Applications of Mathematics

Nikolai I. Chernov; Genadij A. Ososkov; Luc Pronzato

3-dimensional multivertex reconstruction from 2-dimensional tracks observations using likelihood inference

Applications of Mathematics, Vol. 37 (1992), No. 6, 437–452

Persistent URL: `http://dml.cz/dmlcz/104522`

37 (1992) APPLICATIONS OF MATHEMATICS No. 6, 437–452

3-DIMENSIONAL MULTIVERTEX RECONSTRUCTION FROM 2-DIMENSIONAL TRACKS OBSERVATIONS USING LIKELIHOOD INFERENCE

N. I. Chernov, G. A. Ososkov and L. Pronzato

(Received December 17, 1990)

Summary. Let $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_k$ be vertices in the XYZ-space, each vertex producing several tracks (straight lines) emanating from it within a narrow cone with a small angle about a fixed direction (Z-axis). Each track is detected (by drift chambers or other detectors) by its projections on XY and YZ views independently with small errors. An automated method is suggested for the reconstruction of vertices from noisy observations of the tracks projections. The procedure is based on the likelihood inference for mixtures. An illustrative example is considered.

1. Introduction and problem statement

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_v}$, $\mathbf{v}_i = (v_{i_x}, v_{i_y}, v_{i_z})^T$, $i = 1, \ldots, N_v$, be the original vertices in the 3-dimensional XYZ-space producing tracks (straight lines) emanating from them within a narrow cone with a small angle about the direction of the Z-axis. Not only the locations of these tracks are unknown but also their number N_v. The projections of the tracks are detected by drift chambers or other detectors in the XZ and YZ views independently. Reconstruction of vertices by visual inspection, although feasible, is rather tedious and we suggest an automated procedure for that purpose.

We shall assume throughout the paper that the problem of allocation of noisy observations among tracks is satisfactorily solved. Let $\mathbf{x}_j$, $\mathbf{z}_j$ ($\mathbf{y}_j$, $\mathbf{z}_j$) be the vectors of the coordinates of the N_0 observations corresponding to the j-th track, $j = 1, \ldots, N_t$, in the XZ- (the YZ-) plane. It is not possible to recognize *a priori* which track in one plane corresponds to a given track in the other plane. The 3-dimensional problem is then split into two subproblems, one associated with tracks in the XZ plane, the other with tracks in the YZ plane.

The method to be used to recover the original vertices relies on the maximum likelihood estimation for mixtures, briefly presented in Section 2. An algorithmic procedure is described in Section 3. The notion of atomic likelihood vectors is the central point of this procedure, and two possible choices, based on different approximations, are suggested in Section 4. In Sections 2–4 we only discuss the problem in the XZ-plane, the approach being quite similar in the YZ-plane. The choice of the admissible domains for vertex candidates in the XZ- and YZ-planes is considered in Section 5. Finally, Section 6 presents an illustrative example.

2. Likelihood inference and vertex reconstruction

Consider a distribution of vertices $\pi(\mathbf{v})$ in a domain $\mathscr{D}_x$ of the XZ-plane, and let $f(\mathbf{x}, \mathbf{z}, \mathbf{v})$ be the probability density function of the observations $(\mathbf{x}, \mathbf{z})$ when the associated track is assumed to be produced by $\mathbf{v}$ ($f(\mathbf{x}, \mathbf{z}, \mathbf{v})$ will be chosen continuous with respect to $\mathbf{v}$ given $\mathbf{x}$, $\mathbf{z}$). The problem consists in estimating the distribution π via maximum likelihood. We observe random variables $\mathbf{x}$, $\mathbf{z}$ with the mixture marginal density

$$f(\mathbf{x}, \mathbf{z}, \pi) = \int_{\mathscr{D}_x} f(\mathbf{x}, \mathbf{z}, \mathbf{v}) \pi(d\mathbf{v}). \tag{1}$$

Using independence of observations between tracks, one obtains for the likelihood

$$L(\pi) = \prod_{i=1}^{N_t} \int_{\mathscr{D}_x} f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) \pi(d\mathbf{v}), \tag{2}$$

and $\ln(L(\pi))$ is a concave functional on the set $\mathscr{P}$ of all probability measures on $\mathscr{D}_x$. This formulation stresses the similarity between the vertex reconstruction problem and optimal experimental design, widely considered in the literature (see e.g. [Fedorov, 1972; Silvey, 1980; Pázman, 1986]). The following geometrical interpretation can also be formulated [Lindsay, 1983].

For any $\mathbf{v}$ in $\mathscr{D}_x$, consider the N_t-dimensional vector (atomic likelihood vector)

$$\mathbf{f}(\mathbf{v}) = \big(f(\mathbf{x}_1, \mathbf{z}_1, \mathbf{v}), f(\mathbf{x}_2, \mathbf{z}_2, \mathbf{v}), \ldots, f(\mathbf{x}_{N_t}, \mathbf{z}_{N_t}, \mathbf{v})\big)^T. \tag{3}$$

The likelihood curve is then the function from $\mathscr{D}_x$ to $\mathscr{R}^{N_t}$ defined by $\mathbf{v} \to \mathbf{f}(\mathbf{v})$, and the trace $\mathscr{G}$ of this curve represents all possible values of atomic likelihood vectors.

Any vector of densities

$$\mathbf{f}(\pi) = \big(f(\mathbf{x}_1, \mathbf{z}_1, \pi), f(\mathbf{x}_2, \mathbf{z}_2, \pi), \ldots, f(\mathbf{x}_{N_t}, \mathbf{z}_{N_t}, \pi)\big)^T, \tag{4}$$

where $f(\mathbf{x}, \mathbf{z}, \pi)$ is given by (1), can then be written as a convex combination of elements of $\mathscr{G}$. The maximization of $L(\pi)$ (2) with respect to π is then equivalent to the maximization of the functional $\prod_{i=1}^{N_t} f_i$, where $\mathbf{f}(\pi) = (f_1, f_2, \ldots, f_{N_t})^T$ belongs to the convex hull $\mathrm{conv}(\mathscr{G})$ of the set $\mathscr{G}$. Note that this is equivalent to the maximization of the volume of the right N_t-hedron with opposite vertices distant of $||\mathbf{f}||$.

Remark 1. Two intersections of estimated tracks in the XZ- and YZ-planes that correspond to the same original vertex must have the same Z-coordinate (approximately because of measurement errors). This property will be used in Section 5 to define the admissible domains $\mathscr{D}_x$ and $\mathscr{D}_y$. Now, if it were possible to associate pairs of tracks in the XZ- and YZ-planes *a priori*, so that the observations $\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{z}_i$ would correspond to the i-th track in the XYZ-space, one could reconstruct vertices in both planes simultaneously with an equality constraint for the Z-coordinates. Due to the independence of observations between the two planes, the function $f(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{v})$ could then be chosen as $f_x(\mathbf{x}, \mathbf{z}, \mathbf{v}_x) \cdot f_y(\mathbf{y}, \mathbf{z}, \mathbf{v}_y)$, with $\mathbf{v}_x = (v_x, v_z)^T$, $\mathbf{v}_y = (v_y, v_z)^T$.

An algorithmic procedure based on likelihood inference is now presented, to be used for vertex reconstruction in the plane.

3. Algorithmic procedure

Let $\pi_{\mathbf{v}}$ be the one-point measure which puts its mass 1 at $\mathbf{v}$, and define $Q(\alpha)$ as

$$Q(\alpha) = \ln\big(L\big((1-\alpha)\pi + \alpha\pi_{\mathbf{v}}\big)\big). \tag{5}$$

This corresponds to a move of $\mathbf{f}(\pi)$ (4) in the direction of $\mathbf{f}(\pi_{\mathbf{v}}) = \mathbf{f}(\mathbf{v})$ (3). The first derivative of $Q(.)$ with respect to α at $\alpha = 0$ is equal to

$$Q'(0) = d(\mathbf{v}, \pi) - N_t \tag{6}$$

with

$$d(\mathbf{v}, \pi) = \sum_{i=1}^{N_t} \frac{f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v})}{L_i(\pi)}, \tag{7}$$

and

$$L_i(\pi) = \int_{\mathscr{D}_x} f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v})\pi(d\mathbf{v}). \tag{8}$$

$\hat{\pi}$ is then a maximum likelihood estimator of π if and only if [Lindsay, 1983]

$$\max_{\mathbf{v} \in \mathscr{D}_x} d(\mathbf{v}, \hat{\pi}) = N_t. \tag{9}$$

Various algorithms have been suggested in the literature to optimize $L(\pi)$ (see e.g. [Lindsay, 1983; Torsney, 1983, 1988; Mallet, 1986; Böhning, 1989] and the references therein). They differ depending on whether $\mathscr{D}_x$ is a discrete or a continuous set. We suggest to use the vertex-exchange method, combined with the trapezoidal step-length rule [Böhning, 1989], which was experimentally found to provide good convergence properties whether $\mathscr{D}_x$ is discrete or continuous. The procedure is the following one:

1: Set $k = 1$, $\pi^1 = \pi^0$, where π^0 is an initial discrete distribution of vertices. When $\mathscr{D}_x$ is a discrete set, as suggested in Section 5, a possible choice consists of a uniform distribution of masses in $\mathscr{D}_x$.

2: Evaluate

$$d_M = \max_{\mathbf{v} \in \mathscr{D}_x} d(\mathbf{v}, \pi^k) \tag{10}$$

if $d_M - N_t < \varepsilon$ stop,
otherwise go step 3, with ε a small tolerance number.

3: Let

$$\mathbf{v}^+ = \arg \max_{\mathbf{v} \in \mathscr{D}_x} d(\mathbf{v}, \pi^k), \tag{11}$$

and

$$\mathbf{v}^- = \arg \min_{\mathbf{v} \in \mathrm{supp}(\pi^k)} d(\mathbf{v}, \pi^k), \tag{12}$$

where $\mathrm{supp}(\pi)$ denotes the set of supporting points of the distribution π. Compute the scalars w_i, $i = 1, \ldots, N_t$, defined by

$$w_i = \pi^k(\mathbf{v}^-)\frac{f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}^+) - f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}^-)}{L_i(\pi^k)}, \tag{13}$$

where $\pi(\mathbf{v}^-)$ is the mass attached to $\mathbf{v}^-$ in the distribution π, and where $L_i(\pi)$ is given by (8). The function $Q(\alpha)$ corresponding to this vertex exchange algorithm is now given by $\ln\big(L\big(\pi + \alpha\pi(\mathbf{v}^-)(\pi_{\mathbf{v}^+} - \pi_{\mathbf{v}^-})\big)\big)$. Evaluate

$$Q'(1) = \sum_{i=1}^{N_t} \frac{w_i}{1+w_i}. \tag{14}$$

If

$$Q'(1) \geqslant 0, \text{ set } \alpha^* = 1 \text{ (exchange } \mathbf{v}^- \text{ and } \mathbf{v}^+\text{)} \tag{15}$$

otherwise evaluate

$$Q'(0) = \sum_{i=1}^{N_t} w_i, \tag{16}$$

$$Q''(0) = -\sum_{i=1}^{N_t} w_i^2, \tag{17}$$

$$Q''(1) = -\sum_{i=1}^{N_t} \left(\frac{w_i}{1+w_i}\right)^2. \tag{18}$$

If

$$Q''(1) \leqslant Q''(0) \text{ set } \alpha^* = -b + \sqrt{b^2 - c}, \tag{19}$$

if

$$Q''(1) > Q''(0) \text{ set } \alpha^* = -b - \sqrt{b^2 - c}, \tag{20}$$

where

$$b = \frac{Q''(0)}{Q''(1) - Q''(0)}, \tag{21}$$

and

$$c = \frac{2Q'(0)}{Q''(1) - Q''(0)}. \tag{22}$$

Perform a vertex-exchange step of length α^*,

$$\pi^{k+1} = \pi^k + \alpha^*\pi^k(\mathbf{v}^-)(\pi_{\mathbf{v}^+} - \pi_{\mathbf{v}^-}), \tag{23}$$

increase k by 1 and go to step 2.

R e m a r k 2. If $\mathbf{v}^-$, as defined by (12), is "far-away" from all other tracks except the i-th one (i.e. $f(\mathbf{x}_j, \mathbf{z}_j, \mathbf{v}^-) = 0$ for $j \neq i$) and if $\mathbf{v}^+$, as defined by (11), is "far-away" from the i-th track (i.e. $f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}^+) = 0$), w_i, as given by (13), is equal to -1. $Q'(1)$ (14) and $Q''(1)$ (18) are then no longer defined, and the step length α^* cannot be chosen according to (15), (19) or (20). In this case, we suggest to choose $\alpha^* = \frac{1}{k}$, which corresponds to Wynn's step-length sequence for the optimal experimental design [Wynn, 1970]. Note that one then looses monotonicity for the evolution of $L(\pi^k)$, which is obtained when the procedure (10–23) is used [Böhning, 1989].

R e m a r k 3. Other step-length sequences than the trapezoidal one defined by (15–22) are suggested by Böhning [1989], e.g.

$$\alpha^* = -\frac{Q'(0)}{\min\{Q''(0), Q''(1)\}},$$

or

$$\alpha^* = -\frac{Q'(0)}{Q'(1) - Q'(0)}.$$

One could also perform the vertex exchanges (23) corresponding to each one of these step-lengths, and select the best one with respect to the criterion $L(\pi)$ (2). Finally, a one-dimensional maximization of $L(\pi^{k+1})$ with respect to α^* could also be considered, with π^{k+1} given by (23).

R e m a r k 4. Any multiplicative factor in $f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v})$ that does not depend on $\mathbf{v}$ plays no role in the expressions of $d(\mathbf{v}, \pi)$ (7) and w_i (13). Such factors can therefore be omitted.

R e m a r k 5. The number $N_t \cdot \pi^k(\mathbf{v})$, where $\pi^k(\mathbf{v})$ is the mass attached to a reconstructed vertex $\mathbf{v}$, gives an estimate of the number of tracks passing through $\mathbf{v}$.

R e m a r k 6. The algorithmic procedures in the XZ- and YZ-planes are independent. They can therefore be performed simultaneously on two different computers, thereby saving computational time.

4. Choice of the atomic likelihood vectors

The i-th component of the atomic likelihood vector (3) expresses the probability density function of the observations $\mathbf{x}_i$, $\mathbf{z}_i$ conditional to the vertex position $\mathbf{v}$. Approximations will be introduced to derive the expression of $f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v})$. We first consider the case when the measurement errors can be neglected.

4.1 Negligible measurement errors. Let θ_i be the parameters corresponding to the i-th track. The probability density function of the observations $\mathbf{x}_i$, $\mathbf{z}_i$ given $\mathbf{v}$ can be written as

$$f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) = \int p(\mathbf{x}_i, \mathbf{z}_i \,|\, \mathbf{v}, \theta_i) p(\theta_i \,|\, \mathbf{v}) \, d\theta_i. \tag{24}$$

When there are no measurement errors, $p(\theta_i \,|\, \mathbf{v})$ is equal to zero provided $\mathbf{v}$ does not belong to the track $T(\theta_i)$ associated with θ_i. However, in this case there is no density of θ_i because the Lebesgue measure of the set $\{\theta_i \in \mathscr{R}^2; \mathbf{v} \in T(\theta_i)\}$ is zero. Hence we can suppose that $p(\theta_i \,|\, \mathbf{v})$ is nonzero also when $\mathbf{v}$ does not belong to $T(\theta_i)$ but is close to it. A natural choice then is

$$p(\theta_i \,|\, \mathbf{v}) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp - \frac{d^2(\mathbf{v}, \theta_i)}{2\sigma_i^2}, \tag{25}$$

where $d(\mathbf{v}, \theta_i)$ denotes the distance from $\mathbf{v}$ to the track $T(\theta_i)$. Note that $p(\theta_i \,|\, \mathbf{v})$ does not depend on the orientation of the track, i.e. this involves no physical assumption. When the equation of the track is formulated as

$$x = \theta_{i_1} z + \theta_{i_2}, \tag{26}$$

one has

$$d^2(\mathbf{v}, \theta_i) = \frac{(v_x - \theta_{i_1} v_z - \theta_{i_2})^2}{1 + \theta_{i_1}^2}. \tag{27}$$

Equation (24), with the hypothesis of negligible measurement errors, gives

$$f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) = p(\hat{\theta}_i(\mathbf{x}_i, \mathbf{z}_i) \,|\, \mathbf{v}), \tag{28}$$

where $\hat{\theta}_i(\mathbf{x}_i, \mathbf{z}_i)$ is the estimated value of θ_i obtained from the data $\mathbf{x}_i$, $\mathbf{z}_i$. Assuming that these data are related by

$$\mathbf{x}_i = Z_i \theta_i + \varepsilon_i \tag{29}$$

with

$$Z_i = (\mathbf{z}_i, \mathbf{u}_{N_0}), \tag{30}$$

where $\mathbf{u}_{N_0}$ is the N_0-dimensional vector given by $\mathbf{u}_{N_0} = (1, 1, \ldots, 1)^T$, and that the measurement errors ε_i are normally distributed $\mathcal{N}(0, \Sigma_i)$, one obtains for the maximum likelihood estimator of θ_i,

$$\hat{\theta}_i(\mathbf{x}_i, \mathbf{z}_i) = (Z_i^T \Sigma_i^{-1} Z_i)^{-1} Z_i^T \Sigma_i^{-1} \mathbf{x}_i. \tag{31}$$

Note that this includes the assumption that the coordinates $\mathbf{z}_i$, $i = 1, \ldots, N_t$, are known without errors, which can be justified in practical situations. Using (25–31) and taking Remark 4 into account, one can finally choose

$$f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) = \exp -\frac{2}{2\sigma_i^2} \frac{\left(\omega_z^T (Z_i^T \Sigma_i^{-1} Z_i)^{-1} Z_i^T \Sigma_i^{-1} \mathbf{x}_i - v_x\right)^2}{\left\| \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \frac{1}{(Z_i^T \Sigma_i^{-1} Z_i)^{-1} Z_i^T \Sigma_i^{-1} \mathbf{x}_i} \right\|^2} \tag{32}$$

with

$$\omega_z = (v_z, 1)^T. \tag{33}$$

We now consider an alternative approach, to be used when measurement errors cannot be neglected. In this case, the vertex reconstruction via the estimation of the tracks parameters can be highly approximative, and a derivation of the atomic likelihood vectors directly based on the observations is advisable.

4.2 Non-informative prior for the track parameters and vertices considered as ordinary additional observations. Consider again equation (24), with the model structure (29–30) and measurement errors ε_i normally distributed $\mathcal{N}(0, \Sigma_i)$. When these errors can no longer be neglected, the conditional probability density $p(\mathbf{x}_i, \mathbf{z}_i | \theta_i)$ is given by

$$p(\mathbf{x}_i, \mathbf{z}_i | \theta_i) = \frac{1}{(2\pi)^{N_0/2} \det^{1/2} \Sigma_i} \times \exp -\frac{1}{2} (\mathbf{x}_i - Z_i \theta_i)^T \Sigma_i^{-1} (\mathbf{x}_i - Z_i \theta_i). \tag{34}$$

Replacing $p(\theta_i | \mathbf{v})$ and $p(\mathbf{x}_i, \mathbf{z}_i | \theta_i)$ by their expressions (25) and (34), respectively, the integral (24) can be evaluated, at least theoretically. However, this seems untracktable from the computional point of view, and we will adopt a Bayesian approach together with some simplifying approximations.

Using the Bayes rule, we obtain

$$f(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) = \frac{p(\mathbf{x}_i, \mathbf{z}_i | \theta_i) p(\theta_i | \mathbf{v})}{p(\theta_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{v})}, \tag{35}$$

which permits to replace the difficulty of evaluating an integral by that of evaluating the densities $p(\theta_i|\mathbf{v})$ and $p(\theta_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{v})$. Assuming a prior density p_{0_i} for the parameters θ_i, we obtain for these posterior densities

$$p(\theta_i|\mathbf{v}) = \frac{p(\mathbf{v}|\theta_i) p_{0_i}(\theta_i)}{\int p(\mathbf{v}|\theta_i) p_{0_i}(\theta_i)\, \mathrm{d}\theta_i}, \tag{36}$$

and

$$p(\theta_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}) = \frac{p(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}|\theta_i) p_{0_i}(\theta_i)}{\int p(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}|\theta_i) p_{0_i}(\theta_i)\, \mathrm{d}\theta_i}. \tag{37}$$

They can easily be derived when

(i) p_{0_i} is assumed to be normal $\mathcal{N}(\theta_{0_i}, \Omega_i)$;

(ii) $\mathbf{v} = (v_x, v_z)^T$, given θ_i, is considered as a simple additional observation to the data $\mathbf{x}_i$, $\mathbf{z}_i$, with a measurement error normally distributed $\mathcal{N}(0, \sigma_i^2)$ independently of ε_i (note that this implies an approximation since, contrary to what happens for the actual observations $\mathbf{x}_i$, $\mathbf{z}_i$, the Z-coordinate v_z of $\mathbf{v}$ is uncertain, too).

We have in this case

$$p(\mathbf{v}|\theta_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp -\frac{1}{2\sigma_i^2}(v_x - \omega_z^T\theta_i)^2 \tag{38}$$

with ω_z given by (33), and

$$p(\mathbf{x}_i, \mathbf{z}_i, \mathbf{v}|\theta_i) = p(\mathbf{x}_i, \mathbf{z}_i|\theta_i) p(\mathbf{v}|\theta_i) \tag{39}$$

with $p(\mathbf{x}_i, \mathbf{z}_i|\theta_i)$ given by (34). We thus obtain after an elementary calculation

$$p(\theta_i|\mathbf{v}) = \frac{\det^{1/2}\left(\Omega_i^{-1} + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)}{2\pi} \times \exp -\frac{1}{2}(\theta_i - \hat{\theta}_{Bi}(\mathbf{v}))^T \left(\Omega_i^{-1} + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)(\theta_i - \hat{\theta}_{Bi}(\mathbf{v})) \tag{40}$$

with

$$\hat{\theta}_{Bi}(\mathbf{v}) = \left(\Omega_i^{-1} + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)^{-1}\left(\omega_z\frac{v_x}{\sigma_i^2} + \Omega_i^{-1}\theta_{0_i}\right), \tag{41}$$

$$p(\theta_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{v}) = \frac{\det^{1/2}\left(\Omega_i^{-1} + Z_i^T\Sigma_i^{-1}Z_i + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)}{2\pi} \tag{42}$$

$$\times \exp -\frac{1}{2}(\theta_i - \hat{\theta}_{Bi}^{+}(\mathbf{v}))^T\left(\Omega_i^{-1} + Z_i^T\Sigma_i^{-1}Z_i + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)(\theta_i - \hat{\theta}_{Bi}^{+}(\mathbf{v}))$$

with

$$\hat{\theta}_{Bi}^{+}(\mathbf{v}) = \left(\Omega_i^{-1} + Z_i^T\Sigma_i^{-1}Z_i + \frac{\omega_z\omega_z^T}{\sigma_i^2}\right)^{-1}\left(Z_i^T\Sigma_i^{-1}\mathbf{x}_i + \omega_z\frac{v_x}{\sigma_i^2} + \Omega_i^{-1}\theta_{0_i}\right). \tag{43}$$

We can then replace these posterior densities by their expressions in (35).

When the prior mean θ_{0_i} and the covariance matrix Ω_i cannot be chosen according to some physical grounds, a non-informative density can be considered by choosing $\Omega_i = kI_2$ and letting k tend to infinity. A series development of $f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v})$ gives, after lengthy but elementary algebraic manipulations,

$$f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v}) = \frac{1}{\sqrt{k}}\frac{1}{(2\pi)^{N_0/2}\det^{1/2}\Sigma_i\det^{1/2}Z_i^T\Sigma_i^{-1}Z_i} \tag{44}$$

$$\times \frac{\|\omega_z\|}{\left(\sigma_i^2 + \omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}\omega_z\right)^{1/2}}$$

$$\times \exp -\frac{1}{2}\mathbf{x}_i^T(\Sigma_i^{-1} - \Sigma_i^{-1}Z_i(Z_i^T\Sigma_i^{-1}Z_i)^{-1}Z_i^T\Sigma_i^{-1})\mathbf{x}_i$$

$$\times \exp -\frac{1}{2}\frac{\left(\omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}Z_i^T\Sigma_i^{-1}\mathbf{x}_i - v_x\right)^2}{\sigma_i^2 + \omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}\omega_z} + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

Following Remark 4, $f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v})$ can finally be chosen as

$$f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v}) = \frac{\|\omega_z\|}{\left(\sigma_i^2 + \omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}\omega_z\right)^{1/2}} \tag{45}$$

$$\times \exp -\frac{1}{2}\frac{\left(\omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}Z_i^T\Sigma_i^{-1}\mathbf{x}_i - v_x\right)^2}{\sigma_i^2 + \omega_z^T(Z_i^T\Sigma_i^{-1}Z_i)^{-1}\omega_z}.$$

Remark 7. When the measurement errors ε_{i_j}, $j = 1, \ldots, N_0$, are independently identically normally distributed $\mathcal{N}(0,\sigma_i^2)$, so that $\Sigma_i = \sigma_i^2 I_{N_0}$, an estimated value $\hat{\sigma}_i^2$ for σ_i^2 can be obtained from the residuals,

$$\hat{\sigma}_i^2 = \frac{1}{N_0 - 1}\|\mathbf{x}_i - Z_i(Z_i^TZ_i)^{-1}Z_i^T\mathbf{x}_i\|^2, \tag{46}$$

to be used in the expressions (32, 45) of $f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v})$. However, this value $\hat{\sigma}_i^2$ was experimentally found to yield functions $f(\mathbf{x}_i,\mathbf{z}_i,\mathbf{v})$ particularly sharp, so that some

w_i's (13) achieved the value -1, thereby slowing down the algorithm of Section 3 (see Remark 2). Intuitively, σ_i in (32, 45) can be related to the minimal acceptable distance between two distinct vertices. Values between diam($\mathcal{D}_x$)/100 and diam($\mathcal{D}_x$)/10 then seem reasonable in practice.

5. Choice of the admissible domain for vertex candidates

The complexity and the performances of the vertex reconstruction procedure strongly depend on the choice of the domains $\mathcal{D}_x$ and $\mathcal{D}_y$ to be used in the optimization step (10–11) of the algorithm described in Section 3. The original vertices can be located *a priori* anywhere in a continuous part of the 3-dimensional space limited by experimental constraints. In such situations, continuous admissible domains are usually considered because they give more accurate results than discrete sets, though they lead to heavier computations. However, the vertex reconstruction problem in the plane cannot be solved without ambiguity when appropriate constraints are not set on $\mathcal{D}_x$ and $\mathcal{D}_y$. Consider for instance the situation presented in Figure 1 concerning the XZ-plane.

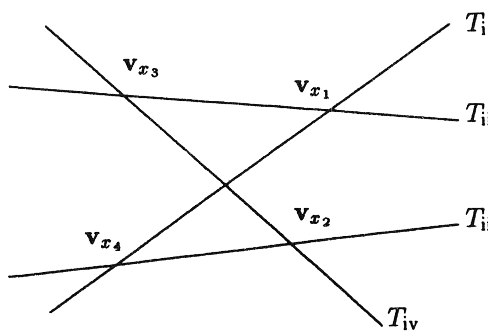

Fig. 1: Ambiguity caused by the 2-dimensional projection of a 3-dimensional event. $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$ correspond to the original vertices, $\mathbf{v}_{x_3}$ and $\mathbf{v}_{x_4}$ are extraneous vertices.

Let the original vertices correspond to $\mathbf{v}_{x_1}$ and $\mathbf{v}_{x_2}$, producing respectively the projected tracks T_{i}, T_{ii} and T_{iii}, T_{iv}. The probability that the 3-dimensional tracks with T_{i}–T_{iv} have other intersections than the original vertices is zero. $\mathbf{v}_{x_3}$ and $\mathbf{v}_{x_4}$ are then extraneous vertices, caused by the 2-dimensional projection of a 3-dimensional picture. Yet, if the problem in the XZ-plane is treated independently of the problem in the YZ-plane, $\mathbf{v}_{x_3}$ and $\mathbf{v}_{x_4}$ could as well be considered as projections of the original vertices producing respectively T_{ii}, T_{iv} and T_{i}, T_{iii}. In order to limit this ambiguity as far as possible, we shall take Remark 1 into account: any original vertex producing tracks has the same Z-coordinate v_z in both projections. On the YZ-projection corresponding to Figure 1, there will hopefully be no vertex $\mathbf{v}_y$ with a Z-coordinate close to $v_{x_{3_z}}$ or $v_{x_{4_z}}$. Discrete feasible sets $\mathscr{D}_x$ and $\mathscr{D}_y$ will then be constructed according to the following steps.

1: Construct initial sets of intersections $\mathscr{I}_x$ and $\mathscr{I}_y$.

Consider the estimated tracks, characterized by parameters $\hat{\theta}_{x_i}$, $i = 1, \ldots, N_t$, in the XZ-plane and parameters $\hat{\theta}_{i_y}$, $i = 1, \ldots, N_t$, in the YZ-plane. The sets $\mathscr{I}_x = \{\mathbf{v}_{x_1}, \mathbf{v}_{x_2}, \ldots, \mathbf{v}_{x_{N_x}}\}$, ($\mathscr{I}_y = \{\mathbf{v}_{y_1}, \mathbf{v}_{y_2}, \ldots, \mathbf{v}_{y_{N_y}}\}$) of the intersections of all tracks in the XZ-plane (in the YZ-plane) are then constructed (only the vertices falling in the experimental region of interest are considered).

2: Discard extraneous candidates from $\mathscr{I}_x$ and $\mathscr{I}_y$.

For any vertex $\mathbf{v}_{x_j}$ in $\mathscr{I}_x$ ($\mathbf{v}_{y_k}$ in $\mathscr{I}_y$), let σ_{x_j} (σ_{y_k}) be the minimal distance between $\mathbf{v}_{x_j}$ ($\mathbf{v}_{y_k}$) and any other vertex $\mathbf{v}_x$ ($\mathbf{v}_y$) for these two vertices to be distinguished (see Remark 7). Any vertex $\mathbf{v}_{x_j}$ in $\mathscr{I}_x$ such that

$$(47) \qquad |v_{x_{j_z}} - v_{y_{k_z}}| > \max(\sigma_{x_j}, \sigma_{y_k}) \forall \mathbf{v}_{y_k} \in \mathscr{I}_y$$

cannot correspond to a vertex producing tracks in the XYZ-space and is therefore discarded from $\mathscr{I}_x$. Similarly, any vertex $\mathbf{v}_{y_k}$ in $\mathscr{I}_y$ such that

$$(48) \qquad |v_{x_{j_z}} - v_{y_{k_z}}| > \max(\sigma_{x_j}, \sigma_{y_k}) \forall \mathbf{v}_{x_j} \in \mathscr{I}_x$$

is discarded from $\mathscr{I}_y$.

3: Introduce additional vertices in $\mathscr{I}_x$ and $\mathscr{I}_y$.

When more than two tracks are produced by one vertex, the presence of measurement noise implies that the estimated tracks will not intersect exactly at the same point. Several false vertices are therefore introduced in $\mathscr{I}_x$ and $\mathscr{I}_y$, while a single vertex, combination of these false vertices, would be more appropriate. We shall then introduce new candidates in $\mathscr{I}_x$ and $\mathscr{I}_y$, combinations of candidates already present and close enough to possibly correspond to the noisy reconstruction of a single vertex. We only describe the procedure concerning $\mathscr{I}_x$, the procedure for $\mathscr{I}_y$ being quite similar.

Consider the vertices in $\mathscr{I}_x$, one at a time. For any $\mathbf{v}_{x_j}$ in $\mathscr{I}_x$, let $\mathscr{I}_{x_j}$ be the set defined by

$$(49) \qquad \mathscr{I}_{x_j} = \{\mathbf{v}_{x_{j_k}} \in \mathscr{I}_x, \mathbf{v}_{x_{j_k}} \neq \mathbf{v}_{x_j}, \|\mathbf{v}_{x_{j_k}} - \mathbf{v}_{x_j}\| \leqslant \sigma_{x_j}\},$$

and let n_j be its cardinal. If $n_j = 0$, consider the next vertex in $\mathscr{I}_x$, otherwise consider the $C_{n_j}^m$ subsets $\mathscr{I}_{x_j}(i, m)$ of $\mathscr{I}_{x_j}$ with cardinal m, $i = 1, \ldots, C_{n_j}^m$,

$m = 1, \ldots, n_j$. A new vertex $\mathbf{v}_{x_j}(i,m)$ is associated with each of these subsets according to

$$(50)\quad \mathbf{v}_{x_j}(i,m) = \left(\frac{1}{\sigma_{x_j}} + \sum_{k=1}^{m} \frac{1}{\sigma_{x_{j_k}}}\right)^{-1} \left(\frac{\mathbf{v}_{x_j}}{\sigma_{x_j}} + \sum_{k=1}^{m} \frac{\mathbf{v}_{x_{j_k}}}{\sigma_{x_{j_k}}}\right), \quad \mathbf{v}_{x_{j_k}} \in \mathscr{I}_{x_j}(i,m),$$

($\sum_{m=1}^{n_j} C_{n_j}^m = 2^{n_j} - 1$ such new vertices are defined). When all vertices in $\mathscr{I}_x$ have been considered, these new vertices are introduced in I_x. The same procedure in applied to $\mathscr{I}_y$.

4: Repeat step 2.

Extraneous vertices may be introduced during step 3 so that the discarding policy corresponding to (47–48) must be applied again, to finally yield $\mathscr{D}_x$ and $\mathscr{D}_y$.

Remark 8. Other expressions than (50) could be used to define the next vertices introduced in $\mathscr{D}_x$ and $\mathscr{D}_y$ during step 3. For instance, for each vertex $\mathbf{v}_{x_j}$ in $\mathscr{I}_x$ consider the set $\mathscr{T}_{x_j}$ of all tracks that do not pass through $\mathbf{v}_{x_j}$ but pass at a distance from it less than σ_{x_j}. Let n_j be the number of such tracks. When $n_j > 0$, one can associate a new vertex $\mathbf{v}_{x_j}(i,m)$ with each subset $\mathscr{T}_{x_j}(i,m)$ of m such tracks, $i = 1, \ldots, C_{n_j}^m$, $m = 1, \ldots, n_j$. Considering these m tracks and the two tracks that intersect in $\mathbf{v}_{x_j}$ as intersecting in $\mathbf{v}_{x_j}(i,m)$, one can then estimate $\mathbf{v}_{x_j}(i,m)$ by standard least squares.

6. Example

Three original vertices $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ have been randomly generated in the domain $\mathscr{D}$ given by

$$(51)\qquad -1 \leqslant v_x \leqslant 1, \quad -1 \leqslant v_y \leqslant 1, \quad 0 \leqslant v_z \leqslant 1.$$

Their location is known, and will be used to test the procedure of reconstruction,

$$(52)\qquad \begin{aligned} \mathbf{v}_1 &= (-0.7499, -0.3101, 0.2352)^T, \\ \mathbf{v}_2 &= (0.2365, 0.2458, 0.3943)^T, \\ \mathbf{v}_3 &= (0.4170, -0.6078, 0.3965)^T. \end{aligned}$$

Track passing through these vertices have been randomly generated (2 through $\mathbf{v}_1$, 3 through $\mathbf{v}_2$ and 2 through $\mathbf{v}_3$), together with 11 noisy observations along each of these seven tracks. The model structure which is used is given by (29–30) and the measurement errors ε_{i_j} are independently normally distributed $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 10^{-4}$. All vectors $\mathbf{z}_i$, $i = 1, \ldots, 7$, are taken equal to $(1, 1.1, 1.2, \ldots, 1.9, 2)^T$. Figure 2 presents the associated XZ-view, with the observations in the XZ-plane, the tracks estimated from the observations, and the projections of the original vertices.

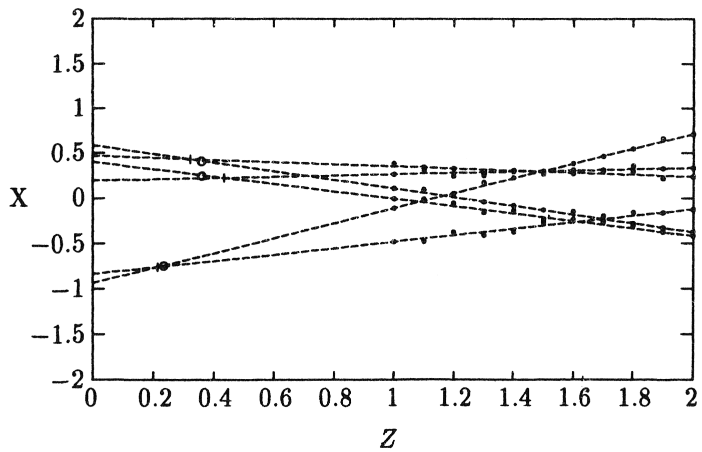

Fig. 2: Projections of the original vertices (∘), noisy observations (•), estimated tracks (— — —) and reconstructed vertices (+) in the XZ-plane.

We first construct the set for the admissible candidates $\mathscr{D}_x$ and $\mathscr{D}_y$, following the procedure described in Section 5. The initial sets of intersections $\mathscr{I}_x$ and $\mathscr{I}_y$ both contain 6 vertices. The σ_{x_j}'s and σ_{y_k}'s in (47–50) are taken equal to 0.1. $\mathscr{D}_x$ and $\mathscr{D}_y$ then both contain 4 vertices.

We apply the algorithmic procedure of Section 3. The atomic likelihood vectors are based on the function f given by (45) with $\Sigma_i = \sigma_i^2 I_{11}$ and $\sigma_i^2 = 0.01$. The initial distribution π^0 (step 1) consists of a uniform distribution of masses (equal to 0.25) in $\mathscr{D}_x$ and $\mathscr{D}_y$. The value of ε (step 2) is taken equal to 0.01. In the XZ- (YZ-) plane the algorithm stops after 6 (8) iterations. The reconstructed vertices are given by

$$
\begin{aligned}
(53) \qquad \mathbf{v}_{x_1} &= (-0.7516, 0.2189)^T, \\
\mathbf{v}_{x_2} &= (0.2306, 0.4381)^T, \\
\mathbf{v}_{x_3} &= (0.4289, 0.3651)^T
\end{aligned}
$$

in the XZ-plane and

$$(54) \qquad \begin{aligned} \mathbf{v}_{y_1} &= (-0.2845, 0.2865)^T, \\ \mathbf{v}_{y_2} &= (0.2784, 0.4504)^T, \\ \mathbf{v}_{y_3} &= (-0.6132, 0.4015)^T \end{aligned}$$

in the YZ-plane. These values can be compared to the coordinates of the original vertices (52). The location of the reconstructed vertices in the XZ-plane is indicated in Figure 2. The estimated numbers of tracks passing through the reconstructed vertices (see Remark 5) are respectively 1.9999, 3.3387 and 1.6614 for $\mathbf{v}_{x_1}$, $\mathbf{v}_{x_2}$ and $\mathbf{v}_{x_3}$, and 2.0852, 2.9826, 1.9329 for $\mathbf{v}_{y_1}$, $\mathbf{v}_{y_2}$ and $\mathbf{v}_{y_3}$. These numbers can be compared to the actual number of tracks, respectively 2, 3 and 2, passing through $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$.

Remark 9. If the reconstructed vertices in the XZ- and YZ-planes are simply paired off by a comparison of their Z-coordinate, depending on the actual value of the measurement noise, confusion is possible when several original vertices have their Z-coordinates almost identical. In particular, in the example $\mathbf{v}_{x_3}$ ($\mathbf{v}_{x_2}$) could be paired off with $\mathbf{v}_{y_2}$ ($\mathbf{v}_{y_3}$) for a different set of noisy observations. When the original vertices which are concerned produce different numbers of tracks, the confusion can be avoided by a comparison of the numbers of estimated tracks passing through the reconstructed vertices in the two planes. The estimated number of tracks given in Remark 5 can then help to rightly pair the reconstructed vertices off.

7. Conclusions

Maximum likelihood inference has been used to develop an automated procedure for vertex reconstruction. One determines the maximum likelihood estimator for the distribution of vertices, given noisy observations of projections of tracks. The optimization algorithm is applied in both planes of projections simultaneously, with convenient admissible domains for vertices. The choice of these admissible domains for vertex candidates has been discussed. It permits to avoid ambiguity that would result from treating problems in both planes independently. A numerical example has illustrated the behavior of the reconstruction policy. The algorithm located the original vertices in a very small number of iterations. However, less favorable situations may be encountered where confusion in pairing off the reconstructed vertices in the two planes cannot be avoided by any procedure. Further investigations are required to determine the most efficient choices for likelihood vectors and numerical constants used in the procedure.

References

[1] *Böhning, D*: Likelihood inference for mixtures: geometrical and other constructions of monotone step-length algorithms, Biometrika *76* no. 2 (1989), 375–383.

[2] *Fedorov, V. V.*: Theory of Optimal Experiments, Academic Press, New York, 1972.

[3] *Lindsay, B. G.*: The geometry of mixture likelihoods: a general theory, Annals of Stat. *11* no. 1 (1983), 86–94.

[4] *Mallet, A.*: A maximum likelihood estimation method for random coefficient regression models, Biometrika *73* no. 3 (1986), 645–656.

[5] *Pázman, A.*: Foundations of Optimum Experimental Design, co-editor VEDA, Bratislava, Reidel, Dordrecht, 1986.

[6] *Silvey S. D.*: Optimal Design, Chapman & Hall, London, 1980.

[7] *Torsney, B.*: A moment inequality and monotonicity of an algorithm, Semi-Infinite Programming and Applications (A. V. Fiacco and K. O. Kortanek, eds.), Springer-Verlag, Berlin, 1983, pp. 249–260.

[8] *Torsney, B.*: Computing optimizing distributions with applications in design, estimation and image processing, Optimal Design and Analysis of Experiments (Y. Dodge, V. V. Fedorov and H. P. Wynn, eds.), North-Holland, Amsterdam, 1988, pp. 361–370.

[9] *Wynn, H. P.*: The sequential generation of D-optimum experimental designs, Annals of Math. Stat. *41* (1970), 1655–1664.

Authors' addresses: *N. I. Chernov, G. A. Ososkov*, Joint Institute for Nuclear Research, Laboratory of Computing Techniques and Automation, Head Post Office, P.O. Box 79, 101000, Moscow, USSR; *L. Pronzato*, Laboratoire des Signaux et Systèmes, CNRS/ESE, Plateau de Moulon, 91192 Gif-sur-Yvette cedex, France.