Bayesian Astrometry

J. B. James, J. S. Bloom, …

Astronomy Department, University of California, Berkeley CA, 94720-3411, USA

Draft dated June 7, 2012

ABSTRACT

We provide a scheme for probabilistic astrometry incorporating uncertainty distributions in matched object locations, as well in the astrometric mapping itself. We consider the case of astrometric calibration involving image alignment with linear transformations—translation, scaling and rotation—followed by non-parametric regression to account for localized distortions. The probabilistic mapping between frames that results can produce non-trivial distributions in the target frame, allowing for MCMC sampling to be used to propagate positions. Additionally, we present an accurate approximation to this process where the image alignment probability distribution can be described analytically, so that an astrometric query can be handled quickly and concisely, returning a location in the target frame that has its uncertainty specified in two spatial coordinates with covariance between them. We discuss generalizing this formalism to incorporate prior knowledge on astrometric components such as proper motion and parallax. An implementation of this scheme, including extensive documentation and examples, is available at

Berian: "Text appearing in these boxes are guides for what to comment on."

Subject headings:

Contents

1. Introduction

(JSB to contribute some opening text)

1.1. Astrometry as a gaussian process

In this work, we treat astrometry as a problem in vector field regression on the celestial sphere, where uncertainties may be present in the observational data and required in output. The regression has two parts: a transformation representing the alignment of frames that is applied uniformly across the image as an affine (parametric) mapping, and localized non-parametric distortions that are a residual to this background. This prescription leads us to model the astrometric mapping as a gaussian process: the map is an element in a space of functions characterized by a mean function—the background mapping—and a covariance representing the distortions to this mean.

A short review of gaussian processes is provided within Section 2. Treating astrometry in this way carries a number of strong benefits: i) uncertainties can be derived that contain covariance between celestial coordinates; ii) all sources of error can be properly tracked, including the uncertain alignment between frames and the otherwise poorly understood distortion map; iii) regression from observed points to arbitrary locations within the field can incorporate variance throughout the image, extending even to highly non-trivial distortion effects; and iv) astrometry can be treated as Bayesian updating, allowing prior knowledge about proper motion and parallax (among other considerations) to be incorporated.

At its core, we require the process of astrometry to provide astronomers with positions transformed from one frame to another, with a concise summary of their uncertainties. While this process will in general yield non-trivial probability distributions that cannot be easily summarized, and being able to show such distributions visually is an important end, we also find that it is often possible to approximate the output simply and accurately as two-dimensional (i.e., bivariate) normal distributions in celestial coordinates, allowing for the inclusion of this technique in the data reduction process common to all members of the astronomical community.

1.2. Previous work

Summary of work within the astronomical community relevant to the goal of probabilistic astrometry (JSB? maybe goes at top of introduction?). Also, a summary of multi-response regression techniques, as we are extending these in interesting ways (JBJ, maybe just gets put in method section, as 'previous work' subsections in the intro are uncommon in astronomy.)

2. Scheme for probabilistic astrometry

The astrometric mapping between a pair or sequence of astronomical data frames should be thought of an uncertain *function*, and mappings between particular locations in each frame as evaluations of that function. This is because the mapping is only made apparent through the motions of matched objects between target frames, about which we might have some previous knowledge (for instance, in the form of parallax or proper motion), but all of which is inherently probabilistic. This article is concerned with how to handle the fundamental tasks of astrometry given this uncertainty.

We are able to make some statements about the nature of probabilistic astrometric mapping: firstly, that part of the transformation involves alignment of images, the result of the fairly simple linear operations of translation, rotation and scaling. Even this analytically tractable component will be inherently probabilistic when the locations of the objects used to define it are uncertain. The result is that an exact location in one frame may be uncertain and even highly non-trivial in another; a source with an uncertain location will compound this effect.

Secondly, imposed on top of this parametric background are localized distortions produced by atmospheric or instrumental effects, whose physical origin is not, and likely cannot, be specified. Conventionally modeled with bivariate Chebyshev or Legendre

polynomials, these distortions are in fact more flexibly treated in a non-parametric manner where the covariance between different positions within the image is specified explicitly. We are led to the choice of representing the astrometric mapping as a gaussian process, a probabilistic function defined by a mean function (the background mapping) and, by analogy with the multivariate normal treatment of data vectors, a covariance between pairs of spatial locations that controls the way a particular instance of the probabilistic function departs from the mean.

The space of mappings between frames is greatly constrained by observations in the form of matched objects in each frame. The problem of building a probabilistic astrometric mapping is then reduced to that of imposing these observations on any initial knowledge, and that of astrometric querying to the task of gaussian process regression. In the language of Bayesian probability, the prior knowledge of the mapping distribution is updated to a posterior distribution that incorporates the information provided by the data with their uncertainties. Then, querying is the result of propagating a point or spatial region in one frame through this function, mapping to a distribution in the target frame.

This scheme provides an estimate of the posterior likelihood for the astrometric solution between frames under the following assumptions:

(i) two frames, A and B, are used as comparators, where by convention A is a reference frame, though this choice has no impact on the method;

(ii) objects within frames are characterized by bivariate normal distributions, with means $\boldsymbol{\mu}_i$ (the 'location' of the object) and uncertainty ellipses defined by the covariance matrix $\boldsymbol{\Sigma}_i$;

(iii) the number of tie objects is known and all such pairs have been matched between the two frames, though the possibility of false matches is not excluded.

Because astrometry is concerned with the motion (displacement) of objects between

images, the uncertain function that is being estimated is a vector field, representing the transformation at each location of interest, with an associated covariance. If one wishes to know the location distribution of an object in the target frame, given its location (and distribution) in the initial frame, this is found by integrating over the vector field and associated covariance.

Estimation of this field proceeds in the two parts outlined above: an affine background mapping component is estimated to account for global differences between the images that result from a relatively restricted set of linear transformations, specified by a seven-parameter model. The form of the likelihood distribution in this seven-parameter space can be left unspecified in general when sampling is used propagate between frames. From each sampling, the residual to the background transformation yields a local distortion field, which is characterized non-parametrically and with full covariance as a two-dimensional gaussian process whose mean is the background transformation. Chaining these two steps and tracking the sources of covariance that arise from them yields a likelihood distribution for the vector distribution field; this is the Bayesian astrometric solution.

Nothing in the formalism that follows requires assumptions about the probability distributions for the locations of observed matched pairs. However, it transpires that the assumption of bivariate normal object distributions makes the likelihood computation especially well behaved, so that many computations can be done without numerical integration. Moreover, this is the choice of distribution implicit when specifying a location with uncertainty in celestial coordinates, which is virtually the only representation of object positions used at the present time. This choice is adopted throughout this work. The remainder of this section gives an overview the parametric (2.1) and non-parametric (2.2) components of the scheme.

2.1. The parametric background

In addition to the purely formal assumptions listed above, the following physical considerations are made:

(i) Image alignment consists of a global transformation that is homogeneous across the frame, i.e., the mapping is the same at every point;

(ii) the basis operations for the space of mappings between the two frames are: translation $\mathbf{x} \to \mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$, rotation about an arbitrary point $\mathbf{x} \to \mathbf{x}' = \mathbf{U}(\theta)\,(\mathbf{x} - \mathbf{x_0}) + \mathbf{x_0}$ and scaling $\mathbf{x} \to \mathbf{x}' = \Lambda\mathbf{x}$, where scaling by negative values is allowed, corresponding to the image coordinates being flipped.

This generates a seven-parameter model $\boldsymbol{\Theta}_b = \{\Delta x, \Delta y, \theta, x_0, y_0, \Lambda_{xx}, \Lambda_{yy}\}$, which transforms an initial point or probability distribution $p(\mathbf{x})$ to an aligned distribution $p_a(\mathbf{x}; \boldsymbol{\Theta}_b) \equiv T_{\boldsymbol{\Theta}_b}\big[p(\mathbf{x})\big]$, where T stands in for the form of the linear mapping transformation that will be made explicit in Section 3; that section also treats the problem of likelihood estimation for this model in detail. Given matched catalogs of object locations with uncertainties, the parametric mapping can be expressed as a probability distribution $p\,(D; \boldsymbol{\Theta}_b)$, the probability of the observed data D for a specific choice of transformation parameters. For sampling, all that is required is the ability to compute the likelihood $\mathscr{L}(\boldsymbol{\Theta}_b | D)$ for any choice of parameter set: an analytic form for the distribution as a whole is not needed.

When the input data have uncertainties associated with them, this background distribution will be spread out about its maximum value, giving rise to non-trivial target frame distributions even before local distortions are considered. For instance, consider the effect of an uncertain rotation parameter θ. A fixed point (other than the centre of rotation) will be drawn out along an arc by this probabilistic mapping, an effect whose extent will increase the further the point is from the centre of rotation; similarly, uncertain scaling parameters will

produce rectilinear distributions. By contrast, the effect on uncertain translation parameters is relatively benign, and in Section 5 we present ways this can be handled analytically in some settings.

But combining these uncertain transformations gives rise to target frame distributions that we should not aim to characterize explicitly. Likelihood sampling will allow us to define a contour in the target frame at some probability threshold that represents how a point or region is distributed. In particular, the exact expression

$$p_t(\mathbf{x}) = \int p_a(\mathbf{x};\mathbf{\Theta}_b)dp(\mathbf{\Theta}_b) = \int T_{\mathbf{\Theta}_b}\left[p(\mathbf{x})\right]p(\mathbf{\Theta}_b)d\mathbf{\Theta}_b \tag{1}$$

for the posterior target frame probability distribution p_t given a distribution p in the initial frame and background mapping distribution p_b can be sampled with the Metropolis-Hastings algorithm and characterized empirically to a desired level of precision.

2.2. The non-parametric residual distortion map

The parametric background is a first-order characterisation of the difference between image frames. The residuals to this background are local distortions that are manifestly non-parametric. We model this distortion map using a two-dimension gaussian process over the vector residuals at each location, training a covariance function to describe how a known mapping (*i.e.*, from each of the tie objects) influences the estimation of the model in its neighborhood.

A standard reference for gaussian processes is Rasmussen & Williams (2006), whose development we follow here. A gaussian process may be thought of as a probability distribution $f(\mathbf{x})$ over the space of functions, characterized by a mean function $m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right]$ and a covariance function $k(\mathbf{x},\mathbf{x}') = \mathbb{E}\left[\left(f(\mathbf{x}) - m(\mathbf{x})\right)\left(f(\mathbf{x}') - m(\mathbf{x}')\right)\right]$; we write

$$f \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')) = m(\mathbf{x}) + \mathcal{GP}(0, k(\mathbf{x},\mathbf{x}')). \tag{2}$$

In probabilistic astrometry, the role of the mean function is played by the background image alignment transformation described in the previous subsection; local distortions are residuals to this mean. The generation of a distortion map proceeds by conditioning the function $k(\mathbf{x}, \mathbf{x}')$ that describes the covariant relationship between pairs of spatial locations, using the observed residuals of the matched catalogs. This process of conditioning is described in Section 4.

The description of a gaussian process as a probability distribution in the space of functions is abstract, and in practice what matters most is the values this function takes at specific locations: particularly, the locations of the observed matches on the one hand, and the locations within the image that we would like query astrometrically on the other. Using a gaussian process means that the values of the astrometric map at these sets of locations are jointly normally distributed, so that we can produce a conditional distribution for one set of locations (the ones we would like to regress) based on the others:

$$
\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix} + \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix} \right).
\tag{3}
$$

The vector and matrix elements above are evaluations of the astrometric mapping at the observation (X) and target (X_*) locations. The vector of means is the background transformation mapping applied at the different sets of locations; the joint normal distribution on the right-hand side has mean zero and a covariance matrix who elements are themselves the matrices of the covariance function $k(\mathbf{x}, \mathbf{x}')$ evaluated at each pair of locations in the two sets X and X_*, that is,

$$
K(X,Y) = \begin{bmatrix} k(\mathbf{x}_1,\mathbf{y}_1) & \cdots & k(\mathbf{x}_1,\mathbf{y}_m) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n,\mathbf{y}_1) & \cdots & k(\mathbf{x}_n,\mathbf{y}_m) \end{bmatrix};
\tag{4}
$$

hereafter, we will write K for $K(X,X)$, K_* for $K(X_*,X) = K(X,X_*)^T$ and K_{**} for $K(X_*,X_*)$.

Because the values at all locations are jointly normally distributed, the astrometric mapping at the new locations X_*, given the observations of the mapping at locations in X, is

itself normally distributed, with mean and covariance

$$\overline{\mathbf{y}_* | \mathbf{y}} = m(X_*) + K_* K^{-1} \mathbf{y}; \quad C(\mathbf{y}_* | \mathbf{y}) = K_{**} - K_*^T K^{-1} K_*. \tag{5}$$

These expressions are the formulae for regressing the values of the astrometric map at query locations, given the observations in the matched catalogs. To reiterate, the mean displacement of the mapping is the background image alignment and the residuals are described completely by the covariance function $k(\mathbf{x}, \mathbf{x}')$, about which we have said very little so far. The detailed consideration of this function will be postponed until Section 4, but its basic properties are that it will be symmetric in inputs, stationary across the image frame and a function of the separation between input locations, so that it might instead be written $k(|\mathbf{x} - \mathbf{x}'|)$. Furthermore, we will presume that it is characterised by a limited set of parameters, called 'hyperparameters' in this context, as they define a model within a space that is abstractly above the spatial plane of the image with which we are actually concerned. Training the covariance function with the observed data will produce a joint likelihood distribution on these parameters, in a manner analogous to the parameters of the background mapping model described in Section 2.1.

Ultimately, the hyperparameters $\mathbf{\Theta}_{\mathrm{HP}}$ are joined to the background model parameters $\mathbf{\Theta}_b$, yielding a higher-dimensional mapping space over which likelihood sampling takes place. This yields a probabilistic solution for the astrometric mapping between frames. Extending Equation (1),

$$p_t(\mathbf{x}) = \int \int y_* \left[p(\mathbf{x}); \mathbf{\Theta}_b, \mathbf{\Theta}_{\mathrm{HP}} \right] p(\mathbf{\Theta}_b) p(\mathbf{\Theta}_{\mathrm{HP}}) d\mathbf{\Theta}_p d\mathbf{\Theta}_{\mathrm{HP}} \tag{6}$$

$$= \int \int \underbrace{T_{\mathbf{\Theta}_b} \left[p(\mathbf{x}) \right]}_{\text{image alignment}} + \underbrace{\left(K_* K^{-1} \right)_{\mathbf{\Theta}_{\mathrm{HP}}} \left(p(\mathbf{x}) - T_{\mathbf{\Theta}_b} \left[p(\mathbf{x}) \right] \right)}_{\text{gaussian process residual distortion map}} dp(\mathbf{\Theta}_b) dp(\mathbf{\Theta}_{\mathrm{HP}}). \tag{7}$$

This is the master equation for the querying of an astrometric mapping defined by a matched pair of catalogs. The next two sections provide details on how the components of the probabilistic mapping are defined and how they can be sampled.

2.3.   Sampling from priors and posteriors

Equation (7) describes an astrometric query as a process by which a possibly uncertain location in one frame is pushed to the target frame by a mapping that is itself uncertain. When little is known about the nature of the distributions for the locations or the mapping beyond a method to evaluate the likelihood for some choice of parameters, a likelihood sampling technique like the Metropolis-Hastings algorithm can be used to generate samples in the target frame.

3.   Image alignment: MCMC likelihood estimation

The previous section described the process through which a probability distribution in the initial frame is propagated into the target frame with a probabilistic mapping that is defined by matched pairs of objects with associated uncertainties. This section describes the first component of this mapping, the linear transformation of image alignment that acts globally across the frame. Except where noted, it is assumed that objects in the matched catalogs have their locations specified by bivariate normal distributions, as this is the only case relevant to astronomers at the present time and makes the presentation substantially more concrete. Nothing within the formalism explicitly requires this assumption, however; for non-gaussian distributions on the observed object locations, the process of likelihood sampling is simply extended to incorporate those distributions as well.

3.1.   Linear transformation of bivariate normal distributions

A bivariate normal distribution is an excellent representation of an uncertain celestial position at the current level of precision available to astronomers. Writing $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the

probability density function for the object's location is

$$p(\mathbf{x}) = \frac{1}{2\pi|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \tag{8}$$

In the context of an image frame, $\mathbf{x}$ might be a vector of the pixel coordinates; or it could represent the right ascension and declination of the object. We assume that all objects in the matched catalogs have location distributions of this form and label the two frames under comparison A and B.

Let $p_A^i \sim \mathcal{N}(\boldsymbol{\mu}_A, \mathbf{\Sigma}_A)$ represent the likelihood distribution in frame A for the ith matched object and similarly for p_B^i. It is straightforward to generate the corresponding frame B distribution under the seven-parameter transformation $\mathbf{\Theta}_b = \{\Delta\boldsymbol{\mu}, \theta, \mathbf{x}_0, \mathbf{\Lambda}\}$ using the spectral decomposition of $\mathbf{\Sigma} = \mathbf{VDV}^{-1}$:

$$p_{\hat{B}}^i(\mathbf{x}) \equiv T_{\mathbf{\Theta}_b}\left[p_B^i(\mathbf{x})\right] \;\; = \;\; \Delta\boldsymbol{\mu} + \mathbf{\Lambda}^{1/2}\mathbf{U}(p_B^i - \mathbf{x}_0) + \mathbf{x}_0 \tag{9}$$

$$\sim \;\; \mathcal{N}\left(U(\mathbf{\Lambda}\boldsymbol{\mu} + \Delta\boldsymbol{\mu} - \mathbf{x}_0) + \mathbf{x}_0, (\mathbf{UV})(\mathbf{\Lambda}^{1/2}\mathbf{D})(\mathbf{UV})^{-1}\right). \tag{10}$$

Here, $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{\Sigma}$ and $\mathbf{D}$ the diagonal matrix of corresponding eigenvalues; the transformation composes scaling, translation and rotation (in that order) and operates on the mean of the gaussian in the way it would on a single point, while the error ellipse is transformed by rotating the principal axes of the error ellipse with the rotation matrix U of angle θ and scaling each of the principal axis sizes, which are given by the eigenvalues. Figure 1 shows the transformations of a template bivariate gaussian generated by the basis translation, rotation and scaling mappings where the center of rotation $\mathbf{x}_0$ has been placed at the centre of the frame.

3.2. Likelihood construction

This section describes how to evaluate the relative likelihood value for a mapping $\mathbf{\Theta}_b$ given the object distributions in the two frames. The B frame distributions are transformed in
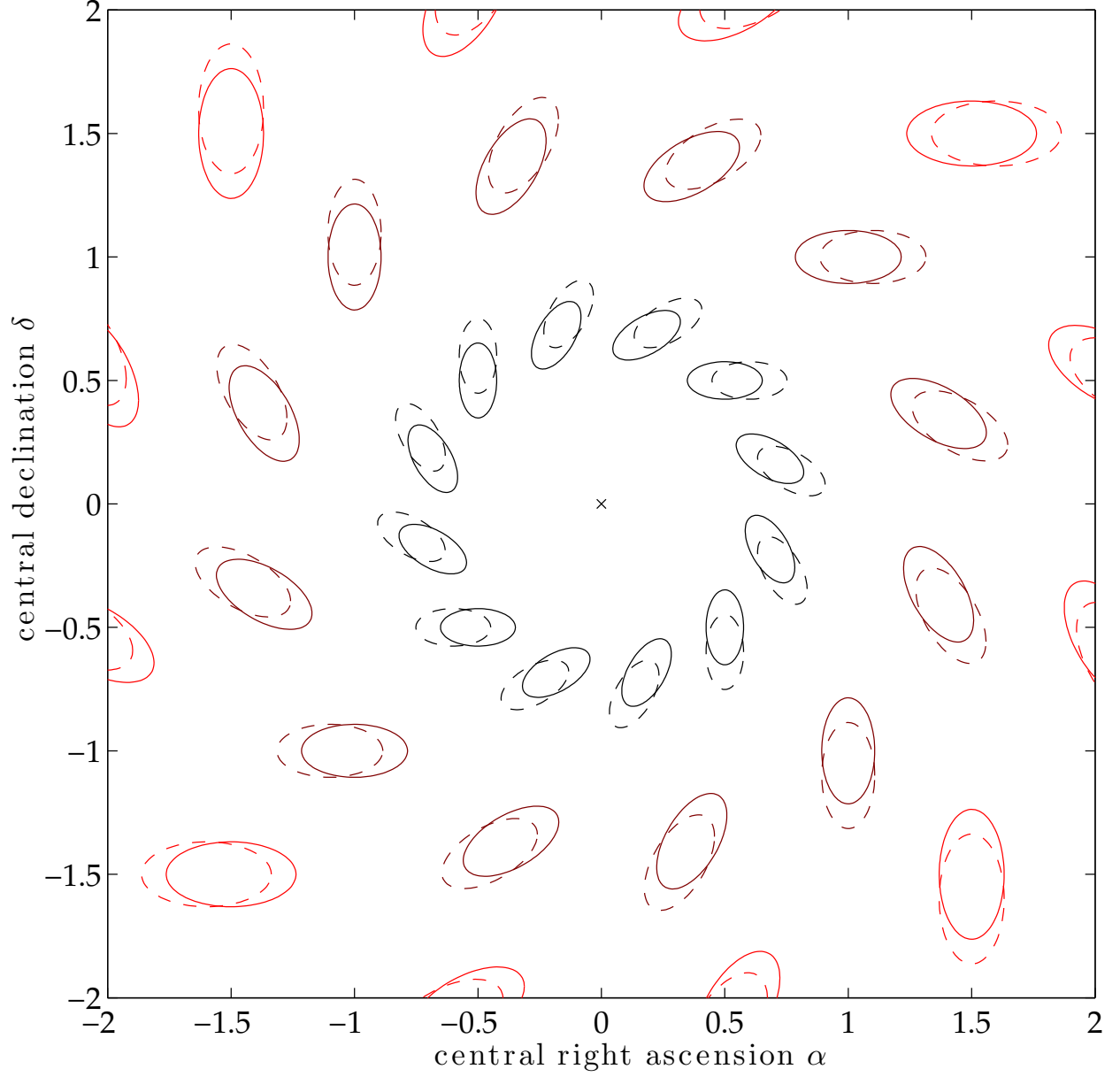
Fig. 1.— Mappings of a bivariate gaussian under the three basis operations considered in this work: translation (dashed), scaling (color) and rotation. In each case the template bivariate $\mathcal{N}\left(\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}0.1 & 0\\0 & .05\end{smallmatrix}\right]\right)$ is transformed according to the prescription of equation (10).

the manner described above, so that it remains to decide how the correspondence between tie objects in the reference and mapped frames is evaluated.

Given a tie object reference distribution p_A^i and a trial mapped distribution $p_{\tilde{B}}^i$, the likelihood contribution from the tie pair is given by a measure of separation between the two distributions. The natural choice of distance function for the two distributions in this context is the Bhattacharaya distance,

$$D_B = -\log\left(\int_{\mathscr{A}} \sqrt{pq}\,d\mathbf{x}\right),\tag{11}$$

i.e., the integral over the frame of the geometric mean distribution. This definition provides $D_B = 0$ when $p = q$ and the integral is interpreted geometrically as a measure of overlap between the two distributions. In the limit where the object's location is known exactly in one frame, the distance gives a normalised measure of the deviation of that value from the centre of the comparison distribution. [1] For a pair of bivariate gaussians distributions, the distance D_B permits analytic expression:

$$D_B(p_A^i, p_{\tilde{B}}^i) = \frac{1}{8}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_{\tilde{B}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_{\tilde{B}}) + \frac{1}{2}\log\left(\frac{|\boldsymbol{\Sigma}|}{\sqrt{|\boldsymbol{\Sigma}_A||\boldsymbol{\Sigma}_{\tilde{B}}|}}\right), \quad \text{where } \boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_{\tilde{B}}}{2},\tag{12}$$

so that for a background mapping $\boldsymbol{\Theta}_b$, the likelihood computation given the data is

$$\mathscr{L}(\boldsymbol{\Theta}_b|D) \propto \exp\left(-\sum_i D_B\left(p_A^i, p_{\tilde{B}(\boldsymbol{\Theta}_b)}^i\right)\Big/ 2\right).\tag{13}$$

---

[1]An alternative is the Kullback-Leibler divergence

$$D_{KL}(p|q) = \int_{-\infty}^{\infty} p(x)\log\frac{p(x)}{q(x)}\,dx$$

that is popular in feature separation. In a Bayesian context it should be interpreted as an evaluation of the information gain in going from the prior to the posterior, which is distinct to what we aim for in astrometry. This, together with it being asymmetric in p and q, leads us to prefer the Bhattacharaya distance; in a practical setting the difference is unimportant. Note, though, that neither of these quantities defines a metric in the formal sense.

4.   Residual distortion map: a gaussian process approach

The work in the previous section addresses the first part of the problem, providing a global mapping between image frames. The task is now to characterize probabilistically the resulting residual distortion map in a manner that can be smoothly joined to the task of estimating the background, so that both parts—parametric and non-parametric—can be carried out in unison.

We argue for treating the residual distortion map as a two-dimensional gaussian process over the vector components at each location, which requires the specification of a covariance function and outputs a mean vector (the mean astrometric solution) and covariance for the vector components. This section describes this covariance function in detail, particularly with regard to gaussian process regression of a vector, which has been the subject of only limited study within the machine learning literature.

4.1.   Vector-output gaussian processes

Much of the relevant literature on gaussian processes deals with the regression of a scalar-valued function. In these cases, the role of the covariance function is to describe how the different spatial locations (termed 'inputs') are correlated so that the values of the regression function at those locations (termed 'outputs') can be relayed to new locations. A particular choice of covariance function, written

$$k(\mathbf{x}, \mathbf{x}') = \sigma_l^2 e^{(\mathbf{x}-\mathbf{x}')^2/2l^2} \equiv \sigma_l^2 \tilde{k}(\mathbf{x}, \mathbf{x}') \tag{14}$$

and termed the 'square exponential' covariance function, is commonly encountered and is the general form that we pursue here; the unweighted covariance function $\tilde{k}$ will be needed when we consider vector output. The two hyperparameters l and σ_l control the scale and amplitude of the correlations between locations, respectively. Using all pairs of points within the sets X

and X_* of the input and regression locations, the matrices populated with scalar values K, K_* and K_{**} can be constructed from evaluations of the covariance function.

When the outputs are not scalar-valued but vector-valued, a consideration of a different kind arises. Now, in addition to the correlation between different spatial locations, the correlation between the components of the outputs must also be evaluated; in the case of astrometry, there are two components for the two spatial directions. We handle this by splitting the role of the scale and amplitude hyperparameters: within the image frame, the scale parameter is isotropic and so remains a scalar l, but the amplitude should be generalized to a symmetric 2×2 matrix, so that the covariance function is matrix valued:

$$
\begin{bmatrix} k_{xx}(\mathbf{x},\mathbf{x}') & k_{xy}(\mathbf{x},\mathbf{x}') \\ k_{yx}(\mathbf{x},\mathbf{x}') & k_{yy}(\mathbf{x},\mathbf{x}') \end{bmatrix} = \begin{bmatrix} \sigma^2_{l,xx} & \sigma_{l,xy} \\ \sigma_{l,yx} & \sigma^2_{l,yy} \end{bmatrix} e^{(\mathbf{x}-\mathbf{x}')^2/2l^2} \equiv \mathbf{\Sigma}_k e^{(\mathbf{x}-\mathbf{x}')^2/2l^2},
\tag{15}
$$

subject to the constraints $\sigma_{l,xy} = \sigma_{l,yx}$ and, unless a clear reason for discriminating between spatial directions is apparent, $\sigma_{l,xx} = \sigma_{l,yy}$.

This matrix represents the correlation between each pair of components at just one particular pair of locations $\mathbf{x}$ and $\mathbf{x}'$. Generalizing the construction of the covariance function matrices K, K_* and K_{**}:

$$
\mathbb{K} = \tilde{K} \otimes \mathbf{\Sigma}_k = \begin{bmatrix} \tilde{k}(\mathbf{x}_1,\mathbf{y}_1) & \cdots & \tilde{k}(\mathbf{x}_1,\mathbf{y}_m) \\ \vdots & \ddots & \vdots \\ \tilde{k}(\mathbf{x}_n,\mathbf{y}_1) & \cdots & \tilde{k}(\mathbf{x}_n,\mathbf{y}_m) \end{bmatrix} \otimes \begin{bmatrix} \sigma^2_{l,xx} & \sigma_{l,xy} \\ \sigma_{l,yx} & \sigma^2_{l,yy} \end{bmatrix}
\tag{16}
$$

$$
= \begin{bmatrix} \tilde{k}(\mathbf{x}_1,\mathbf{y}_1)\mathbf{\Sigma}_k & \cdots & \tilde{k}(\mathbf{x}_1,\mathbf{y}_n)\mathbf{\Sigma}_k \\ \vdots & \ddots & \vdots \\ \tilde{k}(\mathbf{x}_n,\mathbf{y}_1)\mathbf{\Sigma}_k & \cdots & \tilde{k}(\mathbf{x}_n,\mathbf{y}_n)\mathbf{\Sigma}_k \end{bmatrix};
\tag{17}
$$

here, $\otimes$ is the Kronecker matrix product, so that each element of $\mathbb{K}$ is the 2×2 matrix of covariance between spatial components weighted by the unnormalised scalar covariance function value. While the nested matrices appear conceptually, in implementation $\mathbb{K}$ is simply

a $2n \times 2n$ matrix, for n input locations.[2] Likewise, if regression is occurring to m new locations, $\mathbb{K}_*$ and $\mathbb{K}_{**}$ will be $2n \times 2m$ and $2m \times 2m$ respectively.

The gaussian process regression equations are straightforwardly generalized

$$\overline{\mathbf{y}_* | \mathbf{y}} = m(X_*) + \mathbb{K}_* \mathbb{K}^{-1} \mathbf{y}; \quad \mathbb{C}(\mathbf{y}_* | \mathbf{y}) = \mathbb{K}_{**} - \mathbb{K}_*^T \mathbb{K}^{-1} \mathbb{K}_*, \tag{18}$$

where

$$\mathbf{y} = \left[(y_{1,x}, y_{1,y}), (y_{2,x}, y_{2,y}), \ldots, (y_{n,x}, y_{n,y}) \right]^T \tag{19}$$

is reshaping of the $n \times 2$ observed output spatial components at the n observed input locations. Similarly, the regression yields a vector of length $2m$ ordered as pairs of spatial components at each of the m regression locations, with all correlations between and vector components and locations accounted for. Finally, each element of $\mathbb{C}$ is a 2×2 covariance matrix, and the diagonal of $\mathbb{C}$ is a sequence of m 2×2 covariance matrices describing the uncertainty in the regressed vector mapping at each output location.

4.2. Measurement uncertainties on matched catalogue data

When measurement uncertainties are present in the observations, this information must be propagated through the estimation of the astrometric mapping. For a pair of matched object locations (p_A^i, p_B^i) that are both bivariate normal distributions, the mapping vector they define is also bivariate normal, with mean $\boldsymbol{\mu}_A^i - \boldsymbol{\mu}_B^i$ with uncertainty $\Sigma_A^i + \Sigma_B^i$. For a particular background mapping parameter choice $\boldsymbol{\Theta}_b$, the residual to the background mapping—which is

---

[2]This formulation is most similar to one suggested by Bonilla et al. (2008), who note that when the observed data are known precisely the gaussian process regression decouples into two independent processes—one on each of the vector components—a property known as autokrigeability in the field of geostatistics.

$\mathbf{y}_i$ in the gaussian process—will remain bivariate normal, with mean vector $(\boldsymbol{\mu}_A^i - \boldsymbol{\mu}_B^i) - T_{\boldsymbol{\Theta}_b}\left[\boldsymbol{\mu}_A^i\right]$ and covariance $T_{\boldsymbol{\Theta}_b}\left[\Sigma_A^i + \Sigma_B^i\right]$.

This transformed covariance matrix Σ_r^i, representing the uncertainty in the vector residual to the background mapping, can be incorporated into the gaussian process regression using a formulation that is standard for scalar-output gaussian processes:

$$\overline{\mathbf{y}_* | \mathbf{y}} = m(X_*) + \mathbb{K}_*(\mathbb{K} + \Sigma_r\mathbb{I})^{-1}\mathbf{y}; \quad \mathbb{C}(\mathbf{y}_* | \mathbf{y}) = \mathbb{K}_{**} - \mathbb{K}_*^T(\mathbb{K} + \Sigma_r\mathbb{I})^{-1}\mathbb{K}_*, \tag{20}$$

where $\Sigma_r\mathbb{I}$ represents a diagonal matrix whose n diagonal elements are the residual mapping uncertainties Σ_r^i for each of the n pairs of matched object locations. Both the regressed vector field and the estimate of its uncertainty will now incorporate the uncertainty in the initial observations.

4.3. Covariance hyperparameters

In the vector-output gaussian process described above, there are three hyperparameters: l, the isotropic scale of the covariance function; $\sigma_l^2 \equiv \sigma_{xx}^2 = \sigma_{yy}^2$, the amplitude of gaussian process regression; and the new parameter σ_{xy} describing the amplitude of the correlation between the vector output components. Though gaussian process regression may be thought of as non-parametric in the space of the image, it is in fact hyperparametric and these three parameters can be conditioned in just the same way as the parameters to the background mapping. The log-likelihood for a choice of hyperparameters $\boldsymbol{\Theta}_{\mathrm{HP}}$ is

$$-2\log\mathscr{L}(\boldsymbol{\Theta}_{\mathrm{HP}}|D) = \mathbf{y}^T(\mathbb{K} + \Sigma_r\mathbb{I})^{-1}\mathbf{y} + \log|\mathbb{K} + \Sigma_r\mathbb{I}| + \log 2\pi, \tag{21}$$

which can be used both to define a maximum-likelihood choice of hyperparameters and in likelihood sampling to generate samples of points from the distribution in the target frame.

This completes the description of the necessary formalism for Bayesian astrometric sampling. As presented, an input set of matched catalogues whose data are object

locations expressed as bivariate normal distributions will define a total likelihood expression $\mathscr{L}(\mathbf{\Theta}_b|D) + \mathscr{L}(\mathbf{\Theta}_{\mathrm{HP}}|D)$. Sampling this likelihood, perhaps with the additional step of sampling the location in the initial frame when it is specified as a distribution, will produce a sequence of locations in the target frame that can be treated as an estimate of the posterior astrometric mapping solution.

5. Normal approximation to astrometric mapping

Likelihood sampling is the most comprehensive method of generating an astrometric mapping in the target frame, but frequently an approximation the distribution in the target frame will be desired so that the astrometric query can return a concise result. This can be achieved with the addition of two assumptions, neither of which is burdensome:

 (i) When the uncertainties on the observational data are small relative to the scale of the image, the background mapping probability distribution will be well approximated by a seven-dimensional multivariate normal distribution in the parameters $\mathbf{\Theta}_b$;

 (ii) furthermore, when the background mapping distribution is well constrained about a maximum likelihood value $\mathbf{\Theta}_b^{\mathrm{ML}}$, the gaussian process hyperparameters can be conditioned just once, setting $p(\mathbf{\Theta}_{\mathrm{HP}})$ to a delta function at $\mathbf{\Theta}_{\mathrm{HP}}^{\mathrm{ML}}$.

If these requirements are satisfied, then the master astrometric regression equation (7) is analytically tractable. With the observational data and their uncertainties expressed as probabilistic vectors between matched objects locations, with mean vectors $\mathbf{y}_i$ and stacked covariances $\mathbf{\Sigma}_r^i$, then from an initial frame distribution $p(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, the output distribution in the target frame $p_t(\mathbf{x})$ will be well approximated by a bivariate normal distribution with mean mapping

$$\boldsymbol{\mu}_t = T_{\mathbf{\Theta}_b^{\mathrm{ML}}}[\boldsymbol{\mu}] + \left(\mathbb{K}_* (\mathbb{K} + \mathbf{\Sigma}_r \mathbb{I})^{-1} \right)_{\mathbf{\Theta}_{\mathrm{HP}}^{\mathrm{ML}}} \left(\mathbf{y} - T_{\mathbf{\Theta}_b^{\mathrm{ML}}}[\mathbf{y}] \right), \tag{22}$$

and covariance

$$\mathbf{\Sigma}_t = \left[\mathbb{K}_{**} - \mathbb{K}_*^T (\mathbb{K} + \mathbf{\Sigma}_r \mathbb{I})^{-1} \mathbb{K}_* \right]_{\mathbf{\Theta}_{\mathrm{HP}}^{\mathrm{ML}}} + \mathbf{\Sigma}_b + \mathbf{\Sigma}, \tag{23}$$

where the terms in this expression correspond (from left-to-right) to the covariance contributions of the gaussian process, the uncertain background mapping and the initial frame uncertainty.

5.1.   Image alignment probability distribution

All of the terms in the expressions (22) and (23) are defined by either the observational data $(\mathbf{y}, \mathbf{\Sigma}_r)$, the location of the query $(\boldsymbol{\mu}, \mathbf{\Sigma})$ or the parameter distributions $p(\mathbf{\Theta}_b) \sim \mathcal{N}(\mathbf{\Theta}_b^{\mathrm{ML}}, \mathbf{\Sigma}_{\mathbf{\Theta}_b})$ and $p(\mathbf{\Theta}_{\mathrm{HP}}) \sim \delta(\mathbf{\Theta}_{\mathrm{HP}}^{\mathrm{ML}})$. Proceeding with the assumption that the distribution for the background mapping parameters will be multivariate normal, we must decide how to extract an estimate of the covariance contribution within the image plane $\mathbf{\Sigma}_b$ from the covariance matrix of the larger parameter space $\mathbf{\Sigma}_{\mathbf{\Theta}_b}$.

The covariance matrix for the background mapping parameters, $\mathbf{\Sigma}_{\mathbf{\Theta}_b}$ can be estimated using the likelihood expression (13), by sampling the likelihood values along each parameter axis in the vicinity of the maximum. In particular, the elements of the precision matrix $L = \left[-\mathbf{\Sigma}_{\mathbf{\Theta}_b} \right]^{-1}$ are given by

$$L''_{ij} \equiv \frac{d^2 L}{d\theta_i d\theta_j} = \quad \left[\mathscr{L}\left(\mathbf{\Theta}_b^{\mathrm{ML}} + \epsilon \hat{e}_i + \epsilon \hat{e}_j \right) - \mathscr{L}\left(\mathbf{\Theta}_b^{\mathrm{ML}} + \epsilon \hat{e}_i - \epsilon \hat{e}_j \right) - \tag{24}$$

$$\mathscr{L}\left(\mathbf{\Theta}_b^{\mathrm{ML}} - \epsilon \hat{e}_i + \epsilon \hat{e}_j \right) + \mathscr{L}\left(\mathbf{\Theta}_b^{\mathrm{ML}} - \epsilon \hat{e}_i - \epsilon \hat{e}_j \right) \right] / 4\epsilon^2 , \tag{25}$$

where $\hat{e}_k$ is the unit parameter vector for a step along the kth dimension and ϵ is a small number. This procedure of numerical differentiation returns a precision matrix whose inverse fully characterizes the normal approximation to the background mapping distribution (Gelman et al. 2004).

Even when the background mapping distribution is represented in this manner, the propagation of this uncertainty into the target frame cannot be derived analytically. A special case is when the uncertainty in the translation parameters is the primary source of uncertainty, in which case the conditional covariance matrix constructed for these two parameters (i.e., given the maximum likelihood values for the others) will be a good approximation to $\mathbf{\Sigma}_b$.

More commonly, the covariance contribution from the background mapping should be estimated by sampling the multivariate normal $p(\mathbf{\Theta}_b)$ some number of times and estimating the covariance in the target frame of the resulting point set.

5.2. Covariance stacking

6. Applications

How can the construction of these distributions be painlessly used to improve the quality of science that is carried out?

6.1. Mock data

A pair of mock frames consisting of $n = 100$ tie objects is generated in the following manner

(i) the set of object locations in frame A are drawn using $\boldsymbol{\mu}_A^i \sim U\left([-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]\right)$;

(ii) each location is assigned bivariate gaussian uncertainty distribution $\mathcal{N}\left(\boldsymbol{\mu}_A^i, \mathbf{\Sigma}_A^i\right)$, where the covariance matrix $\mathbf{\Sigma}_A^i$ is determined from the widths $\left(\sigma_x^i, \sigma_y^i\right) \sim U[0.001, 0.1]$ and rotated by angle $\theta_A^i \sim U[0, 2\pi)$;

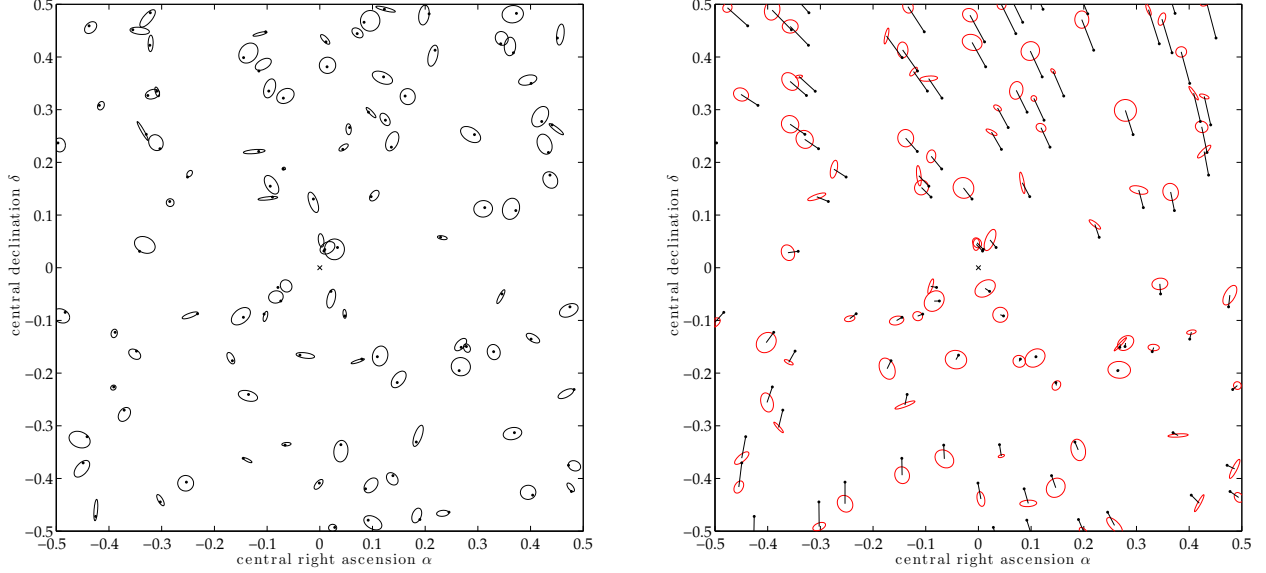(iii) a random point is drawn from each bivariate gaussian distribution to provide the starting

Fig. 2.— The mock data used in the application of the Bayesian astrometry scheme. *Left:* A set of 100 bivariate gaussians (ellipses) are placed within the image frame and sampled (points) to generate the starting locations for the mapping. *Right:* The starting locations (points) are transformed according to equation (10) and used as the central positions for objects in the new image frame, which are then assigned random bivariate gaussian covariance matrices (ellipses).

location $\boldsymbol{\mu}_B^i$ for the mapping into frame B, as shown in the left panel of Figure 2;

(iv) in the new image frame, these points are subject to the same random transformation, giving $\boldsymbol{\mu}_{\tilde{B}}^i \sim \mathbf{U}(\boldsymbol{\Lambda}\boldsymbol{\mu} + \Delta\boldsymbol{\mu})$, where $\mathbf{U}$ is the matrix of rotation by angle $\theta \sim U(0, \pi/50]$ about the origin, the scaling parameters are $(\Lambda_{xx}, \Lambda_{yy}) \sim U[0.9, 1, 1]$ and the translation is $\Delta\boldsymbol{\mu} \sim U([-0.02, 0.02] \times [-0.02, 0.02])$;

(v) the transformed object locations are assigned bivariate gaussian uncertainty distributions $\mathcal{N}\left(\boldsymbol{\mu}_{\tilde{B}}^i, \boldsymbol{\Sigma}_{\tilde{B}}^i\right)$ following the same prescription as before.

The two mock images with object locations and uncertainty distributions are shown in Figure 2.
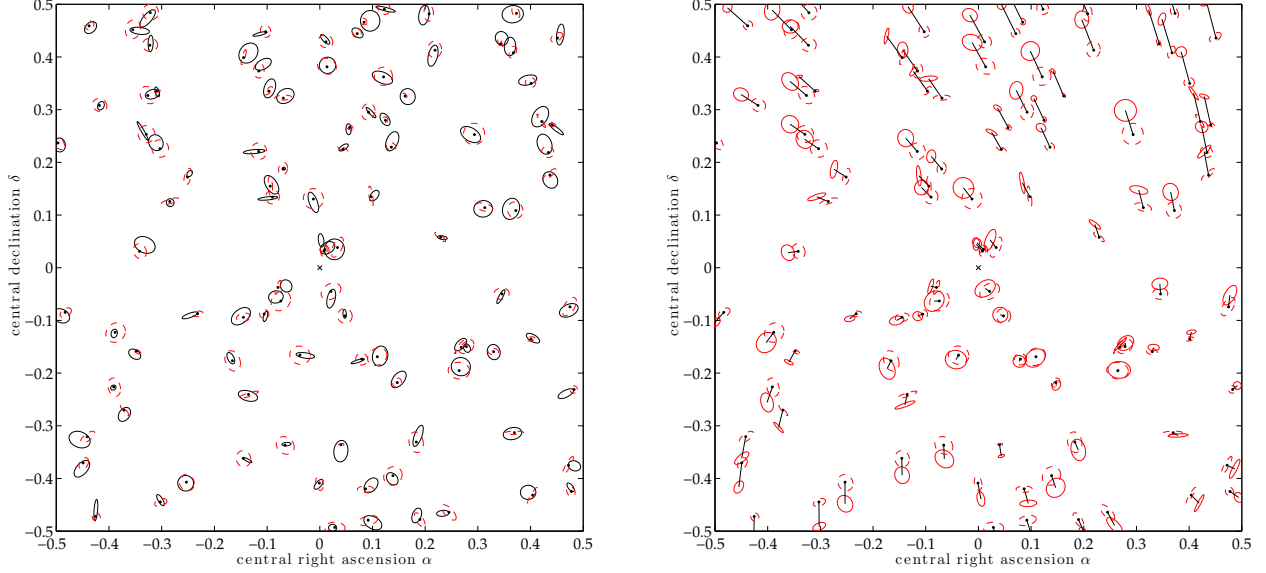
Fig. 3.— The maximum likelihood (ML) solution for the mapping between frames given the mock test data. *Left:* The *A* frame object distributions (black) with the ML map of the *B* frame tie object distributions superimposed. *Right:* The starting locations (points) are transformed according to equation (10) and used as the central positions for objects in the new image frame, which are then assigned random bivariate gaussian covariance matrices (ellipses).

Before considering the production of the the full likelihood distribution for the parameter set $\boldsymbol{\Theta}_b$, the 'best' mapping is examined in the form of the maximum likelihood value of $\mathcal{L}(\boldsymbol{\Theta}_b)$. This peak is found by minimizing $-2\log\mathcal{L}$, starting from the identity map $(\mathbf{0},0,\mathbf{1})$, a computation that is $\mathcal{O}(n)$ in the number of tie objects. [3]

This section described how this work incorporates the full MCMC estimate of the likelihood distribution. A chain on 50,000 steps was run on the mock data with loglikelihood

---

[3]As is usual for Nelder-Mead optimization, the resulting extremum is somewhat sensitive to the choice of starting point. The best-fitting translation seems quite robust, but the maximum likelihood scaling and rotation do vary a little. This is irrelevant in the MCMC prescription.

function D_B, with a burn-in period of 6,000 steps. The prior distributions for each of the five parameters $(\Delta x, \Delta y, \theta, \Lambda_x, \Lambda_y)$ are uniform with limits set equal to the limits on the input uniform random distributions in steps (ii) and (iv) of the mock data production scheme.

[4]. Figure 4 shows the progress of each parameter throughout the chain.

6.1.1. *Maximum likelihood solution*

The barycentric mean of the joint parameter sets over the length of the chain provides an estimate of the maximum likelihood solution:

$$\hat{\boldsymbol{\mu}} = \left[\widehat{\Delta x}, \widehat{\Delta y}, \hat{\theta}, \widehat{\Lambda_x}, \widehat{\Lambda_y}\right] = [0.0105, -0.0080, -0.0383, 0.9761, 0.9136];$$

$$\hat{\boldsymbol{\sigma}} \times 10^2 = [0.1414, 0.1384, 0.3425, 0.4924, 0.4396]$$

for reference, this mapping is satisfyingly close to the (inverse of) the random mapping used to transform the *A* frame object locations (prior to their being assigned *B* frame uncertainty distributions), which is

$$\left[-\Delta x, -\Delta y, -\theta, 1/\Lambda_x, 1/\Lambda_y\right] = [0.0102, -0.0086, -0.0391, 0.9782, 0.9116].$$

The task now is to convert the chain into a generalized mapping distribution that convert object locations or distributions in one image frame to distributions in the other. The approach taken here is to assume that the five-dimensional likelihood distribution of the mapping parameters is multivariate normal, so that the marginalized parameter distributions and pairwise covariance between parameters will determine the likelihood distribution in full.

---

[4]H. Haario, M. Laine, A. Mira and E. Saksman (2006), *DRAM: Efficient adaptive MCMC,* Statistics and Computing **16**, pp. 339-354; P. J. Green and A. Mira (2001b), *Delayed rejection in reversible jump Metropolis-Hastings,* Biometrika, Vol. 88, pp. 1035-1053; H. Haario, E. Saksman and J. Tamminen (2001), *An adaptive Metropolis algorithm,* Bernoulli **7**, pp. 223-242

Covariance matrix of likelihood distribution   The joint likelihood distributions of each

pair of mapping parameters is shown in Figure 5. Under the assumption that the mapping

likelihood distribution is multivariate normal, the covariance matrix elements Σ_{ij} are equal to

the covariance σ_{ij} between all pairs of parameters. This covariance matrix is:

$$C(i,j) \times 10^6 = \begin{pmatrix} 2.0065 & -0.075375 & 0.29147 & 0.22954 & -0.28978 \\ -0.075375 & 1.924 & -0.0012628 & -0.40617 & -0.9144 \\ 0.29147 & -0.0012628 & 11.7927 & -0.31664 & 0.65299 \\ 0.22954 & -0.40617 & -0.31664 & 24.0556 & -1.1182 \\ -0.28978 & -0.9144 & 0.65299 & -1.1182 & 19.3468 \end{pmatrix}$$

It is apparent that the parameters are mostly uncorrelated:

$$\frac{C(i,j)}{\sqrt{C(i,i)C(i,j)}} = \begin{pmatrix} 1 & -0.038363 & 0.059918 & 0.033039 & -0.04651 \\ -0.038363 & 1 & -0.00026512 & -0.059703 & -0.14988 \\ 0.059918 & -0.00026512 & 1 & -0.0188 & 0.043231 \\ 0.033039 & -0.059703 & -0.0188 & 1 & -0.051835 \\ -0.04651 & -0.14988 & 0.043231 & -0.051835 & 1 \end{pmatrix}$$

as is confirmed visually by the likelihood contours in the joint distribution, shown in Figure 5.

Nevertheless, it should be possible to proceed using the off-diagonal elements of the covariance

matrix in the computation of the mapping distribution for an arbitrary image point (x_0, y_0).

6.1.2. Distortion mapping

With the chain that has been generated, any point within the image plane can be mapped

to a set of transformed points, one for each step in the chain. The spread of these mapped

points within the image plane can be used to define the distortion distribution, here assumed

to be bivariate normal. Figure 6 shows this transformation for many such points within the

plane. [5]

6.2. High proper motion white dwarf

6.3. Adam's data

7. Discussion

7.1. Likelihood folding for proper motion and parallax

7.2. Bayesian cross-validation and false matches

8. Conclusion

We have produced a scheme for determining the likelihood distribution of mappings composed of translations, scalings and rotations between two frames in which tie objects are characterised by known bivariate normal distributions. MCMC is used to explore the likelihood distribution, which is found to be well described by a multivariate gaussian in the mapping parameters. With this model for the likelihood, a map of the distortion distribution within the image plane has been produced.

---

[5]The MCMC run should be 'thinned' using only every nth member of the chain for some $n > 1$. The motivation for this is the adjacent steps in the chain may be correlated and so the inclusion of all steps is non representative of the variance across the entire run. With this distortion map, the scale of the ellipses does increase as the chain is thinned. The map shown here is with quite substantial thinning; if no thinning is used the scale of the distortion distribution for each point is much smaller. I'm not sure whether this is to be regarded as correct (in that correlation in the chain is being removed) or wrong (in that the distortion map should be insensitive to thinning).
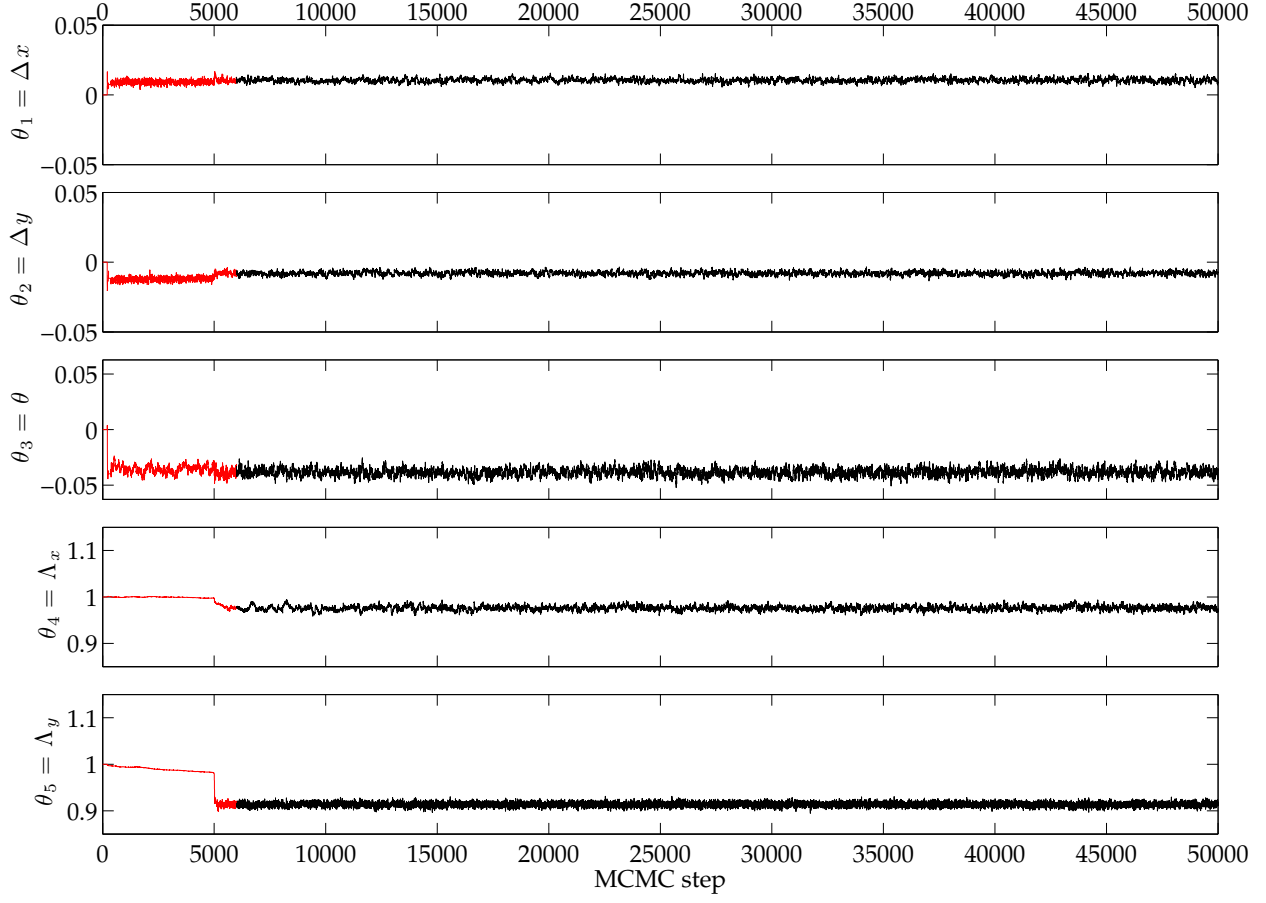
Fig. 4.— Chains for the translation $(\Delta x, \Delta y)$, rotation θ and scaling (Λ_x, Λ_y) parameters, with the burn-in period indicated with red colouring. The range of the parameter value axis has been narrowed from the width of the uniformed prior so that variability within the chain is visible.
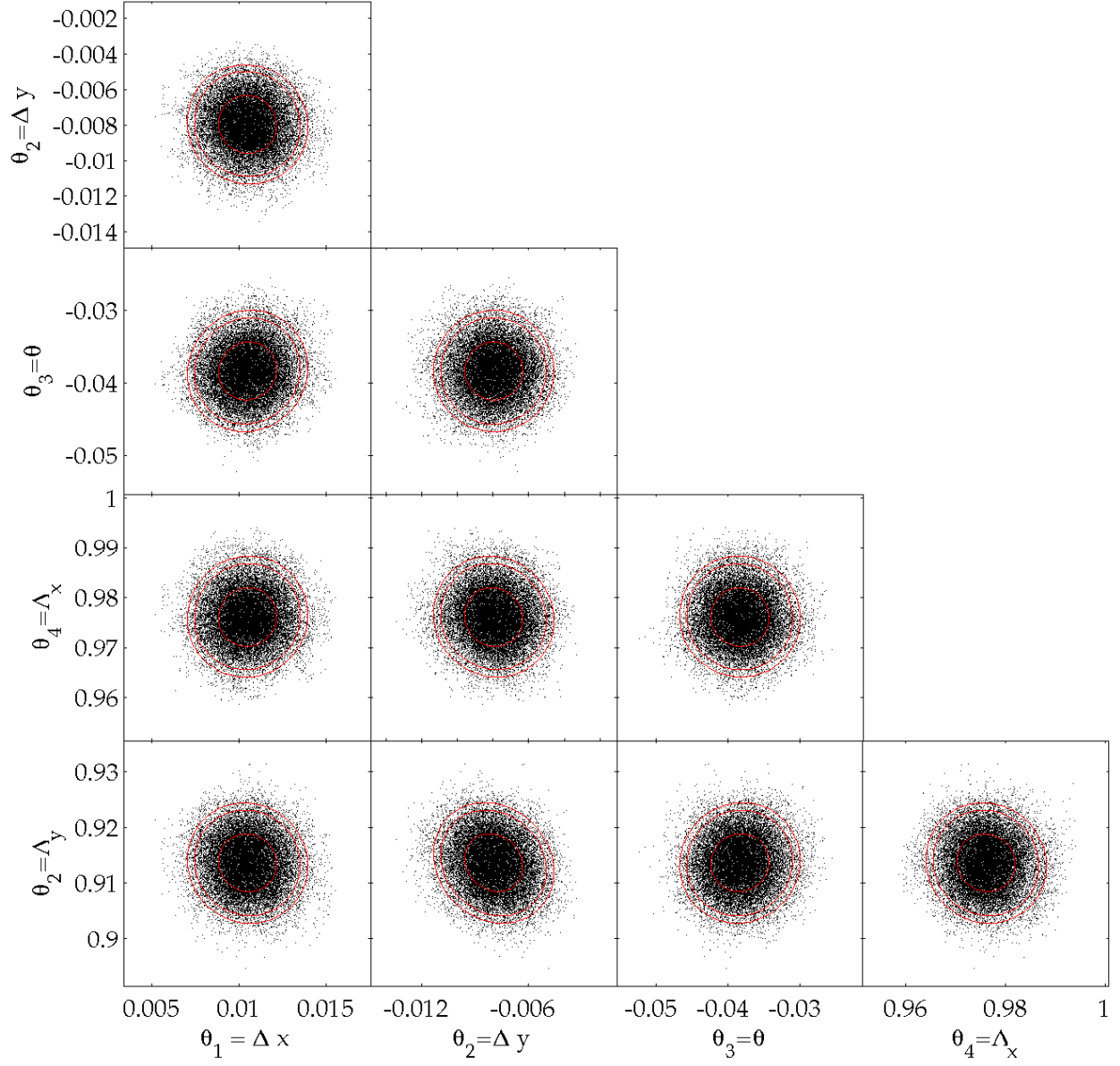
Fig. 5.— Pairwise joint likelihood distributions for transformation parameters, with each point representing a step in the chain and ellipses containing 50%, 90% and 95% of the distribution.
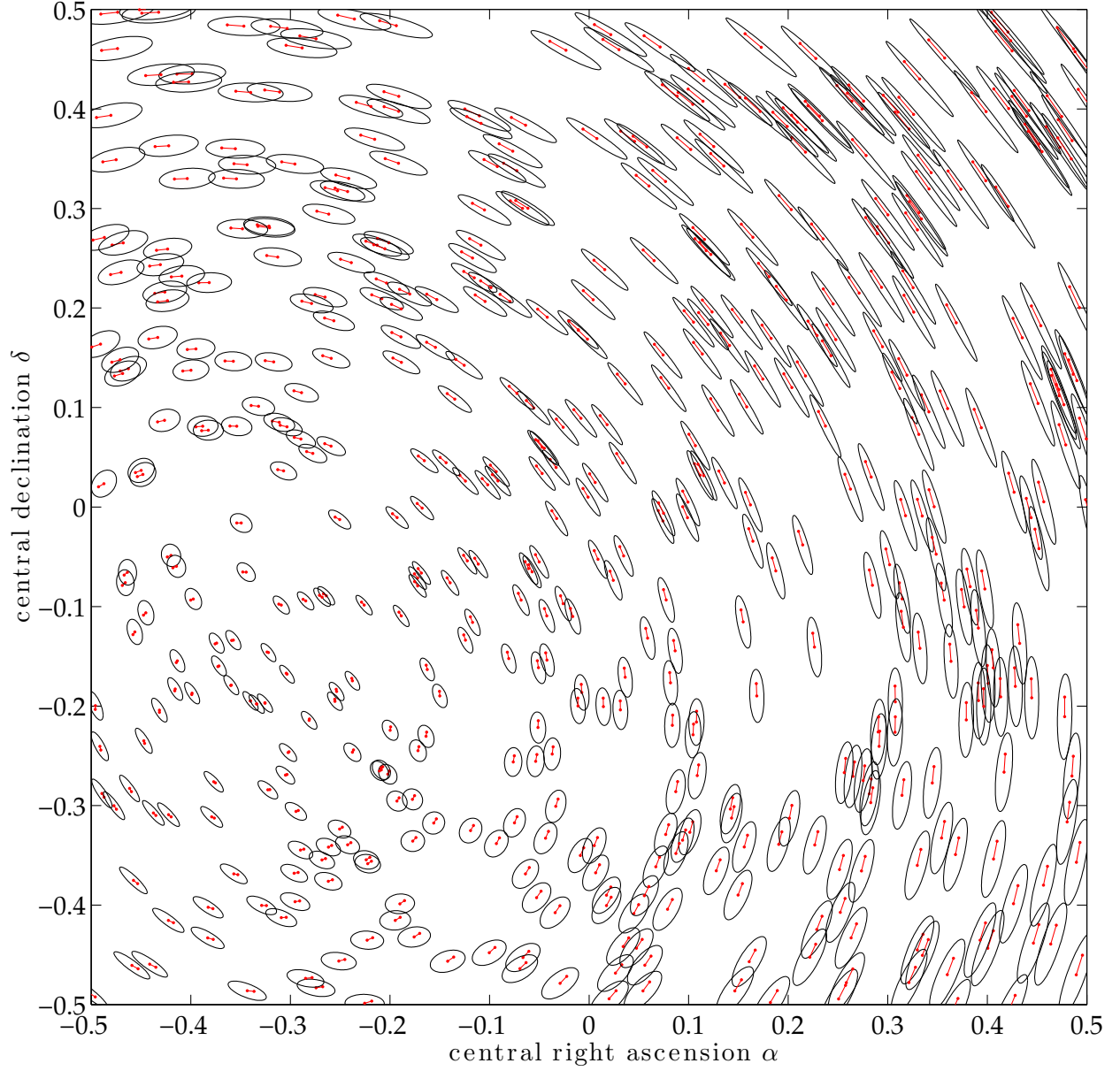
Fig. 6.— Distortion map corresponding to the likelihood distribution computer for the mock data. A random distribution of 500 points within the plane (off-centred red points) are mapped to a set of distributions (ellipses) with widths determined by the uncertainty in the mapping space.

REFERENCES

Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. 2008, in Advances in Neural Information Processing Systems 20, ed. J. Platt, D. Koller, Y. Singer, & S. Roweis (Cambridge, MA: MIT Press)

Gelman, A., Carlin, J., Stern, H., & Rubin, D. 2004, Bayesian Data Analysis (Chapman & Hall/CRC)

Rasmussen, C., & Williams, C. 2006, Gaussian Processes for Machine Learning, Adaptive Computation And Machine Learning (MIT Press)

A. Tutorial: probabilistic astrometry software

A.1. Interactive mode

A.2. Command line mode

B. Implementation details

This manuscript was prepared with the AAS LATEX macros v5.2.