple expression for the acceptance probability and to reasonable acceptance rates. Simulating 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 $c = d/\ln 2$. 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).

3 Results of the simulation experiment

It turns out that the improvement of the momentum resolution depends critically on two model parameters: the number d of radiation lengths traversed, and the ratio of the relative standard errors $\rho = \tau_1/\tau_0$. In order to study the influence of these parameters we have run a batch of 5000 events for various combinations of d 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. Table 1 shows η as a function of d and ρ, for both the GSF and the M–H algorithm. We observe that the results are almost identical.

Table 1
Efficiency of M–H and GSF relative to the Kalman filter

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

It follows that the GSF operates at the Bayes limit, which is gratifying even if somewhat surprising in view of the various approximations 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 effiency when ρ is small, especially for relatively thin layers, where the distribution of z is much more skew than for thicker layers. The final energy of the electron is normally measured by a calorimeter in which the energy resolution scales as $E^{-1/2}$. At least for high energy electrons we can therefore expect ρ to be reasonably small. In such a case using a non-linear estimator is certainly worth the effort.

4 Discussion and outlook

We have shown that at least in our simple model the momentum resolution of electrons can be improved by using a non-linear 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, 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. In the case of the M–H algorithm, the variance has to be computed from the sample, taking into account autocorrelations of higher orders. Given the fact that the GSF operates virtually at the Bayes limit, the arguments in favour of the GSF are 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 and R. Frühwirth, Comp. Phys. Comm. 79 (1994) 157.

[2] G. Kitagawa, Ann. Inst. Statist. Math. 46 (1994) 605.

[3] R. Frühwirth, Comp. Phys. Comm. 100 (1997) 1.

[4] S. Chib and E. Greenberg, The American Statistician 49 (1995) 327.

[5] F. James, MINUIT – Function Minimization and Error Analysis, CERN Program Library Long Writeup D506, CERN, Geneva (1994).