The Fisher-Rao Metric for Projective Transformations of the Line

STEPHEN J. MAYBANK

School of Computer Science and Information Systems, Birkbeck College, Malet Street, London, WC1E 7HX, UK
sjmaybank@dcs.bbk.ac.uk

Abstract. A conditional probability density function is defined for measurements arising from a projective transformation of the line. The conditional density is a member of a parameterised family of densities in which the parameter takes values in the three dimensional manifold of projective transformations of the line. The Fisher information of the family defines on the manifold a Riemannian metric known as the Fisher-Rao metric. The Fisher-Rao metric has an approximation which is accurate if the variance of the measurement errors is small. It is shown that the manifold of parameter values has a finite volume under the approximating metric.

These results are the basis of a simple algorithm for detecting those projective transformations of the line which are compatible with a given set of measurements. The algorithm searches a finite list of representative parameter values for those values compatible with the measurements. Experiments with the algorithm suggest that it can detect a projective transformation of the line even when the correspondences between the components of the measurements in the domain and the range of the projective transformation are unknown.

Keywords: asymptotic expansion, canonical volume, Fisher-Rao metric, heat equation, probability of false detection, projective transformation of the line, Riemannian manifold.

1 Introduction

A projective transformation of the line is a bilinear map from the projective line $\mathbb{P}^1$ to itself or from $\mathbb{P}^1$ to another copy of $\mathbb{P}^1$. These transformations have a number of applications in computer vision, especially in stereo vision. If two images of a straight line are taken from different viewpoints, then the map between corresponding image points is a projective transformation of the line. If two cameras take images of the same scene from different viewpoints, then the epipolar lines in the first image comprise a one parameter family which has the structure of $\mathbb{P}^1$ in that each epipolar line is associated with a point of $\mathbb{P}^1$. The correspondence between the epipolar lines in the first image and the epipolar lines in the second image is a projective transformation of the line (Faugeras, 1993; Hartley and Zisserman, 2000). The terminology in this area varies from one publication to another. Hartley and Zisserman (2000) reserve the term 'projective transformation' for projective transformations of the plane. They use the term 'homography' for projective transformations of the line. Faugeras (1993) uses the term 'epipolar collineation' for the correspondence between epipolar lines.

A projective transformation of the line can be estimated using measurements of points in the domain and in the range of the transformation. A single measurement is a pair of points, (x_1, x_2), in which x_1 is in the domain and x_2 is in the range. In the noise free case the projective transformation maps x_1 to x_2 exactly, and three measurements are in general sufficient to determine the projective transformation uniquely.

If the measurements are subject to noise, then there is in general no projective transformation exactly compatible with four or more measurements. The usual procedure in this case, and the one taken here, is to use a probabilistic description of the measurement errors. If the measurement errors are samples from independent Gaussian random variables, then a projective transformation θ compatible with the measurements is found by making a least squares fit to the measurements. The transformation θ is referred to as an explanation or model for the measurements. Any other projective transformation sufficiently close to θ is also a suitable model for the measurements.

Two questions arise:

i) How large is the set $B(\theta)$ of projective transformations which are sufficiently close to θ to provide suitable models for the measurements?

ii) Is it possible to find a finite set of models $\theta(1), \ldots, \theta(n)$ such that for any projective transformation θ there is at least one model $\theta(i)$ such that $\theta(i)$ is in $B(\theta)$?

The aim of this paper is to answer both questions and to show how the answers lead to a simple algorithm for detecting projective transformations of the line. Question *(i)* is answered in Section 4.2 and question *(ii)* is answered in Section 5.1.

The following geometric and probabilistic framework is used. The geometrical part of the framework consists of a parameterisation of the manifold $\mathrm{PSL}(2, \mathbb{R})$ of projective transformations of the line. The probabilistic part consists of a probability density function for the errors in the measurement $x = (x_1, x_2)$ of the points x_1, x_2 which correspond under a projective transformation. The whole framework is summarised by the family of conditional probability density functions, $p(x|\theta)$.

Let θ, ψ be points of $\mathrm{PSL}(2, \mathbb{R})$. If θ, ψ are sufficiently close together, then it is unlikely that a single measurement x will contain enough information to distinguish between θ and ψ. If $p(x|\theta)$ is large, then θ is a likely model for x, but $p(x|\psi)$ is large as well, therefor ψ is also a likely model for x. The idea of 'sufficiently close' is made quantitative using the Fisher information, $J(\theta)$, which defines a Riemannian metric on $\mathrm{PSL}(2, \mathbb{R})$, known in statistics as the Fisher-Rao metric (Amari, 1985; Fisher, 1922; Kotz and Johnson, 1992; Rao, 1945). The points θ, ψ are 'sufficiently close' if they are close under the Fisher-Rao metric.

Let $B(\theta)$ be the set of points of $\mathrm{PGL}(2, \mathbb{R})$ close to θ under the Fisher-Rao metric. The point θ is thought of as a single representative model for all of the points in $B(\theta)$ (Myung et al., 2000). It is shown that $\mathrm{PGL}(2, \mathbb{R})$ is covered by the union of a finite number of sets $B(\theta(i))$, $1 \leq i \leq n_s$. The number n_s is a measure of the complexity of the task of detecting a projective transformation of the line compatible with a given set of measurements. A projective transformation can be found by checking each $\theta(i)$ in turn for compatibility with the measurements.

If the measurement noise is low, then the Fisher-Rao metric J can be replaced by a computationally tractable approximation K. The less the measurement noise, the more accurately K approximates to J. It is shown that $\mathrm{PSL}(2, \mathbb{R})$ has a finite volume under K. The sets $B(\theta(i))$, $1 \leq i \leq n_s$ all have similar volumes under K. The ratio of the volume of T to the volume of a set $B(\theta(i))$ is a lower bound for n_s. The method used to obtain K from J is applicable to a wide range of structure detection problems. An application to line detection is described in Maybank (2004).

A simple algorithm is given for finding the models $\theta(i)$ compatible with a given set of measurements. The algorithm takes as input a set of N measurements and a threshold $r \leq N$. It detects a projective transformation $\theta(i)$ of the line if there are r or more measurements compatible with $\theta(i)$. The threshold r is needed to reduce the probability of a false detection. If the noise level is very small, then n_s is large and it is no longer possible to check each $\theta(i)$ in a reasonable time. In such cases the detection algorithm can be speeded up by using a multiresolution approach.

All of the theoretical results are independent of the choice of parameterisation of $\mathrm{PSL}(2, \mathbb{R})$. The reason for this independence is the following transformation rule for the Fisher-Rao metric: if ξ is an alternative parameter for $\mathrm{PSL}(2, \mathbb{R})$ such that the family of probability density functions $\xi \mapsto p(x|\theta(\xi))$ has the Fisher-Rao metric $J'(\xi)$, then

$$J'_{ij}(\xi) = \sum_{k,l=1}^{3} \frac{\partial \theta_k}{\partial \xi_i} J_{kl}(\theta) \frac{\partial \theta_l}{\partial \xi_j}, \qquad 1 \leq i, j \leq 3. \tag{1}$$

It follows from Eq. (1) that $J(\xi)$ and $J(\theta)$ define the same metric on $\mathrm{PSL}(2, \mathbb{R})$. A consequence of Eq. (1) is that the parameterisation of $\mathrm{PSL}(2, \mathbb{R})$ can be and is chosen to simplify the calculation of $J(\theta)$ and $K(\theta)$.

The geometrical and probabilistic framework is described in Section 2. The approximation to the Fisher information on $\mathrm{PSL}(2, \mathbb{R})$ is obtained in Section 3. The number of models is estimated in Section 4. An algorithm for detecting the models compatible with a given set of measurements is described in Section 5. The experimental results are described in Section 6 and some concluding remarks are made in Section 7.

2 Geometric and Probabilistic Framework

The geometric and probabilistic framework is similar to that described in Maybank (2004). A different method of defining a Fisher-Rao metric on $\mathrm{PSL}(2, \mathbb{R})$ can be found in Maybank (2003).

2.1 The Group of Projective Transformations of the Line

Let $(u_1, u_2)^\top$ be coordinates for the real projective line $\mathbb{P}^1$. As usual, at least one of u_1, u_2 is required to be non-zero, and $(u_1, u_2)^\top$, $(v_1, v_2)^\top$ are coordinates for the same point of $\mathbb{P}^1$ if and only if there exists a number $s \neq 0$ such that $u_i = sv_i$ for $i = 1, 2$ (Faugeras, 1993; Semple and Kneebone, 1952). Let H be a 2×2 non-singular matrix with real entries,

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}.$$

The matrix H defines a projective transformation of the line,

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mapsto H \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} h_{11}u_1 + h_{12}u_2 \\ h_{21}u_1 + h_{22}u_2 \end{pmatrix}. \tag{2}$$

If $s \neq 0$, then H and sH define the same projective transformation. The projective transformations represented by matrices H with $\det(H) = 1$ form a Lie group, $\mathrm{PSL}(2, \mathbb{R})$, which has dimension 3 as a manifold. From now on only the projective transformations in $\mathrm{PSL}(2, \mathbb{R})$ are considered.

The singular value decomposition (Golub and Van Loan, 1996) of H is

$$H = \begin{pmatrix} \cos(b) & -\sin(b) \\ \sin(b) & \cos(b) \end{pmatrix}^\top \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{pmatrix}. \quad (3)$$

The parameters $(a, b - \pi, \lambda)$ and $(a - \pi, b, \lambda)$ both yield the matrix $-H$. The matrices H and $-H$ give rise to the same projective transformation, thus the ranges of a, b can be and are restricted to $0 \leq a, b < \pi$. The parameter λ has the range $\lambda \geq 1$. The parameter λ is replaced by the parameter ϕ defined such that

$$\lambda^2 = \cot(\phi), \qquad 0 < \phi \leq \pi/4. \quad (4)$$

The reason for the replacing λ by ϕ will become apparent in Section 3. It simplifies the formulae for the entries of the approximation K to the Fisher-Rao metric.

Let S_r be the circle with radius r. The parameter vector for H is $\theta = (a, b, \phi)$ and the manifold T of values of θ is chosen such that $T = S_{1/2} \times S_{1/2} \times (0, \pi/4)$. The manifold T does not include $S_{1/2} \times S_{1/2} \times \{\pi/4\}$. The reason for omitting $S_{1/2} \times S_{1/2} \times \{\pi/4\}$ from T is that the correspondence between parameter values and elements of $\mathrm{PSL}(2, \mathbb{R})$ ceases to be one to one when $\phi = \pi/4$ or equivalently $\lambda = 1$. The manifold T is open and dense in $\mathrm{PSL}(2, \mathbb{R})$,thus T is sufficient for calculating probabilities and estimating parameter values.

The projective line $\mathbb{P}^1$ is given an angular coordinate $x_1 \in [-\pi/2, \pi/2)$ chosen such that $(u_1, u_2)^\top = (\sin(x_1), \cos(x_1))^\top$. If x_1 is the angular coordinate in the domain of the projective transformation and x_2 is the angular coordinate in the range, then Eq. (2) becomes

$$x_2 = \tan^{-1}\left(\frac{h_{11}\tan(x_1) + h_{12}}{h_{21}\tan(x_1) + h_{22}}\right), \qquad -\pi/2 \leq x_1, x_2 < \pi/2. \quad (5)$$

The points x_1, x_2 are said to correspond under H.

2.2 Measurement space

It follows from the description of the measurements given in Sections 1 and 2.1 that the measurement space D is $D = \mathbb{P}^1 \times \mathbb{P}^1$. Each measurement is a point $x = (x_1, x_2)$ of D, where x_1, x_2 are angles such that $-\pi/2 \leq x_1, x_2 < \pi/2$. The real projective line $\mathbb{P}^1$ has the topology of a circle, thus D has the topology of a torus.

Let $x = (x_1, x_2)$ be a noise free measurement exactly compatible with the projective transformation $\theta = (a, b, \phi)$. It follows from Eqs. (3), (4) and (5) that x is on the submanifold $M(\theta)$ of D defined by $f(x, \theta) = 0$ where

$$f(x, \theta) \equiv x_2 - b - \tan^{-1}(\cot(\phi)\tan(x_1 - a)), \qquad x \in D. \quad (6)$$

The measurement space D is compact and $M(\theta)$ is a closed submanifold of D. Examples of the manifolds $M(\theta)$ for two different values of θ are shown in Fig. 1.

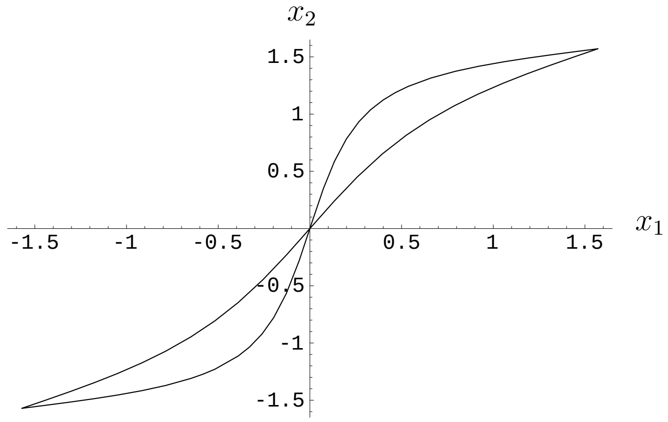

Figure 1. Manifolds $M(\theta)$ for (a) $\theta = (0, 0, 0.2)$ - steep curve; and (b) $\theta = (0, 0, 0.5)$ - shallow curve.

The space D usually has a metric, g, which is closely related to its use as a measurement space. The simplest choice, and the one adopted in this paper, is to choose g to be the flat metric on D. The flat metric agrees with the Euclidean metric within any sufficiently small region of D. Under the flat metric, any differences in the values of two measurements have the same significance over the whole of D. If x, x' are two measurements which are close together in D, then $x + y$, $x' + y$ are also close together for any choice of y in D. Under a different metric the distance between $x + y$ and $x' + y$ might depend on y.

2.3 Probability density function for the measurements

The probability density function $p(x|\theta)$ is constructed in two steps. The first step is to define $p(x|\tilde{x})$ where $\tilde{x}$ is a noise free or true measurement and x is the measurement obtained when $\tilde{x}$ is perturbed by noise. The second step is to combine the $p(x|\tilde{x})$ obtained as $\tilde{x}$ takes values in $M(\theta)$. The construction is described here for projective transformations of the line but in fact it applies much more widely. The key formulae, Eqs. (13) and (25), for the approximation K to the Fisher-Rao metric J hold for any family of conditional densities $p(x|\theta)$ depending smoothly on x, θ and for which the $M(\theta)$ are submanifolds of the measurement space.

Each measurement x is the result of a perturbation of an unknown true measurement $\tilde{x}$. The perturbation is modelled as a diffusion or heat flow (Chavel, 1984) in D and $p(x|\tilde{x})$ is obtained as a solution to the heat equation. Let s be the time parameter, let the initial $(s = 0)$ condition be the delta function centred at $\tilde{x}$ and let

$$(s, x) \mapsto p_s(x|\tilde{x}), \qquad\qquad s > 0, x \in D,$$

be the resulting solution to the heat equation on D. Let σ^2 be an estimate of the variance of the measurement errors. If $2t = \sigma^2$ and if t is small, then $p_t(x|\tilde{x})$ is closely approximated by a Gaussian density with expected value $\tilde{x}$ and variance σ^2. For this reason, $p(x|\tilde{x})$ is defined by $p(x|\tilde{x}) = p_t(x|\tilde{x})$. If t is large, then $p_t(x|\tilde{x})$ is no longer closely approximated by a Gaussian density, but in this application it is always assumed that t is small.

In order to combine the densities $p_t(x|\tilde{x})$, it is necessary to choose a distribution dh for $\tilde{x}$ on $M(\theta)$. If prior information about the distribution of $\tilde{x}$ on $M(\theta)$ is available, then dh

can incorporate this information. In the absence of any prior information, dh is obtained by normalizing the Riemannian metric induced on $M(\theta)$ by g.

Let $(s, x) \mapsto p_s(x|\theta)$ be the solution to the heat equation on D for which the initial condition at $s = 0$ is dh. The density $p(x|\theta)$ is defined by $p(x|\theta) = p_t(x|\theta)$. In effect, the true measurements $\tilde{x}$ are distributed on $M(\theta)$ with a density dh, and $p_t(x|\theta)$ is a weighted sum of contributions from all of the $p_t(x|\tilde{x})$. Further details are given in Maybank (2004), and a similar approach is described by Werman and Keren (1999).

The above choices of $p_t(x|\tilde{x})$ and dh follow a maximum entropy principle. If t is small then $p(x|\tilde{x})$ is very similar to the Gaussian density with variance $2t$. The Gaussian density has the maximum entropy for a given variance and a given expected value. In the case of dh, there is an analogy between dh and the uniform density which has the maximum entropy among all densities confined to a given subset of D with a finite, non-zero area. Let $\tilde{x}$ be given a uniform density on D and then let $\tilde{x}$ be conditioned to be in the open subset $U(\theta, \xi)$ of D consisting of those points strictly within a small distance $\xi > 0$ of $M(\theta)$. The set $U(\theta, \xi)$ is closely approximated as a Riemannian manifold by $M(\theta) \times (-\xi, \xi)$. The marginal density induced on $M(\theta)$ is approximately equal to dh and the approximation of the marginal density to dh becomes more accurate as $\xi \to 0$.

To simplify the notation, the subscript t is omitted from $p_t(x|\theta)$ in subsequent sections.

3 Fisher information

The notation ∂_{θ_i} is used for the partial derivative with respect to the component θ_i of θ. If f is a function of θ, then $\partial_{\theta_i} f = \partial f / \partial\theta_i$ and $\partial_\theta f$ is the vector with components $\partial_{\theta_i} f$. The double derivative of f is $\partial^2_{\theta_i,\theta_j} f$. Similar remarks apply to $\partial_x f$.

The Riemannian metric g defines an inner product on the tangent space $T_x D$ of D at x. Let u, v be tangent vectors in $T_x D$. The inner product of u, v under g is written as $g_x(u, v)$. The geodesic distance between points x, y of D is written as $d_g(x, y)$. The flat torus D is geodesically complete, thus $d_g(x, y)$ is the length of a shortest geodesic of D that starts at x and finishes at y (Gallot et al., 1990). The metric g is flat, thus if $d_g(x, y)$ is sufficiently small, then y can be identified with a vector u in $T_x D$, and $d_g(x, y)^2 = g_x(u, u)$. The tangent space $T_x D$ is given Euclidean coordinates and a Euclidean norm $u \mapsto \|u\|$ such that $g_x(u, u) = \|u\|^2$.

3.1 Asymptotic approximation to the Fisher information

The Fisher information $J(\theta)$ at θ in T is the 3×3 matrix $J(\theta)$ defined by

$$J_{ij}(\theta) = -\int_D \left(\partial^2_{\theta_i,\theta_j} \ln p(y|\theta)\right) p(y|\theta)\, dy, \qquad 1 \leq i, j \leq 3, \theta \in T. \tag{7}$$

Let $w(y, \theta)$ be the function defined by

$$w(y, \theta) = \min\{d_g(x, y) \mid x \in M(\theta)\}, \qquad y \in D, \theta \in T. \tag{8}$$

The manifold D is compact and $M(\theta)$ is a closed submanifold of D, thus there exists a point $x(y)$ in $M(\theta)$ such that $w(y, \theta) = d_g(x(y), y)$.

It can be shown that the leading order term in the asymptotic expansion of $\ln p(y|\theta)$ as a function of t is given by

$$\ln p(y|\theta) \sim -\frac{1}{4t} w(y,\theta)^2 + O(t^0). \tag{9}$$

The notation $O(t^0)$ is used rather than $O(1)$ in order to emphasize the role of t in the approximation to $\ln p(y|\theta)$. It follows from Eqs. (7) and (9) that

$$J_{ij}(\theta) = \frac{1}{4t} \int_D \left(\partial^2_{\theta_i,\theta_j} w(y,\theta)^2\right) p(y|\theta)\, dy + O(t^0), \qquad 1 \le i,j \le 3, \theta \in T. \tag{10}$$

At small values of t, the term $p(y|\theta)\, dy$ in Eq. (10) can be approximated by the measure dh on $M(\theta)$. On carrying out this approximation, Eq. (10) reduces to

$$J_{ij}(\theta) = \frac{1}{4t} \int_{M(\theta)} \left(\partial^2_{\theta_i,\theta_j} w(y,\theta)^2\right)_{y=x} dh(x) + O(t^0). \tag{11}$$

Let $u(y)$ be the vector in the tangent space $T_y D$ identified with the point $y - x(y)$ of D. It follows that

$$w(y,\theta)^2 = g_y(u(y), u(y)),$$

thus

$$J_{ij}(\theta) = \frac{1}{4t} \int_{M(\theta)} \left(\partial^2_{\theta_i,\theta_j} g_y(u(y), u(y))\right)_{y=x} dh(x) + O(t^0), \qquad 1 \le i,j \le 3, \theta \in T. \tag{12}$$

Let $K(\theta)$ be the 3×3 matrix defined by

$$K_{ij}(\theta) = \frac{1}{4t} \int_{M(\theta)} \left(\partial^2_{\theta_i,\theta_j} g_y(u(y), u(y))\right)_{y=x} dh(x), \qquad 1 \le i,j \le 3, \theta \in T. \tag{13}$$

It follows from Eqs. (12) and (13) that $K(\theta)$ is the leading order term in the asymptotic expansion of $J(\theta)$.

In the application of Eq. (13) to projective transformations of the line $M(\theta)$ is a hypersurface (codimension 1 submanifold) of D defined by an equation $f(x,\theta) = 0$ where f is given by Eq. (6). Let y be a point of D near to $M(\theta)$ and let x be the nearest point of $M(\theta)$ to y. Then $x = y + u(y)$ where to a first approximation, $u(y)$ is parallel to $\partial_y f(y,\theta)$. It follows that

$$f(y + u(y), \theta) = f(x,\theta) = 0. \tag{14}$$

On taking a Taylor expansion of Eq. (14) about y it follows that

$$f(y,\theta) + u(y).\partial_y f(y,\theta) = O(\|u(y)\|^2),$$

thus

$$u(y) = \frac{-f(y,\theta)\partial_y f(y,\theta)}{\|\partial_y f(y,\theta)\|^2} + O(\|u(y)\|^2),$$

and

$$g_y(u(y), u(y)) = \|u(y)\|^2 = \frac{f(y,\theta)^2}{\|\partial_y f(y,\theta)\|^2} + O(\|u(y)\|^3), \tag{15}$$

It follows from Eqs. (13) and (15) that

$$K_{ij}(\theta) = \frac{1}{4t}\int_{M(\theta)} \left(\frac{\partial^2_{\theta_i,\theta_j} f(y,\theta)^2}{g_y(\partial_y f(y,\theta), \partial_y f(y,\theta))}\right)_{y=x} dh(x), \qquad 1 \le i,j \le 3, \theta \in T. \tag{16}$$

Let N be the number of measurements. The matrix $N\,K(\theta)$ is closely related to the matrix $\bar{\boldsymbol{M}}_\infty$ defined in Section 14.4 of Kanatani (1996). Kanatani identifies the inverse of $\bar{\boldsymbol{M}}_\infty$ as an asymptotic approximation, as $N \to \infty$, to the Cramer-Rao lower bound for the covariance of an unbiased estimate of θ.

3.2 The Fisher-Rao metric as a function of ϕ

It is shown that $J(\theta)$ and $K(\theta)$ are independent of the first two components of $\theta = (a, b, \phi)$.

It is convenient to gather a, b into a vector $z = (a, b)$. The function $x \mapsto x - z$ defines an isometry on D, in that $d_g(x, y) = d_g(x - z, y - z)$. Let $\jmath : (0, \pi/4) \mapsto T$ be the embedding $\jmath(\phi) = (0, 0, \phi)$. It follows from (6) that under the isometry $x \mapsto x - z$ the submanifold $M(\theta)$ is mapped to $M(\jmath(\phi))$. The measure dh is induced on $M(\theta)$ by g, thus

$$dh(x, (a, b, \phi)) = dh(x - z, \jmath(\phi)). \tag{17}$$

Let $k_s(x, y)$ be the heat kernel for D (Chavel, 1984). The function $(s, x) \mapsto k_s(x, y)$ is the solution to the heat equation on D for which the initial condition at $s = 0$ is a delta function at y. The heat kernel $k_s(x, y)$ satisfies

$$k_s(x, y) = k_s(x + z, y + z), \qquad x, y \in D, 0 < s. \tag{18}$$

It follows from the properties of the heat kernel and the definition of $p_s(x|\theta)$ as a solution to the heat equation that

$$p(x|\theta) = \int_{y\in D} k_t(x, y)\, dh(y, \theta). \tag{19}$$

It follows from Eqs. (17) and (19) that

$$\begin{aligned} p(x|\theta) &= \int_{y\in D} k_t(x - z, y)\, dh(y, \jmath(\phi)), \\ &= p(x - z|\jmath(\phi)). \end{aligned} \tag{20}$$

It follows from Eqs. (7) and (20) that

$$\begin{aligned} J_{ij}(\theta) &= -\int_D (\partial^2_{\theta_i,\theta_j} \ln p(x|\theta)) p(x|\theta)\, dx, \\ &= -\int_D \left(\partial^2_{\theta_i,\theta_j} \ln p(x - z|\jmath(\phi))\right) p(x - z|\jmath(\phi))\, dx, \\ &= -\int_D \left(\partial^2_{\theta_i,\theta_j} \ln p(x|\jmath(\phi))\right) p(x|\jmath(\phi))\, dx, \\ &= J_{ij}(\jmath(\phi)), \qquad 1 \le i, j \le 3, \theta \in T. \end{aligned}$$

The result $K(\theta) = K(\jmath(\phi))$ is proved using Eqs. (16), (17) and the identity

$$f(x, (a, b, \phi)) \equiv f(x - z, \jmath(\phi)), \qquad x \in D, (a, b, \phi) \in T,$$

which follows immediately from the definition Eq. (6) of f.

3.3 Expression for $K(\theta)$

It is assumed from now on that $a = b = 0$, or equivalently $\theta = \jmath(\phi)$. It follows from Eq. (6) that the partial derivatives of f are

$$
\begin{aligned}
(\partial_\theta f)_{a=0,b=0} &= \left(\frac{\tan(\phi)\sec^2(x_1)}{\tan^2(x_1)+\tan^2(\phi)}, -1, \frac{\sec^2(\phi)\tan(x_1)}{\tan^2(x_1)+\tan^2(\phi)}\right), \\
(\partial_x f)_{a=0,b=0} &= \left(-\frac{\tan(\phi)\sec^2(x_1)}{\tan^2(x_1)+\tan^2(\phi)}, 1\right). \qquad (21)
\end{aligned}
$$

Eq. (16) for $K(\theta)$ is simplified as follows. Let $x = (x_1, x_2)$. On the curve $f(x, \theta) = 0$ the differentials dx_1, dx_2 satisfy

$$
(\partial_{x_1} f)\, dx_1 + (\partial_{x_2} f)\, dx_2 = 0. \qquad (22)
$$

Let ds be the arc length measure on $M(\theta)$ induced by the metric g on D,

$$
ds = (dx_1^2 + dx_2^2)^{1/2}. \qquad (23)
$$

Let $V(\theta)$ be the 1-dimensional volume of $M(\theta)$ under ds, ie. $V(\theta)$ is the length of $M(\theta)$. The length $V(\theta)$ is finite because $M(\theta)$ is compact. The measure dh in Eq. (16) is given by $dh = V(\theta)^{-1}\, ds$. It follows from Eqs. (22) and (23) that

$$
\begin{aligned}
ds &= \left(1 + \left(\frac{dx_2}{dx_1}\right)^2\right)^{1/2} dx_1, \\
&= \|\partial_x f\| (\partial_{x_2} f)^{-1} dx_1, \\
&= \|\partial_x f\|\, dx_1. \qquad (24)
\end{aligned}
$$

It follows from Eqs. (16), (24) and the choice of dh that

$$
K_{ij}(\theta) = \frac{1}{2tV(\theta)} \int_{M(\theta)} \frac{(\partial_{\theta_i} f)(\partial_{\theta_j} f)}{\|\partial_x f\|}\, dx_1, \qquad 1 \le i, j \le 3, \theta \in T. \qquad (25)
$$

Equations similar in form to Eq. (25) hold for a wide range of detection problems.

3.4 The entries $K_{11}(\theta)$, $K_{22}(\theta)$ of $K(\theta)$

Simple expressions are obtained for the entries $K_{11}(\theta)$, $K_{22}(\theta)$ of K. It follows from Eqs. (21), (24) and (25) that

$$
K_{11}(\theta) + K_{22}(\theta) = \frac{1}{2tV(\theta)} \int_{M(\theta)} \frac{(\partial_a f)^2 + (\partial_b f)^2}{\|\partial_x f\|^2}\, dx_1 = \frac{1}{2t}. \qquad (26)
$$

It is shown next that $K_{11}(\theta) = K_{22}(\theta)$. It follows from Eq. (21) that

$$
K_{22}(\theta) = \frac{1}{t\, V(\theta)} \int_0^{\pi/2} \frac{dx_1}{\|\partial_x f\|}. \qquad (27)
$$

The entries of $K(\theta)$ depend on the function $\theta \mapsto M(\theta)$, but they do not depend on the choice of f. The function f can be replaced by any other smooth function which has $M(\theta)$ as its zero set and which has a non zero gradient at each point of $M(\theta)$. Let $\tilde{f}$ be the function defined by

$$\tilde{f}(x,\theta) = x_1 - a - \tan^{-1}(\tan(\phi)\tan(x_2 - b)).$$

The function $x \mapsto \tilde{f}(x,\theta)$ is zero on $M(\theta)$. It follows that

$$K_{11}(\theta) = \frac{1}{t\,V(\theta)} \int_0^{\pi/2} \frac{dx_2}{\|\partial_x \tilde{f}\|}. \tag{28}$$

A short calculation shows that

$$(\partial_x \tilde{f})_{a=0,b=0} = \left(1, \frac{\tan(\phi)\mathrm{cosec}^2(x_2)}{\cot^2(x_2) + \tan^2(\phi)}\right). \tag{29}$$

It follows from (21) and (29) that

$$\|(\partial_x \tilde{f}(x_2))_{a=0,b=0}\| = \|(\partial_x f(\pi/2 - x_2))_{a=0,b=0}\|, \tag{30}$$

It follows from Eqs. (27), (28) and (30) that $K_{11}(\theta) = K_{22}(\theta)$. The values of $K_{11}(\theta)$, $K_{22}(\theta)$ are now obtained by applying Eq. (26),

$$K_{11}(\theta) = K_{22}(\theta) = \frac{1}{4t}. \tag{31}$$

3.5 The remaining entries of $K(\theta)$

The integrals in Eq. (25), including the integral for $V(\theta)$, are evaluated in terms of elliptic integrals. In order to carry out these evaluations it is convenient to use the substitution $x_1 \mapsto x_1(\xi)$, where ξ is defined by

$$\xi = \tan^{-1}\left(\frac{\tan(\phi)\sec^2(x_1)}{\tan^2(x_1) + \tan^2(\phi)}\right), \qquad 0 \le x_1 \le \pi/2, 0 < \phi < \pi/4. \tag{32}$$

It follows from Eq. (32) that in the range $\phi \le \xi \le \pi/2 - \phi$

$$\begin{aligned}
\tan(x_1) &= \left(\frac{\tan(\phi)(1 - \tan(\xi)\tan(\phi))}{\tan(\xi) - \tan(\phi)}\right)^{1/2}, \\
\frac{d\xi}{dx_1} &= -2\sin(\xi)\left(\frac{\sin(2\xi) - \sin(2\phi)}{\sin(2\phi)}\right)^{1/2}.
\end{aligned}$$

On substituting for x_1 in Eq. (21) the following expressions for $\partial_\theta f$, $\partial_x f$ are obtained,

$$\begin{aligned}
(\partial_\theta f)_{a=0,b=0} &= \left(\tan(\xi), -1, \sec(2\phi)\sec(\xi)\left(\frac{\sin(2\xi) - \sin(2\phi)}{\sin(2\phi)}\right)^{1/2}\right), \\
(\partial_x f)_{a=0,b=0} &= (-\tan(\xi), 1).
\end{aligned} \tag{33}$$

The following expression for $V(\theta)$ is obtained,

$$\begin{aligned} V(\theta) &= 2\int_{x_1=0}^{x_1=\pi/2} ds \\ &= 2\int_0^{\pi/2} \|\partial_x f\|\, dx_1 \\ &= 2(\sin(2\phi))^{1/2}\int_{\phi}^{\pi/2-\phi} \frac{d\xi}{\sin(2\xi)(\sin(2\xi)-\sin(2\phi))^{1/2}}. \end{aligned} \tag{34}$$

Let F be the elliptic integral of the first kind, let K be the complete elliptic integral of the first kind and let Π be the elliptic integral of the third kind (Abramovitz and Stegun, 1965). Let m be defined by

$$m = \frac{1}{2}\left(1-\sin(2\phi)\right), \qquad 0 < \phi < \pi/4, \tag{35}$$

and note that with m given by Eq. (35), it follows from Abramovitz and Stegun (1965), Section 17.4.15 that

$$\begin{aligned} F\left(\pi/4-\phi|m^{-1}\right) &= m^{1/2}K(m), \\ \Pi\left(2,\pi/4-\phi|m^{-1}\right) &= m^{1/2}\Pi(2m|m). \end{aligned}$$

The integral in the third line of Eq. (34) is evaluated using Mathematica (Wolfram, 2003), to yield

$$\begin{aligned} V(\theta) &= 4(\sin(2\phi))\sec(\phi)\Pi(2,\pi/4-\phi|m^{-1}), \\ &= (8(1-2m))^{1/2}\Pi(2m|m). \end{aligned} \tag{36}$$

Similar calculations lead to the following expressions for the remaining entries of $K(\theta)$.

$$\begin{aligned} K_{12}(\theta) &= -\frac{K(m)}{4t\Pi(2m|m)}, \\ K_{13}(\theta) &= \left(4\sqrt{2}\,t(\sin(2\phi))^{1/2}\cos(2\phi)\Pi(2m|m)\right)^{-1}\ln\left(\frac{\operatorname{cosec}(\phi)+\cot(\phi)}{\sec(\phi)+\tan(\phi)}\right), \\ K_{23}(\theta) &= -K_{13}(\theta), \\ K_{33}(\theta) &= (t\,\sin(4\phi)\cos(2\phi))^{-1}\left(\frac{K(m)}{\Pi(2m|m)}-\sin(2\phi)\right). \end{aligned} \tag{37}$$

It is noted that $\left(1/\sqrt{2},1/\sqrt{2},0\right)^{\top}$ is an eigenvector of $K(\theta)$ with eigenvalue $(4t)^{-1}+K_{12}(\theta)$.

4 Volume of the Parameter Manifold

The metric K defines a volume measure $\tau(\theta)\,d\theta$ on T. This measure is sometimes referred to as the canonical measure defined by K (Gallot et al., 1990). The volume, $V(B,K)$, of any measurable subset B of T under the canonical measure is independent of the choice of parameterisation of T. The expression for the canonical measure is

$$\tau(\theta)\,d\theta = |\det(K(\theta))|^{1/2}\,d\theta.$$

The term $\tau(\theta)$ is the product of a function of ϕ and the term $t^{-3/2}$.

4.1 Calculation of $V(T, K)$

The volume $V(T, K)$ is found by integrating $\tau(\theta)\, d\theta$ over T. The integration of $\tau(\theta)$ over the components a, b of θ is straightforward because $\tau(\theta)$ does not depend on a or b. The integration of $\tau(\theta) = \tau(\jmath(\phi))$ over ϕ is more difficult.

It is necessary to prove that the integral of $\tau(\jmath(\phi))$ over ϕ is defined and finite. The components of $K(\theta)$ are continuous functions of ϕ for $0 < \phi < \pi/4$ and they are bounded for ϕ in any closed interval $[\delta_1, \pi/4 - \delta_2]$ where δ_1, δ_2 are small strictly positive constants. It follows that $\phi \mapsto \tau(\jmath(\phi))$ is continuous for $0 < \phi < \pi/4$ and bounded for ϕ in $[\delta_1, \pi/4 - \delta_2]$. Thus the integral of $\tau(\jmath(\phi))$ over $[\delta_1, \pi/4 - \delta_2]$ exists and is finite.

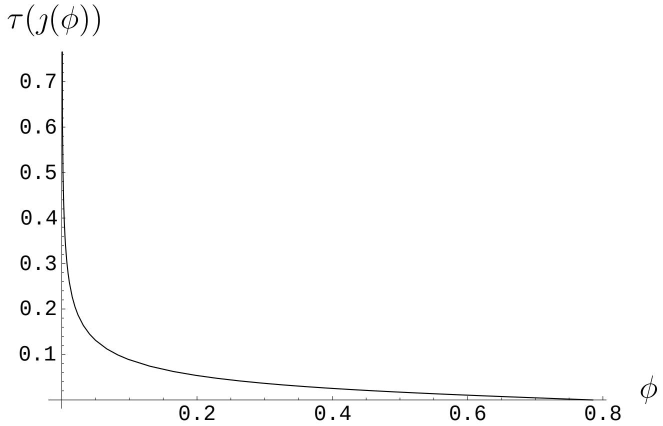

Figure 2. Graph of $\phi \mapsto \tau(\jmath(\phi))$ with $t = 1$.

It remains to check the behaviour of $\tau(\jmath(\phi))$ for ϕ near to 0 and for ϕ near to $\pi/4$. A graph of the function $\phi \mapsto \tau(\jmath(\phi))$ is shown in Fig. 2 with t set equal to the nominal value $t = 1$. The graph suggests that $\tau(\jmath(\phi))$ increases without bound as $\phi \to 0$ and that $\tau(\jmath(\phi))$ tends to zero as $\phi \to \pi/4$. To check these suggestions, suppose first that $\phi = \pi/4 - \delta$, where δ is a small strictly positive number. It follows that $m \equiv (1/2)(1 - \sin(2\phi)) = O(\delta^2)$. A series expansion of the coefficients of $K(\theta)$ yields

$$K(\jmath(\pi/4 - \delta)) = \frac{1}{4t} \begin{pmatrix} 1 & -1 & 2\pi^{-1} \\ -1 & 1 & -2\pi^{-1} \\ 2\pi^{-1} & -2\pi^{-1} & 2^{-1} \end{pmatrix} + O(\delta^2),$$

from which it follows that $\tau(\jmath(\pi/4 - \delta)) = O(\delta)$.

Suppose, secondly, that ϕ is near to 0. The expression for $\Pi(2m|m)$ in Abramovitz and Stegun (1965), Section 17.7.14 yields

$$\begin{aligned} \Pi(2m|m) &= \frac{\pi}{(2(1-2m))^{1/2}} + O(\phi^0), \\ &= \pi(2\sin(2\phi))^{-1/2}(1 + O(\phi)) + O(\phi^0), \\ &= \frac{\pi}{2\phi^{1/2}}(1 + O(\phi)) + O(\phi^0), \qquad 0 < \phi < \pi/4. \end{aligned} \tag{38}$$

The leading order term in Eq. (38) is obtained from the term $\pi\delta_2/2$ defined by Abramovitz and Stegun (1965). On using Mathematica to substitute the right hand side of Eq. (38) for $\Pi(2m|m)$ into $K(\theta)$ and taking a series expansion of $|\det(K(\theta))|^{1/2}$ about $\phi = 0$ the following approximation to $\tau(\jmath(\phi))$ is obtained,

$$\tau(\jmath(\phi)) = \frac{(K(1/2))^{1/2}}{4t^{3/2}\phi^{1/4}} + O\left((\ln\phi)^2\phi^{1/4}\right).$$

It follows from the above approximations to $\tau(\jmath(\phi))$ for ϕ near to 0 and ϕ near to $\pi/4$ and the continuity of $\tau(\jmath(\phi))$ in $[\delta_1, \pi/4 - \delta_2]$ that $\tau(\jmath(\phi))$ is bounded above on the open interval $(0, \pi/4)$ by an integrable function $C\phi^{-1/4}$ where C is a sufficiently large positive constant. The function $\tau(\jmath(\phi))$ is thus integrable over $0 < \phi < \pi/4$ and the value of the integral is finite.

The volume $V(T, K)$ is estimated by numerical integration of $\tau(\theta)\, d\theta$ over T,

$$\begin{aligned} V(T,K) &= \int_T \tau(\theta)\, da\, db\, d\phi, \\ &= \pi^2 \int_0^{\pi/4} \tau(\jmath(\phi))\, d\phi, \\ &= 0.349409...t^{-3/2}. \end{aligned} \tag{39}$$

4.2 Number of models

Some variations are possible in the exact number of models, depending on the details of the definitions employed. In all cases the main idea is to choose enough models in T to ensure that every $\theta \in T$ is close enough to at least one model.

Let $B(\theta) \subset T$ be defined by

$$B(\theta) = \{\psi \mid \psi \in T \text{ and } (\psi - \theta)^\top K(\theta)(\theta - \psi) \le 1\}. \tag{40}$$

The definition of $B(\theta)$ in Eq. (40) is closely linked to the Kullback-Leibler distance, $D(\theta||\psi)$, between the probability density functions $p(x|\theta)$ and $p(x|\psi)$ (Amari, 1985; Cover and Thomas, 1991), The distance $D(\theta||\psi)$ is defined by

$$D(\theta||\psi) = \int_D \ln\left(p(x|\theta)/p(x|\psi)\right) p(x|\theta)\, dx.$$

A Taylor series expansion of the function $\psi \mapsto D(\theta||\psi)$ about $\psi = \theta$ yields

$$D(\theta||\psi) = \frac{1}{2}(\psi - \theta)^\top K(\theta)(\psi - \theta) + O(\|\psi - \theta\|^3). \tag{41}$$

It follows from Eqs. (40) and (41) that

$$B(\theta) \approx \{\psi \mid \psi \in T \text{ and } D(\theta||\psi) \le 1/2\}.$$

Thus the average value of the likelihood ratio $\ln\left(p(x|\theta)/p(x|\psi)\right)$ is less than or equal to $1/2$ for $\psi \in B(\theta)$. The set $B(\theta)$ consists of those ψ for which $p(x|\psi)$ is difficult to distinguish from $p(x|\theta)$ given a single measurement x. The definition Eq. (40) answers question *(i)* in Section 1.

Let $u = K(\theta)^{1/2}(\theta - \psi)$. It follows from Eq. (40) and the definition of $V(B(\theta), K)$ that

$$\begin{aligned} V(B(\theta), K) &= \int_{B(\theta)} \tau(\theta)\, d\theta, \\ &\approx \int_{u.u \le 1} du, \\ &\approx 4\pi/3. \end{aligned}$$

The number $n(T, K)$ of models is defined by

$$\begin{aligned} n(T, K) &= V(T, K)/(4\pi/3), \\ &= c_T t^{-3/2}. \end{aligned} \tag{42}$$

where c_T is a numerical constant. A calculation with Mathematica shows that

$$c_T = 0.0294918\ldots.$$

The number $n(T, K)$ of models is a measure of the complexity of the task of finding values of θ for which $p(x|\theta)$ is compatible with a given measurement x. This task can be solved simply by searching a set of $O(n(T, K))$ candidate models.

If t is very small, then $n(T, K)$ may be so large that it is no longer feasible to check every one of the $O(n(T, K))$ models. In such cases the following multiresolution method can be used: let $t' > t$ be a nominal noise level and let K' be the approximation to the Fisher-Rao metric corresponding to t'. The parameter space T is sampled at $O(n(T, K'))$ points $\theta(i)$. The points $\theta(i)$ with the largest number of inliers are identified and the subsets $B(\theta)$ of T sampled at the higher resolution defined by t. Further details are given in Section 5.2 below.

5 Detection of projective transformations of the line

Let $S = \{x(1), \ldots, x(N)\}$ be a set of measurements and let r be an integer threshold. The projective transformation θ of the line is detected if there are r or more measurements compatible with θ. In applications it is often the case that $r < N$. The reason is that S may contain measurements which are due to random noise or which are associated with different projective transformations or with other image structures. The term 'compatible' has the following meaning: a measurement $x(i)$ is compatible with θ if $f(x(i), \theta(i)) = 0$ for some $\theta(i) \in B(\theta)$. Such measurements are also referred to as inliers. The measurements not compatible with θ are classified as outliers (Torr and Murray, 1993).

The aim of this section is to describe an algorithm in which the task of detecting a projective transformation of the line reduces to a search through a finite set of points in T. This will answer question *(ii)* in Section 1.

The algorithm, as given below in Sections 5.2-5.4, has some similarity to the Hough transform (Forsyth and Ponce, 2003; Gonzalez and Woods, 2002) in that T is analogous to the accumulator space for the Hough transform and $B(\theta)$ is analogous to a single accumulator. The main difference between this algorithm and the Hough transform is the use of the Fisher-Rao metric on T to provide a statistically sound method for deciding if two points θ, ψ of T should be counted as a single model.

5.1 Discrete approximation to the parameter manifold

A simple method is given for sampling T at a finite set of points G. The main condition on G is that every set $B(\theta) \subset T$ should contain one or more points of G. Ideally, the size $|G|$ of G should be of the same order as the number $n(T, K)$ of models. In fact G is chosen such that $|G|$ is significantly larger than $n(T, K)$, in order to ensure that there is a high probability that every set $B(\theta)$ contains at least one element of G.

The set G is constructed as follows. Small strictly positive numbers Δ_a, Δ_b, Δ_ϕ are chosen. The manifold T is divided into cuboids of size $\Delta_a \times \Delta_b \times \Delta_\phi$ in a, b, ϕ space. There are approximately $\pi^3/(4\Delta_a\Delta_b\Delta_\phi)$ such cuboids. Then points are chosen within each cuboid. The decision on the number of points to choose in a given cuboid is made randomly in the following way. First, define α by

$$\alpha = -\ln\left(1 - (95/100)^{1/n(T,K)}\right). \tag{43}$$

Let c be any one of the cuboids, let $\theta(c)$ be the point of T at the centre of c and define $n(c)$ by

$$n(c) = \frac{\alpha n(T, K)\tau(\theta(c))\Delta_a\Delta_b\Delta_\phi}{V(T, K)}.$$

A number $u(c)$ is chosen randomly and uniformly in $[0, 1]$. If $u(c) \leq n(c) - \lfloor n(c) \rfloor$, then $\lfloor n(c) \rfloor$ points are chosen randomly and uniformly in c and added to G. If $u(c) > n(c) - \lfloor n(c) \rfloor$, then $\lceil n(c) \rceil$ points are chosen randomly and uniformly in c and added to G. The set G is the union of points chosen in this way over all the cuboids c. It turns out that in most cases $\lfloor n(c) \rfloor = 0$.

The value Eq. (43) of α is justified as follows. The number of cuboids c in each set $B(\theta)$ is approximately

$$\frac{V(B(\theta), K)}{\tau(\theta(c))\Delta_a\Delta_b\Delta_\phi} = \frac{V(T, K)}{n(T, K)\tau(\theta(c))\Delta_a\Delta_b\Delta_\phi} = \alpha n(c)^{-1}.$$

The probability that none of the cuboids in $B(\theta)$ contains a point of G is approximately

$$(1 - n(c))^{\alpha n(c)^{-1}} \approx e^{-\alpha}.$$

It is assumed, as a first approximation, that the infinite number of sets $B(\theta)$, $\theta \in T$ can be replaced by a finite set of $n(T, K)$ representatives. The probability that at least one of the representatives $B(\theta)$ contains no point of G is approximately

$$1 - (1 - e^{-\alpha})^{n(T,K)}. \tag{44}$$

The expression Eq. (43) for α is obtained by setting Eq. (44) equal to 5/100.

The advantage of this method of choosing G is that it can be implemented easily. The density of the points of G in T is approximately proportional to the canonical density $\tau(\theta)\, d\theta$, but with some random variations. The coverage of T by G was tested by selecting points θ uniformly in $[0, \pi) \times [0, \pi) \times (0, \pi/4)$, and testing to see if $\theta \in B(\theta(i))$ for some $\theta(i) \in G$. In a typical test, 50 points were selected. Every point was contained in at least one of the $B(\theta(i))$ for $\theta(i) \in G$.

The above algorithm for constructing G was implemented in Mathematica, version 5 (Wolfram, 2003), with $t = 1/1000$ and $\Delta_a = \Delta_b = \Delta_\phi = \sqrt{t}$. The following results were obtained.

$$\begin{aligned} \lfloor n(c) \rfloor &= 0, \qquad \text{for all } c, \\ \alpha &= 9.80821\ldots, \\ |G| &= 12,181. \end{aligned}$$

The size of G can be reduced by making better use of the Riemannian geometry of T, but this topic is beyond the scope of this paper.

5.2 Multiresolution method

In this subsection and the following subsections the dependence of entities on the noise level t is included in the notation where appropriate. For example, $G(t)$ is the set G constructed as in Section 5 and $B(t, \theta)$ is the set defined by (40).

If the noise level t is not too small, then it is straightforward to search for a projective linear transformation of the line which fits a given set of measurements. The parameter space T is sampled at a finite set $G(t)$ of points and each $\theta \in G(t)$ is checked to see if it is supported by a sufficient number of the measurements.

If t is very small, then the size $|G(t)|$ of $G(t)$ is very large and it is no longer possible to check all the elements of $G(t)$ in a reasonable time. The problem is solved using a multiresolution approach. Let t_2 be a given small value for the noise level t and let $t_1 > t_2$ be a value of t for which $G(t_1)$ is small enough to allow the checking of all its elements in a reasonable time. Every measurement which is an inlier for θ at resolution t_2 is also an inlier for θ at the lower resolution (or higher noise level) t_1.

Let $\theta(i)$, $1 \leq i \leq m$ be the elements in $G(t_1)$ which are detected at resolution t_1. Each subset $B(t_1, \theta(i))$, $1 \leq i \leq m$ of T is searched at resolution t_2. The search at resolution t_2 is simplified by assuming that t_1 is small enough to ensure that the variation in $K(t_2, \theta)$ as θ ranges over $B(t_1, \theta(i))$ is negligible. Let H be the symmetric positive 3×3 matrix defined by $H = K(t_1, \theta(i))^{1/2}$. The function $\theta \mapsto f(\theta) \equiv H(\theta - \theta(i))$ maps $B(t_1, \theta(i))$ to the ball $C(1, 0)$ in $\mathbb{R}^3$ with radius 1 and centred at the origin 0. The metric pushed forward from $B(t_1, \theta(i))$ to $C(1, 0)$ by f coincides with the Euclidean metric on $\mathbb{R}^3$.

The image of $B(t_2, \theta(i))$ under f is the ball $C((t_2/t_1)^{1/2}, 0)$. Let Δ be the length of the sides of a cube inscribed in $C((t_2/t_1)^{1/2}, 0)$. It follows from Pythagoras' theorem that $\Delta = 2(t_2/(3t_1))^{1/2}$. Let $W(i)$ be the set of of all points in $C(1, 0)$ of the form $(i\Delta, j\Delta, k\Delta)$, where i, j, k are integers. The set $B(t_1, \theta(i))$ is sampled at the set $f^{-1}(W(i)) \cap T$. Any elements of $\cup_{i=1}^m f^{-1}(W(i))$ with a sufficient number of inliers are detected at resolution t_2.

5.3 False detection

In some cases a projective transformation of the line is detected even though the measurements do not arise from any such transformation. For example, if the measurements are sampled randomly and uniformly from D, then there is a small but non-zero probability that the measurements will be inliers to a projective transformation of the line. In these circumstances a false detection is said to occur. The probability of a false detection is reduced by using a threshold $r(N)$ which depends on the total number N of measurements. A projective transformation of the line is detected only if it has $r(N)$ or more inliers.

The best value of $r(N)$ depends on the nature of the measurements. A useful default is to choose $r(N)$ large enough to ensure that there is only a small probability of a false detection when the measurements are randomly and uniformly distributed in D. The threshold $r(N)$ should not be too large, otherwise projective transformations of the line may not be detected even though they are present.

5.4 Algorithms

In this subsection three algorithms are described: Inliers, CoarseSearch and Refine. Inliers finds those measurements inlying to a given $\theta \in T$. CoarseSearch implements the low resolution search at noise level t_1. Refine implements the high resolution search at a noise level $t_2 \ll t_1$.

Each measurement x contains two components, $x = (x_1, x_2)$ where x_1 is in the domain of the projective transformation and x_2 is in the range of the projective transformation. Let $S(1)$ be the set of components in the domain and let $S(2)$ be the set of components in the range. If an element $x_1(i)$ of $S(1)$ is known to correspond to an element $x_2(j)$ in $S(2)$, $x_1(i) \leftrightarrow x_2(j)$, then $(x_1(i), x_2(j))$ is an element of the set S of measurements.

The correspondence between $S(1)$ and $S(2)$ is not assumed to be known and it is not assumed that each point of $S(1)$ corresponds to a point of $S(2)$. It is assumed that the points of $S(1)$, $S(2)$ are ordered and that the correct correspondences preserve the order. More formally, if $x_1(i)$, $x_1(j)$ are in $S(1)$, if $x_2(k)$, $x_2(l)$ are in $S(2)$, if the correspondences $x_1(i) \leftrightarrow x_1(j)$, $x_2(k) \leftrightarrow x_2(l)$ are correct and if $i \leq j$, then $k \leq l$.

The algorithm Inliers requires the distance function $x \mapsto w(x, \theta)$ defined by Eq. (8) or at least an approximation to $w(x, \theta)$ accurate near to $M(\theta)$. It takes as input t, an element θ of T and the sets $S(1)$, $S(2)$. It returns the set $C(\theta)$ of inliers for θ. Each element of $C(\theta)$ is a pair $(x_1(i), x_2(j))$, for which the correspondence $x_1(i) \leftrightarrow x_2(j))$ is found by the algorithm Inliers.

Algorithm: Inliers

Inputs: t, θ, $S(1)$, $S(2)$;
Outputs: $C(\theta)$;

1. $\sigma \leftarrow (2t)^{1/2}$;
2. $C(\theta) \leftarrow \emptyset$;
3. for($i = 1$; $i \leq |S(1)|$; $i++$)
   - 3.1. $j_m = \text{argmin}\{j \mapsto w((x_1(i), x_2(j)), \theta)$ where $x_2(j) \in S(2)\}$;
   - 3.2. if $w((x_1(i), x_2(j_m)), \theta) \leq 2\sigma$,
     - 3.2.1. $C(\theta) \leftarrow C(\theta) \cup \{(x_1(i), x_2(j_m))\}$;
     - 3.2.2. $S(2) \leftarrow S(2) \setminus \{x_2 \mid x_2 \in S(2)$ and $x_2 \leq x_2(j_m)\}$;

3.3. endif;

4. endfor;

5. Output $C(\theta)$;

6. Halt.

CoarseSearch takes as inputs a value t_1 for the noise level, a set $G(t_1)$ sampled from the parameter space T, a set $S(1)$ of measurement components from the domain and a set $S(2)$ of measurement components from the range. The quantity $2t_1$ is the variance of the measurement errors in the domain and the range. CoarseSearch outputs a projective transformation of the line $\theta(i_m)$ chosen from $G(t_1)$ and a set $C(\theta(i_m)) \subset S(1) \times S(2)$ consisting of inliers to $\theta(i_m)$. The criterion for choosing $\theta(i_m)$ is that the size $|C(\theta(i_m))|$ is the largest among the sizes of the different sets of inliers, $C(\theta')$, $\theta' \in G(t_1)$, examined by CoarseSearch.

Algorithm: CoarseSearch

Inputs: t_1, $G(t_1)$, $S(1)$, $S(2)$;
Outputs: $\theta(i_m)$, $C(\theta(i_m))$;

1. $i_m = \text{argmax}\{i \mapsto |\text{Inliers}(t, \theta(i), S(1), S(2))| \text{ where } \theta(i) \in G(t_1)\}$;

2. $C(\theta(i_m)) = \text{Inliers}(t, \theta(i_m), S(1), S(2))$;

3. Output $\theta(i_m)$, $C(i_m)$;

4. Halt.

The algorithm Refine takes as input two values $t_1 > t_2$ for the noise level, a single point θ in T and the two sets $S(1)$, $S(2)$ of components of measurements from the range and the domain of the projective linear transformation. It outputs a point $\theta(i_m)$ sampled from the region of T near to θ and a set $C(\theta(i_m))$ of inliers for $\theta(i_m)$. The criterion for choosing $\theta(i_m)$ is that $|C(\theta(i_m))|$ is the largest among the sizes of the different sets of inliers examined by Refine.

Algorithm: Refine

Inputs: t_1, t_2, θ, $S(1)$, $S(2)$;
Outputs: $\theta(i_m)$, $C(\theta(i_m))$;

1. $H \leftarrow K(t_1, \theta)^{1/2}$;

2. $\Delta \leftarrow 2(t_1/(3t_2))^{1/2}$;

3. $W \leftarrow \{z \mid z = (i\Delta, j\Delta, k\Delta) \text{ where } i, j, k \text{ are integers and } z.z \le 1\}$;

4. $G(\theta, t_1, t_2) \leftarrow \theta + H^{-1}(W) \cap T$;

5. $i_m = \text{argmax}\{i \mapsto |\text{Inliers}(t_2, \theta(i), S(1), S(2))| \text{ where } \theta(i) \in G(\theta, t_1, t_2)\}$;

6. $C(\theta(i_m)) = \text{Inliers}(t_2, \theta(i_m), S(1), S(2))$;

7. Output $\theta(i_m)$, $C(i_m)$;

8. Halt.

6 Experiments

6.1 Images and preprocessing

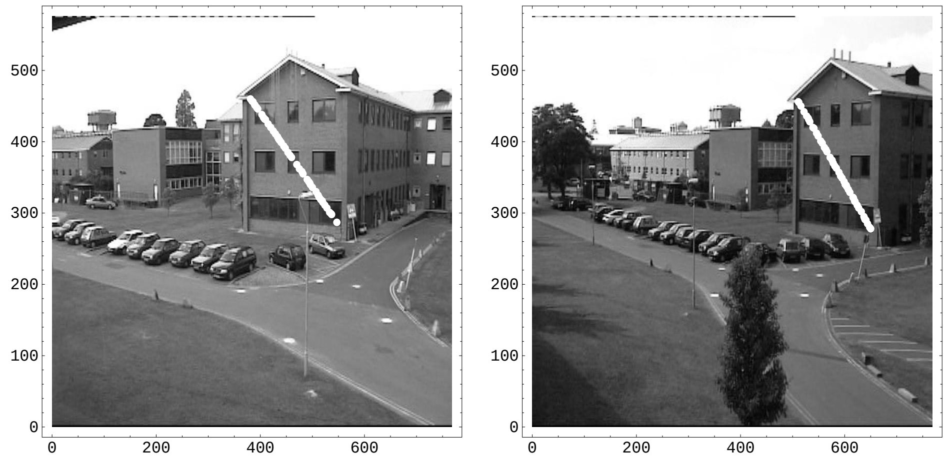

Figure 3. Left: domain and right: range of the projective transformation of the line.

The algorithms described in Section 5.4 were applied to the measurements obtained from the regions covered by the two white lines in the images shown in Fig. 3. The original colour images were taken from the PETS'2001 data base (Ferryman, 2001) and converted to grey scale using the Microsoft Photo Editor. The left hand image was obtained from number 0017, Camera 1. The right hand image was obtained from number 0017, Camera 2. The start point and the end point of the white line in each image were chosen by hand.

Each 2D grey level image was smoothed using the $n \times n$ mask $u \otimes u$ where $u_i = 2^{-n+1}\binom{n-1}{i}$, $0 \leq i \leq n-1$. The mask is a discrete approximation to the 2D Gaussian density with covariance $((n-1)/4)I$ where I is the 2×2 identity matrix. In these experiments $n = 7$. The Sobel edge operator was applied and the gradient magnitudes calculated at each pixel. Let $p(1)$, $q(1)$ be the start and end points of the line in the first image, and let v be a unit vector in the direction $p(1)$ to $q(1)$. The Sobel magnitudes were read into an array $X(1)$ defined such that the ith entry $X_i(1)$ is equal to the Sobel magnitude at the pixel Round$[p(1) + iv]$, $0 \leq i \leq \|q(1) - p(1)\|$. The measurements in the first image were the local maxima of $X(1)$. An integer i specifies the location of a local maximum in $X(1)$ if $X_i(1) > X_{i-1}(1)$ and $X_i(1) > X_{i+1}(1)$. The Sobel magnitudes $X(2)$ for the line in the second image were obtained in a similar way. The measurements are shown in the two images as small white dots which merge together to give the appearance of lines. The two arrays $X(1)$, $X(2)$ of Sobel magnitudes are shown as graphs in Fig. 4.

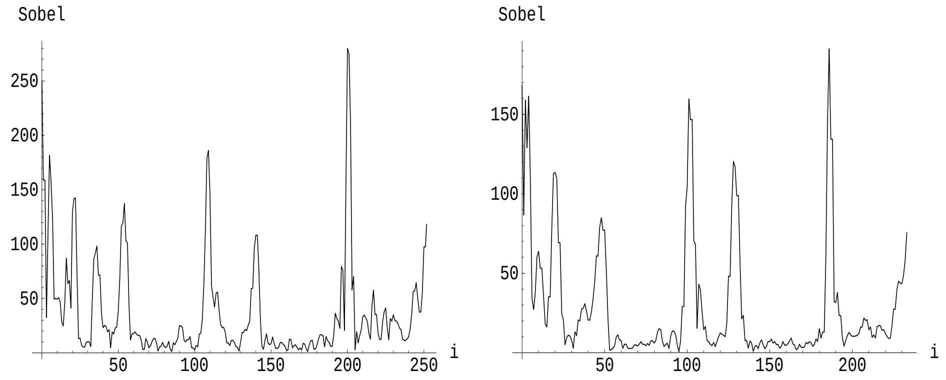

Figure 4. Sobel magnitudes from left: the domain, $i \mapsto X_i(1)$ and right: the range, $i \mapsto X_i(2)$.

In order to apply the algorithms in Section 5.4, the measurements on each line were mapped to the arc of a circle. Let $L(1)$, $L(2)$ be the lengths of the two lines. The mappings were

$$\begin{aligned} X_i(1) &\mapsto x_i(1) \equiv \tan^{-1}(2(X_1(i)/L(1)) - 1), \\ X_i(2) &\mapsto x_i(2) \equiv \tan^{-1}(2(X_i(2)/L(2)) - 1). \end{aligned}$$

It follows that $-\pi/4 \leq x_i(1), x_i(2) \leq \pi/4$.

With the chosen value $n = 7$, the Gaussian smoothing of each line has variance $(n-1)/4 = 3/2$. The standard deviations of the noise of the measurements $x_i(1)$, $x_i(2)$ on the unit circle were estimated at $(3/2)^{1/2}\pi/(2\min\{L(1), L(2)\})$. The resulting value for t_2 is

$$t_2 = \frac{3\pi^2}{16\min\{L(1), L(2)\}^2}$$

In the experiments, $L(1) = 251.8$, $L(2) = 233.0$ and $t_2 = 3.41 \times 10^{-5}$. The value of t_1 was set at $t_1 = 10^{-3}$. The number of elements in $G(t_1)$ was 12,236. If one of S(1), S(2) has more elements than the other, then the elements with the lowest Sobel magnitudes are discarded, until $S(1)$, $S(2)$ have the same number of elements. In this experiment, $|S(1)| = |S(2)| = 45$.

6.2 Results

After applying the algorithms CoarseSearch and Refine to $S(1)$, $S(2)$, 39 inliers were found and the projective transformation $i \mapsto k(i)$ from $X(1)$ to $X(2)$ was estimated to be

$$k(i) = \frac{-10.7531 + 0.998851\, i}{0.919794 + 0.000628765\, i}. \tag{45}$$

The projective transformation in Eq. (45) is nearly affine because the coefficient of i in the denominator is very small. The effect of the transformation $i \mapsto k(i)$ on the Sobel magnitudes in the array $X(1)$ is shown in Fig. 5. The left hand graph in Fig.

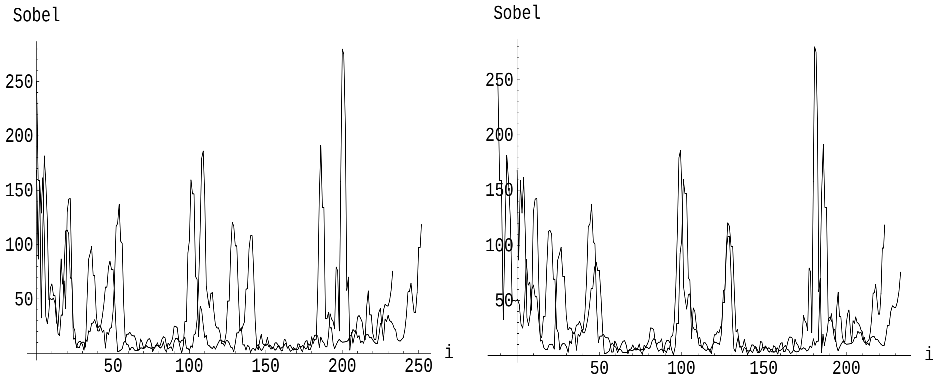

Figure 5. Left: Superposition of Sobel magnitudes $i \mapsto X_i(1)$ and $i \mapsto X_i(2)$. Right: Superposition of Sobel magnitudes from the range, $i \mapsto X_2(i)$ and Sobel magnitudes mapped from the domain to the range, $i \mapsto X_{k^{-1}(i)}(1)$ where k is given by Eq. (45).

5 shows the Sobel magnitudes from the range superposed onto the Sobel magnitudes from the domain. The right hand graph in Fig. 5 shows the Sobel magnitudes from the range superposed on the Sobel magnitudes mapped from the domain to the range by the projective transformation in Eq. (45). The application of Eq. (45) improves the matching between the peaks in the two sets of Sobel magnitudes.

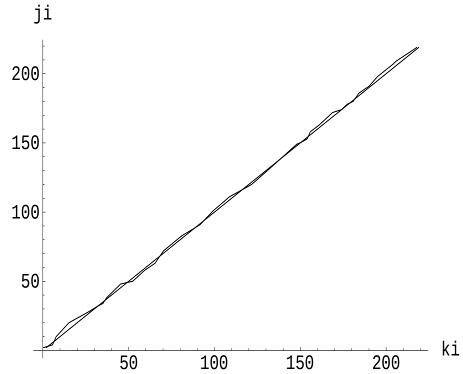

Figure 6. The irregular line joins the points $(k(i), j(i))$ where $k(i)$ is a predicted range measurement and $j(i)$ is the corresponding actual range measurement. The straight line is the diagonal, $k(i) = j(i)$.

The accuracy of the projective transformation $i \mapsto k(i)$ is illustrated in Fig. 6. Let $(i_1, j(i_1))$, ..., $(i_m, j(i_m))$ be the sequence of points obtained as i ranges over those values for which $X_i(1)$ corresponds to $X_{j(i)}(2)$. The irregular line in Fig. 6 is obtained by joining $(k(i_l), j(i_l))$ and $(k(i_{l+1}), j(i_{l+1}))$ for $1 \leq l < m$. The straight line is the diagonal. As expected, the irregular line is close to the diagonal.

6.3 Threshold

A check with random sets of measurements shows that the identification of 39 inliers in a set of 45 measurements is enough to ensure that there is a very low probability of a false detection. The algorithms CoarseSearch and Refine were applied to sets of N measurements sampled randomly and uniformly from D for $N = 30, 35, 40, 45, 50, 55$. The remaining parameter values for the algorithms were the same as in Section 5.2. Table 1 shows for each value of N, the average of the maximum number of inliers over 5 trials.

Table 1. Number of inliers averaged over five trials for each value of N.

N	30	35	40	45	50	55
inliers	19.2	22.6	27	30.4	35	37.6

7 Conclusion

An approximation K is obtained to the Fisher-Rao metric for projective transformations of the line. The approximation is accurate if the perturbations of the measurements due to noise are small. It is shown that under the metric K the parameter manifold $\mathrm{PSL}(2, \mathbb{R})$ for projective transformations of the line has a finite volume. As a result the problem of finding the elements of $\mathrm{PSL}(2, \mathbb{R})$ compatible with a given set of measurements reduces to the simple problem of checking the number of inliers associated with each element in a finite set of examples chosen from $\mathrm{PSL}(2, \mathbb{R})$. The number of examples becomes very large if the noise level is small but this difficulty can be overcome by using a multiresolution approach. Algorithms for detecting elements of $\mathrm{PSL}(2, \mathbb{R})$ were implemented in Mathematica, and tested on natural imagery.

The experiments showed that elements of $\mathrm{PSL}(2, \mathbb{R})$ can be detected even when the correspondences between the measurement components in the domain and the range of the projective linear transformation are unknown. The only requirement is that the ordering of the measurements in the domain is preserved under the transformation to the range.

There are many research directions arising from this work. Two of the main ones are:

i) To which other image structures can these methods be applied?

ii) What is the best way of sampling the parameter space?

The detection of lines is described in Maybank (2004). The sampling of the parameter space can be made more efficient, in that the number of sample points can be reduced, by making better use of the structure of the sample space as a Riemannian manifold.

Acknowledgement

I thank James Ferryman for permission to use images from the PETS2001 database.

References

1. Abramowitz, M. and Stegun, I.A. (eds.) 1965. *Handbook of Mathematical Functions with formulas, graphs, and mathematical tables.* Dover.
2. Amari, S.-I. 1985. *Differential-Geometrical Methods in Statistics.* Lecture Notes in Computer Science, **28**. Springer.
3. Chavel, I. 1984. *Eigenvalues in Riemannian Geometry.* Academic Press Inc.
4. Cover, T.M. and Thomas, J.A. 1991. *Elements of Information Theory.* John Wiley and Sons.
5. Faugeras, O.D. 1993. *Three-Dimensional Computer Vision.* MIT Press.
6. Ferryman, J.M. 2001. PETS'2001 database. Available at http://www.visualsurveillance.org/PETS2001.
7. Fisher, R.A. 1922. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond.*, Series A, **222**, pp. 309-368.
8. Forsyth, D.A. and Ponce, J. 2003. *Computer Vision, a modern approach.* Prentice Hall.
9. Gallot, S., Hulin, D. and LaFontaine, J. 1990. *Riemannian Geometry.* 2nd edition, Universitext, Springer.
10. Golub, G.H. and Van Loan, C.F. 1996. *Matrix Computations.* The John Hopkins University Press.
11. Gonzalez, R.C. and Woods, R.E. (2002) *Digital Image Processing.* 2nd Edition. Prentice Hall.
12. Hartley, R. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision.* Cambridge University Press.
13. Kanatani, K. 1996 *Statistical Computation for Geometric Optimization: theory and practice.* Elsevier.
14. Kotz, S. and Johnson, N.L. (eds.) 1992. *Breakthroughs in Statistics. Vol. 1. foundations and basic theory.* Springer Series in Statistics, Springer-Verlag.
15. Maybank, S.J. 2003. Fisher information and model selection for projective transformations of the line. *Proceedings of the Royal Society of London*, Series A, **459**, pp. 1-21.
16. Maybank, S.J. 2004. Detection of image structures using the Fisher information and the Rao metric. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **26**, No. 12.
17. Myung, J., Balasubramanian, V. and Pitt, M.A. 2000. Counting probability distributions: differential geometry and model selection. *Proc. National Academy of Science*, **97**, no. 21, pp. 11170-11175.
18. Rao, C.R. 1945. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, **37**, pp. 81-91.
19. Semple, J.G. and Kneebone, G.T. 1952. *Algebraic Projective Geometry.* Clarendon Press.
20. Torr, P.H.S. and Murray, D.W. 1993, Statistical detection of nonrigid motion. *Image and Vision Computing*, **11**, pp. 180-187.
21. Werman, M. and Keren, D. 1999. A novel Bayesian method for fitting parametric and non-parametric models to noisy data. *Proc. Computer Vision and Pattern Recognition*, Fort Collins, Colorado, June 1999, **2**, pp. 552-558.
22. Wolfram, S. 2003. *The Mathematica Book.* 5th Edition, Wolfram Media.