

A probabilistic model to identify exoplanetary companions from multi-epoch astrometry

Coryn A.L. Bailer-Jones

Max Planck Institute for Astronomy, Heidelberg

<http://www.mpa.de/homes/calj>

2 October 2023 (first version 14 March 2023)

Table 1: Nomenclature

| | |
|--------------|---|
| ϖ | parallax |
| μ | proper motion |
| t | time |
| M_b | model: source is background object |
| M_c | model: source is companion |
| x_i, y_i | measured 2D position at time t_i |
| x'_i, y'_i | true 2D position at time t_i |
| s | subscript referring to the <i>source</i> |
| h | subscript referring to the <i>host</i> star |

1 Problem set up

From 2D astrometric measurements over time of a source near a star (hereafter “host star”), we wish to distinguish between two models:

M_c : The source is a gravitationally-bound companion to the host star. We assume that the orbital period is large compared to our span of observations such that we cannot detect orbital motion. Thus in the absence of noise we expect the host star and source to retain the same positions relative to one another over time.

M_b : The source is a background star, so will change its position relative to the host star according to the difference in proper motion and parallax of the host star and the background.

We could of course add further models, e.g. a companion with orbital motion.

We observe two or more 2D angular positions $\{x_i, y_i\}$ of a source in an Earth-based coordinate system at corresponding times $\{t_i\}$ for $i = 1 \dots N$. These could be parallel to the RA and Dec. axes, for example. These are noisy measurements, for which we assume there are corresponding true positions $\{x'_i, y'_i\}$. The times are noise-free.

On the assumption that the source undergoes unaccelerated motion relative to the solar system barycentre (SSB), and that the Earth’s orbit is circular, then the source’s true position at time t_i can be written as

$$x'_i = x'_0 + \mu_x t_i + \varpi g_x \sin(2\pi t_i / P_\oplus + \phi_x) \quad (1a)$$

$$y'_i = y'_0 + \mu_y t_i + \varpi g_y \sin(2\pi t_i / P_\oplus + \phi_y) . \quad (1b)$$

The sinusoidal terms give the projection of the parallax motion¹ along the x and y axes due to the orbit of the Earth (observer) about the Sun with period P_{\oplus} . g_x and g_y are geometric factors that depend on the direction of observation (assumed constant and known) and ϕ_x and ϕ_y are phase angles that are determined by the calendar date at $t = 0$ (i.e. where the source is in its parallax cycle). x'_0 and y'_0 are some zero point, which we will consider in the next section. An implementation of the parallactic model for equatorial coordinates is given in section 6.

2 Probabilistic forward model

If everything on the right side of equation 1 were noise-free, then the “true positions” on the left side would also be noise-free. However, in practice the parallax and proper motions are noisy measurements (which is why their symbols have no accent), which can be represented by a 3D likelihood distribution. Under M_c , the parallax and proper motions in equations 1 are those of the host star. These come directly from the Gaia catalogue for the host star in the form of point estimates, “errors” (actually uncertainties), and correlation coefficients. Together these determine the mean and covariance of a 3D Gaussian distribution. Under M_b the parallax and proper motions in equations 1 are for background stars. Their 3D distribution is obtained by fitting an empirical model to a set of background stars (perhaps of some magnitude or colour range) using a set of Gaia data. I assume this to be a 3D Gaussian distribution, but we could instead use something more complicated (e.g. a mixture of Gaussians).

The nomenclature “true positions” in equation 1 therefore refers to those positions we would obtain in the absence of noise in any measurement of positions, e.g. by measuring the centroid of a PSF on a detector. But there is still noise in the parallax and proper motions, and so $\{x'_i, y'_i\}$ has a distribution.

For both M_b and M_c , the joint parallax and proper motion distribution refers to the epoch of the Gaia catalogue, so the time in the probabilistic model in equations 1 is now the time since the Gaia catalogue epoch. By adopting $t_0 = 0$ as the Gaia epoch when the source is at (x'_0, y'_0) , we must add the offset $-\varpi g_x \sin \phi_x$ to the right side of equation 1a to ensure consistency (because we are not free to choose ϕ_x), and similarly for equation 1b. Those equations then become

$$x'_i = x'_0 + \mu_x t_i + \varpi s_x(t_i) \tag{2a}$$

$$y'_i = y'_0 + \mu_y t_i + \varpi s_y(t_i) \tag{2b}$$

¹If the Earth’s orbit about the Sun were an exact circle, then the parallactic motion of a star anywhere on the sky would in general be an ellipse. In the *ecliptic* coordinate system this ellipse always has one axis parallel to the ecliptic plane: if the star is on the ecliptic equator the ellipse is a line in the plane; if the star is at either ecliptic pole the ellipse is a circle centered on the pole.

where

$$s_x(t) = g_x [\sin(2\pi t/P_{\oplus} + \phi_x) - \sin \phi_x] \quad (3a)$$

$$s_y(t) = g_y [\sin(2\pi t/P_{\oplus} + \phi_y) - \sin \phi_y] . \quad (3b)$$

We have so far only been discussing the noise-free expected positions of the source, not the measurements. To tie this model to the measurement coordinate system, we specify that our first measurement of the source (at t_1) is at $(x_1, y_1) = (x'_1, y'_1)$, the value of which is arbitrary (and we might set it to $(0, 0)$ for convenience). The uncertainties (x'_1, y'_1) will appear in the inference later. The position (x_0, y_0) – a measurement at the Gaia epoch – will never be required as it won't be included in the likelihood below. The corresponding true position (x'_0, y'_0) could be computed from equation 2 as a function of (x'_1, y'_1) , but will also not be required. Hence we are not relying on the Gaia *position* measurements at any epoch, just Gaia's measurements of the parallax and proper motion at epoch $t_0 = 0$. This is important because we do not have Gaia positions for the source.

We can now write the PDF of the true positions – a probabilistic forward model – as

$$P(x'_i | t_i, M(\varpi, \mu)) \quad (4a)$$

$$P(y'_i | t_i, M(\varpi, \mu)) \quad (4b)$$

where M can be M_b or M_c to indicate which parallax and proper motions are meant. The dependence on the constants g_x, g_y, ϕ_x, ϕ_y , and P_{\oplus} in these equations is kept implicit for brevity.

To determine these PDFs from equation 2, we would often have to use cumulative distribution functions or even Monte Carlo. But as the distributions of the parallax and proper motion are Gaussian, then given that equations 2 are linear in parallax and proper motion, the resulting PDFs are also Gaussian, with variances (see e.g. Bailer-Jones 2017, section 2.8.3)

$$\text{Var}(x'_i) = t_i^2 \text{Var}(\mu_x) + s_x(t_i)^2 \text{Var}(\varpi) + 2t_i s_x(t_i) \text{Cov}(\varpi, \mu_x) \quad (5a)$$

$$\text{Var}(y'_i) = t_i^2 \text{Var}(\mu_y) + s_y(t_i)^2 \text{Var}(\varpi) + 2t_i s_y(t_i) \text{Cov}(\varpi, \mu_y) . \quad (5b)$$

Under M_c , $\text{Var}(\varpi)$ is just the square of the “error” term in the Gaia catalogue for the host star, and $\text{Cov}(\varpi, \mu_x)$ is the product of the two “errors” and their covariance, and similarly for the other terms. Note that the variance of the true position grows with time in proportion to the proper motion variance. This is because the fixed uncertainty in the proper motion corresponds to an ever growing uncertainty in position the further in time from the Gaia epoch that we predict the source's position.

Below we will also need the covariance between x'_i and y'_i , which we can compute

using the bilinearity property of covariance²

$$\begin{aligned} \text{Cov}(x'_i, y'_i) &= t_i^2 \text{Cov}(\mu_x, \mu_y) + s_x(t_i)s_y(t_i)\text{Var}(\varpi) + \\ &\quad t_i s_y(t_i)\text{Cov}(\varpi, \mu_x) + t_i s_x(t_i)\text{Cov}(\varpi, \mu_y) \end{aligned} \quad (6)$$

where $\text{Cov}(\varpi, \varpi) \equiv \text{Var}(\varpi)$ and the covariance between a constant a random variable is zero.

2.1 Relative motion

We will usually measure the position of the source relative to the host star, given by

$$\Delta x'_i = x'_{s,i} - x'_{h,i} \quad (7a)$$

$$\Delta y'_i = y'_{s,i} - y'_{h,i} \quad (7b)$$

where the s subscript refers to the source, and the h subscript refers to the host star. In both cases their motion is described by equation 2.

To tie the model to the measurement coordinate system, we specify that our first measurement of the source (at t_1) is at $(\Delta x_1, \Delta y_1) = (\Delta x'_1, \Delta y'_1)$, similar to before.

Under M_c , $(\Delta x'_i, \Delta y'_i) = (\Delta x', \Delta y') = \text{constant}$. This is because even though the true parallax and proper motion of the host star have non-zero variance–covariance, under M_c the true position of the source is fixed, and so has zero variance–covariance. It follows from our alignment of the model and measurement coordinate systems that $(\Delta x', \Delta y') = (\Delta x_1, \Delta y_1)$.

Under M_b , the source moves according to the background proper motion and parallax (call these $\varpi_b, \mu_{b,x}, \mu_{b,y}$), and the host star moves according to its proper motion and parallax (call these $\varpi_h, \mu_{h,x}, \mu_{h,y}$). Thus

$$\Delta x'_i = (x'_{s,0} - x'_{h,0}) + (\mu_{b,x} - \mu_{h,x})t_i + (\varpi_b - \varpi_h)s_x(t_i) \quad (8a)$$

$$\Delta y'_i = (y'_{s,0} - y'_{h,0}) + (\mu_{b,y} - \mu_{h,y})t_i + (\varpi_b - \varpi_h)s_y(t_i) \quad (8b)$$

where $(x'_{s,0}, y'_{s,0})$ and $(x'_{h,0}, y'_{h,0})$ are the positions³ of the source and the host star respectively at time t_0 . As before, these positions will not be required in the inference, because they follow from our choice that at t_1 , $(\Delta x_1, \Delta y_1) = (\Delta x', \Delta y')$.

If we assume there is no covariance between the background astrometry and the host star astrometry⁴, then any covariance terms involving b and h are zero, and also $\text{Var}(\varpi_b -$

²For constants a, b, c, d , $\text{Cov}(aX + bY, cW + dZ) = ac\text{Cov}(X, W) + ad\text{Cov}(X, Z) + bc\text{Cov}(Y, W) + bd\text{Cov}(Y, Z)$.

³The position of the source is not with the b subscript: there is no position of the background as this is a set of sources. The position of the source and host at time t_1 is the same under any model, because this is the epoch we use to tie the measurements to the model predictions.

⁴This is not generally true, given that the model for the background and the data on the host star are both drawn from Gaia, and that Gaia astrometry shows correlations between sources separated by small angles on the sky.

$\varpi_h) = \text{Var}(\varpi_b) + \text{Var}(\varpi_h)$, and likewise for the variance of the two components of the proper motion. The variance and covariance of the true relative positions are therefore

$$\begin{aligned} \text{Var}(\Delta x'_i) &= t_i^2 [\text{Var}(\mu_{b,x}) + \text{Var}(\mu_{h,x})] + s_x(t_i)^2 [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad 2t_i s_x(t_i) [\text{Cov}(\varpi_b, \mu_{b,x}) + \text{Cov}(\varpi_h, \mu_{h,x})] \end{aligned} \quad (9a)$$

$$\begin{aligned} \text{Var}(\Delta y'_i) &= t_i^2 [\text{Var}(\mu_{b,y}) + \text{Var}(\mu_{h,y})] + s_y(t_i)^2 [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad 2t_i s_y(t_i) [\text{Cov}(\varpi_b, \mu_{b,y}) + \text{Cov}(\varpi_h, \mu_{h,y})] \end{aligned} \quad (9b)$$

$$\begin{aligned} \text{Cov}(\Delta x'_i, \Delta y'_i) &= t_i^2 [\text{Cov}(\mu_{b,x}, \mu_{b,y}) + \text{Cov}(\mu_{h,x}, \mu_{h,y})] + \\ &\quad s_x(t_i) s_y(t_i) [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad t_i s_y(t_i) [\text{Cov}(\varpi_b, \mu_{b,x}) + \text{Cov}(\varpi_h, \mu_{h,x})] + \\ &\quad t_i s_x(t_i) [\text{Cov}(\varpi_b, \mu_{b,y}) + \text{Cov}(\varpi_h, \mu_{h,y})] . \end{aligned} \quad (9c)$$

3 Probability of data given model

3.1 Direct measurements

For ease of explanation, let us first consider that we measure directly the position of the source. In subsection 3.2 we'll see the small modifications required when we measure the position of the source relative to the host star. Furthermore, we will just consider the probability of the data at one epoch. In section 4 we'll see how to generalize this to the case of multiple measurement epochs.

Each measured position ($i = 1 \dots N$) can be described by a 2D PDF conditional on the true positions. We'll assume this is a Gaussian with mean (x'_i, y'_i) and covariance C_i , which we write as

$$P(x_i, y_i | x'_i, y'_i) = \mathcal{N}(x_i, y_i; (x'_i, y'_i), C_i) . \quad (10)$$

The covariance comes from the instrument model used to make the position measurements. I'll refer to the above probability as the *likelihood*. Recall that for the first position we set $(x_1, y_1) = (x'_1, y'_1)$ to connect the true and measured coordinate systems.

To distinguish between two or more models using the measurements, we can compute $P(x_i, y_i | t_i, M(\varpi, \mu))$. This comes from the above quantities via a marginalization

$$P(x_i, y_i | t_i, M(\varpi, \mu)) = \int_{x'_i, y'_i} P(x_i, y_i, x'_i, y'_i | t_i, M(\varpi, \mu)) dx'_i dy'_i \quad (11a)$$

$$= \int_{x'_i, y'_i} P(x_i, y_i | x'_i, y'_i) P(x'_i, y'_i | t_i, M(\varpi, \mu)) dx'_i dy'_i \quad (11b)$$

I'll refer to this as the *convolved likelihood*, for a reason that will become clear in a moment. The first term under the integral is the likelihood (equation 10); t_i and $M(\varpi, \mu)$ don't appear in this due to conditional independence. The second term is the joint PDF

of the true measurements, which I postulate to be a 2D Gaussian with mean given by the right side of equation 2, and covariance matrix

$$C'_i = \begin{bmatrix} \text{Var}(x'_i) & \text{Cov}(x'_i, y'_i) \\ \text{Cov}(x'_i, y'_i) & \text{Var}(y'_i) \end{bmatrix} \quad (12)$$

with the terms given by equations 5 and 6.

Equation 11b is a convolution of two Gaussians, which is another Gaussian. In 1D this can be written

$$\int_{x'} f(x - x')h(x')dx' = f(x) * h(x) \quad (13)$$

the mean of which is the sum of the means of f and h , and the covariance the sum of the covariances of f and h . In our case $f(x - x')$ corresponds to the likelihood, but rewritten to have argument $x - x'$ and mean zero. $h(x)$ is the probabilistic forward model. Thus $P(x_i, y_i | t_i, M(\varpi, \mu))$ in equation 11 is a 2D Gaussian in the measurement (x_i, y_i) with mean (x'_i, y'_i) given by equation 2 and covariance matrix $C_i + C'_i$.

Given the measurements (and their uncertainties and any covariance), as well as the parallax and proper motions (mean, variance, covariance) of the model, we can compute the convolved likelihood (a scalar) for both models. Each is the probability density of the data given the model, and their absolute values are meaningless. But their ratio, the *odds ratio*⁵

$$r_{cb} = \frac{P(x_i, y_i | t_i, M_c(\varpi, \mu))}{P(x_i, y_i | t_i, M_b(\varpi, \mu))} \quad (14)$$

expressed which model is preferred by the data: above 1 and M_c is preferred.

From Bayes' theorem the posterior probability of the model is proportional to the product of the convolved likelihood and the model prior probability, i.e.

$$P(M_c(\varpi, \mu) | x_i, y_i, t_i) \propto P(x_i, y_i | t_i, M_c(\varpi, \mu)) P(M_c) \quad (15)$$

where t_i in the second term drops out due to conditional independence. The ratio of these is the *posterior odds ratio*

$$p_{cb} = \frac{P(x_i, y_i | t_i, M_c(\varpi, \mu)) P(M_c)}{P(x_i, y_i | t_i, M_b(\varpi, \mu)) P(M_b)}. \quad (16)$$

This metric is preferred if we have knowledge of the ratio of the prior probabilities of the two models. That prior ratio might reflect the number density of stars in the background, for example. In the unlikely event that M_b and M_c are the only two possible models (so they are mutually exclusive and exhaustive), then the posterior probability

⁵Given that we consider the model parameters to be in the inference, then we could also refer to this as the *Bayes factor*.

of M_c is $(1 + 1/p_{cb})^{-1}$.

Of course, with just one measurement epoch we won't be able to say much useful in practice about the models (especially as we connect the measurement and model reference frames using the first measurement). But everything we have done here forms the basis for generalizing this.

3.2 Relative measurements

If our measurements are now the positions of the source relative to the host star, $(\Delta x_i, \Delta y_i)$, the logic of the previous subsection remains the same, but some changes are required to the expressions.

The x and y terms in equations 10 and 11 become Δx and Δy respectively, for both the primed and unprimed terms. The covariance in equation 10 is now replaced with the covariance in the relative measurement, Γ_i . The second term inside the integral in equation 11b, $P(\Delta x'_i, \Delta y'_i | t_i, M)$ is a Gaussian with mean $(\Delta x'_i, \Delta y'_i)$ given by equation 8. Let us consider the two models.

Under M_c , $(\Delta x'_i, \Delta y'_i) = (\Delta x', \Delta y') = \text{constant}$. The covariance is zero, and so $P(\Delta x'_i, \Delta y'_i | t_i, M_c)$ is just a delta function. The resulting convolution, and hence $P(\Delta x_i, \Delta y_i | t_i, M_c)$, is a Gaussian in the measurement $(\Delta x_i, \Delta y_i)$ with mean $= (\Delta x', \Delta y')$ and covariance Γ_i .

Under M_b , $(\Delta x'_i, \Delta y'_i)$ is given by equation 8. This is the mean of $P(\Delta x'_i, \Delta y'_i | t_i, M_b)$ which has covariance matrix

$$\Gamma'_i = \begin{bmatrix} \text{Var}(\Delta x'_i) & \text{Cov}(\Delta x'_i, \Delta y'_i) \\ \text{Cov}(\Delta x'_i, \Delta y'_i) & \text{Var}(\Delta y'_i) \end{bmatrix}, \quad (17)$$

the terms of which are given by equation 9. Once convolved with the likelihood, we get that $P(\Delta x'_i, \Delta y'_i | t_i, M_b)$ is a Gaussian in the measurement $(\Delta x_i, \Delta y_i)$ with mean $(\Delta x'_i, \Delta y'_i)$ from equation 8 and covariance $\Gamma_i + \Gamma'_i$.

4 Multiple measurement epochs

For a set of N measurements of the source's position over time, $(\mathbf{x}, \mathbf{y}) = \{x_i, y_i\}$ at times $\mathbf{t} = \{t_i\}$, we can generalize equation 11 to be

$$P(\mathbf{x}, \mathbf{y} | \mathbf{t}, M(\varpi, \mu)) = \int_{\mathbf{x}', \mathbf{y}'} P(\mathbf{x}, \mathbf{y} | \mathbf{x}', \mathbf{y}') P(\mathbf{x}', \mathbf{y}' | \mathbf{t}, M(\varpi, \mu)) d\mathbf{x}' d\mathbf{y}' \quad (18)$$

which is a $2N$ -dimensional integral. Even if the N epochs are independently measured, this cannot be written as a product of N 2D integrals, because the true coordinates at the different times all depend on the same parallax and proper motion so are highly correlated. Thus the second term under the integral is a $2N$ -dimensional Gaussian

with mean $(\mathbf{x}', \mathbf{y}')$ given by equation 2 and covariance matrix (which I write in order $(x'_1, y'_1, x'_2, y'_2, \dots)$)

$$\Sigma' = \begin{bmatrix} \text{Var}(x'_1) & \text{Cov}(x'_1, y'_1) & \text{Cov}(x'_1, x'_2) & \text{Cov}(x'_1, y'_2) & \dots \\ & \text{Var}(y'_1) & \text{Cov}(y'_1, x'_2) & \text{Cov}(y'_1, y'_2) & \dots \\ & & \text{Var}(x'_2) & \text{Cov}(x'_2, y'_2) & \dots \\ & & & \text{Var}(y'_2) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (19)$$

where the on-diagonal elements are given by equation 5 and the off-diagonal elements are again computed from equation 2 (expressed also in terms of j) to be

$$\text{Cov}(x'_i, x'_j) = t_i t_j \text{Var}(\mu_x) + s_x(t_i) s_x(t_j) \text{Var}(\varpi) + [t_i s_x(t_j) + t_j s_x(t_i)] \text{Cov}(\varpi, \mu_x) \quad (20a)$$

$$\text{Cov}(y'_i, y'_j) = t_i t_j \text{Var}(\mu_y) + s_y(t_i) s_y(t_j) \text{Var}(\varpi) + [t_i s_y(t_j) + t_j s_y(t_i)] \text{Cov}(\varpi, \mu_y) \quad (20b)$$

$$\text{Cov}(x'_i, y'_j) = t_i t_j \text{Cov}(\mu_x, \mu_y) + s_x(t_i) s_y(t_j) \text{Var}(\varpi) + t_i s_y(t_j) \text{Cov}(\varpi, \mu_x) + t_j s_x(t_i) \text{Cov}(\varpi, \mu_y). \quad (20c)$$

Only the upper triangle of the matrix is shown for clarity: the lower triangle follows from the fact that the covariance matrix is symmetric.

We can then apply the convolution theorem again and determine that $P(\mathbf{x}, \mathbf{y} | \mathbf{t}, M(\varpi, \mu))$ in equation 18 is a $2N$ -dimensional Gaussian with mean $(\mathbf{x}', \mathbf{y}')$ given by equation 2 and covariance matrix $\Sigma + \Sigma'$, where Σ is the $2N$ -dimensional Gaussian covariance of the N measurements, which is given by the instrument model. Assuming the different epochs are measured independently, then Σ would be a banded matrix with most off-diagonal elements zero: only the off-diagonals corresponding to x and y of the same epoch would be non-zero. Given the data and the model, this convolved likelihood again evaluates to a scalar value for each model, from which we can compute the (posterior) odds ratio in the same way as in the single measurement case.

4.1 Relative measurements

The generalization to relative measurements (sections 2.1 and 3.2) is conceptually straight forward. Everything remains Gaussian, we just have different terms in the expressions for the (co)variances. Equation 18 becomes

$$P(\Delta \mathbf{x}, \Delta \mathbf{y} | \mathbf{t}, M(\varpi, \mu)) = \int_{\Delta \mathbf{x}', \Delta \mathbf{y}'} P(\Delta \mathbf{x}, \Delta \mathbf{y} | \Delta \mathbf{x}', \Delta \mathbf{y}') P(\Delta \mathbf{x}', \Delta \mathbf{y}' | \mathbf{t}, M(\varpi, \mu)) d\Delta \mathbf{x}' d\Delta \mathbf{y}'. \quad (21)$$

Under M_c , $(\Delta x'_i, \Delta y'_i) = (\Delta x', \Delta y')$ for all i , so just as with two measurements, $P(\Delta \mathbf{x}', \Delta \mathbf{y}' | \mathbf{t}, M_c)$ is a delta function. The resulting convolution, and hence $P(\Delta \mathbf{x}, \Delta \mathbf{y} | \mathbf{t}, M_c)$, is a Gaus-

sian in the measurement $(\Delta \mathbf{x}, \Delta \mathbf{y})$ with mean $= (\Delta \mathbf{x}', \Delta \mathbf{y}')$ and covariance equal to the covariance in the relative measurements, Σ_Δ .

Under M_b , the second term under the integral in equation 21 is a $2N$ -dimensional Gaussian with mean $(\Delta \mathbf{x}', \Delta \mathbf{y}')$ given by equation 8. The covariance matrix (which I write in order $(\Delta x'_1, \Delta y'_1, \Delta x'_2, \Delta y'_2, \dots)$) is

$$\Sigma'_\Delta = \begin{bmatrix} \text{Var}(\Delta x'_1) & \text{Cov}(\Delta x'_1, \Delta y'_1) & \text{Cov}(\Delta x'_1, \Delta x'_2) & \text{Cov}(\Delta x'_1, \Delta y'_2) & \dots \\ & \text{Var}(\Delta y'_1) & \text{Cov}(\Delta y'_1, \Delta x'_2) & \text{Cov}(\Delta y'_1, \Delta y'_2) & \dots \\ & & \text{Var}(\Delta x'_2) & \text{Cov}(\Delta x'_2, \Delta y'_2) & \dots \\ & & & \text{Var}(\Delta y'_2) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (22)$$

where the on-diagonal elements, as well as the off-diagonals with a common epoch, are given by equations 9, and the other off-diagonal elements, i.e. those between different epochs, we can generalize from equation 20 (or compute directly from equation 8 for two different epochs) to give

$$\begin{aligned} \text{Cov}(\Delta x'_i, \Delta x'_j) &= t_i t_j [\text{Var}(\mu_{b,x}) + \text{Var}(\mu_{h,x})] + \\ &\quad s_x(t_i) s_x(t_j) [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad [t_i s_x(t_j) + t_j s_x(t_i)] [\text{Cov}(\varpi_b, \mu_{b,x}) + \text{Cov}(\varpi_h, \mu_{h,x})] \end{aligned} \quad (23a)$$

$$\begin{aligned} \text{Cov}(\Delta y'_i, \Delta y'_j) &= t_i t_j [\text{Var}(\mu_{b,y}) + \text{Var}(\mu_{h,y})] + \\ &\quad s_y(t_i) s_y(t_j) [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad [t_i s_y(t_j) + t_j s_y(t_i)] [\text{Cov}(\varpi_b, \mu_{b,y}) + \text{Cov}(\varpi_h, \mu_{h,y})] \end{aligned} \quad (23b)$$

$$\begin{aligned} \text{Cov}(\Delta x'_i, \Delta y'_j) &= t_i t_j [\text{Cov}(\mu_{b,x}, \mu_{b,y}) + \text{Cov}(\mu_{h,x}, \mu_{h,y})] + \\ &\quad s_x(t_i) s_y(t_j) [\text{Var}(\varpi_b) + \text{Var}(\varpi_h)] + \\ &\quad t_i s_y(t_j) [\text{Cov}(\varpi_b, \mu_{b,x}) + \text{Cov}(\varpi_h, \mu_{h,x})] + \\ &\quad t_j s_x(t_i) [\text{Cov}(\varpi_b, \mu_{b,y}) + \text{Cov}(\varpi_h, \mu_{h,y})]. \end{aligned} \quad (23c)$$

$P(\Delta \mathbf{x}, \Delta \mathbf{y} | \mathbf{t}, M_b)$ is therefore a $2N$ -dimensional Gaussian with mean $(\Delta \mathbf{x}', \Delta \mathbf{y}')$ given by equation 2 and covariance matrix $\Sigma_\Delta + \Sigma'_\Delta$.

5 A special case: two epochs, neglect parallax

If we assume the parallaxes are negligible, or if we are observing in the ICRS where parallaxes are zero, then equation 2 becomes

$$x'_i = x'_0 + \mu_x t_i \quad (24a)$$

$$y'_i = y'_0 + \mu_y t_i. \quad (24b)$$

Assume further that we only have two measurement epochs $i = 1, 2$, and that we use these to derive a measured proper motion (m_x, m_y) . The true proper motion (m'_x, m'_y)

is

$$m'_x \equiv \frac{x'_2 - x'_1}{t_2 - t_1} = \mu_x \quad (25a)$$

$$m'_y \equiv \frac{y'_2 - y'_1}{t_2 - t_1} = \mu_y. \quad (25b)$$

The joint PDF of these is a 2D Gaussian

$$P(m'_x, m'_y | M(\mu)) = \mathcal{N}(m'_x, m'_y; (\mu_x, \mu_y), C_\mu) \quad \text{where} \quad (26)$$

$$C_\mu = \begin{bmatrix} \text{Var}(\mu_x) & \text{Cov}(\mu_x, \mu_y) \\ \text{Cov}(\mu_x, \mu_y) & \text{Var}(\mu_y) \end{bmatrix}.$$

Our measurements are now derived by taking the difference between two positions and dividing by the time between them, the noise-free model for which is equation 25. To be clear, if we just had one component then the measurement would be $m_x = (x_2 - x_1)/(t_2 - t_1)$. This is linear in the positional measurements so variance propagation is simple, and done as before. The likelihood of the 2D measurement is therefore

$$P(m_x, m_y | m'_x, m'_y) = \mathcal{N}(m_x, m_y; (m'_x, m'_y), C'_\mu) \quad \text{where} \quad (27)$$

$$C'_\mu = \frac{1}{(t_2 - t_1)^2} \begin{bmatrix} \text{Var}(x_1) + \text{Var}(x_2) & \left\{ \begin{array}{l} \text{Cov}(x_1, y_1) + \text{Cov}(x_2, y_2) + \\ \text{Cov}(x_1, y_2) + \text{Cov}(x_2, y_1) \end{array} \right\} \\ \left\{ \begin{array}{l} \text{Cov}(x_1, y_1) + \text{Cov}(x_2, y_2) + \\ \text{Cov}(x_1, y_2) + \text{Cov}(x_2, y_1) \end{array} \right\} & \text{Var}(y_1) + \text{Var}(y_2) \end{bmatrix}$$

All of the elements of the covariance matrix come from the measurement model. If the two epochs are measured independently (no instrument correlations over time), then $\text{Cov}(x_1, y_2) = \text{Cov}(x_2, y_1) = 0$.

The convolved likelihood (cf. equation 11b) is then

$$P(m_x, m_y | M(\mu)) = \int_{m'_x, m'_y} P(m_x, m_y | m'_x, m'_y) P(m'_x, m'_y | M(\mu)) dm'_x dm'_y \quad (28)$$

which again is a convolution, the result of which is a Gaussian of mean (μ_x, μ_y) and covariance $C_\mu + C'_\mu$.

6 Parallaxic model in equatorial coordinates

So far we have considered an arbitrary coordinate system with motion defined by equation 1, where x and y are Cartesian coordinates at a tangent-plane projection on the sky.

Let us now consider the equatorial coordinate system, $(\alpha, \delta) = (\text{RA}, \text{declination})$, so

that for small displacements in the tangent-plane

$$x = \Delta\alpha \cos \delta \quad (29a)$$

$$y = \Delta\delta. \quad (29b)$$

Let (X, Y, Z) be the Cartesian coordinates of the Earth's centre (in units au) in the equatorial coordinate system centered on the SSB. The displacement of a star with parallax ϖ (in units arcseconds) at position (α, δ) as seen from the Earth's centre compared to its barycentric position is (Green 1985, equation 8.15 on p. 189 (also equation 12.17); Kovalevsky & Seidelmann 2004, p. 132 equation 6.21)

$$\Delta\alpha \cos \delta = \varpi(X \sin \alpha - Y \cos \alpha) \quad (30a)$$

$$\Delta\delta = \varpi(X \cos \alpha \sin \delta + Y \sin \alpha \sin \delta - Z \cos \delta). \quad (30b)$$

Note that we are neglecting shifts due to aberration, as we assume that these have already been adjusted for.⁶ These two expressions could now be used in place of $\varpi s_x(t_i)$ $\varpi s_x(t_i)$ in equations 2 and 3. At high precision the Earth's orbit is elliptical and perturbed by the Moon, so there is no simple expression for the time dependence of (X, Y, Z) . But if we approximate the Earth's orbit (strictly the observer's orbit) as circular in the ecliptic plane with an inclination of ε to the equatorial plane, then (Green p. 291)

$$(X, Y, Z) = (\cos \lambda, \sin \lambda \cos \varepsilon, \sin \lambda \sin \varepsilon) \quad (31)$$

where λ is the ecliptic longitude of the Earth measured from the vernal equinox Υ . Substituting this into equation 30 we get

$$\Delta\alpha \cos \delta = \varpi(\cos \lambda \sin \alpha - \sin \lambda \cos \varepsilon \cos \alpha) \quad (32a)$$

$$\Delta\delta = \varpi(\cos \lambda \cos \alpha \sin \delta - \sin \lambda[\sin \varepsilon \cos \delta - \cos \varepsilon \sin \alpha \sin \delta]). \quad (32b)$$

For a circular orbit, λ is just a measure of time from the vernal equinox, t_Υ , which occurs around 20 March. In our probabilistic model (section 2) we want $t_0 = 0$ to be the Gaia epoch, so time needs to be measured in tropical years from the calendar date of the Gaia epoch. If that is 1 January, for example, then $\lambda = 2\pi[t - t_\Upsilon]$, with $t_\Upsilon \simeq 80/365.242$.

We can now rewrite equation 2 as

$$\Delta\alpha' \cos \delta = \Delta\alpha'_0 \cos \delta + \mu_\alpha t_i + \varpi s_\alpha(t_i) \quad (33a)$$

$$\Delta\delta' = \Delta\delta'_0 + \mu_\delta t_i + \varpi s_\delta(t_i) \quad (33b)$$

⁶Aberration is a large effect, $\pm 21''$, so cannot be neglected.

where

$$s_\alpha(t_i) = s_{\alpha,0} + \sin \alpha \cos(2\pi[t - t_\gamma]) - \cos \varepsilon \cos \alpha \sin(2\pi[t - t_\gamma]) \quad (34a)$$

$$s_\delta(t_i) = s_{\delta,0} + \cos \alpha \sin \delta \cos(2\pi[t - t_\gamma]) - [\sin \varepsilon \cos \delta - \cos \varepsilon \sin \alpha \sin \delta] \sin(2\pi[t - t_\gamma]) . \quad (34b)$$

$(\Delta\alpha' \cos \delta, \Delta\delta')$ are the coordinates of the source at time t_i in a tangent-plane projection. I am now using primes for these “true” quantities in our probabilistic model (everything in equations 34 are taken as noise-free). $s_{\alpha,0}$ and $s_{\delta,0}$ have been introduced to ensure that $s_\alpha(t_0) = s_\delta(t_0) = 0$ at $t_0 = 0$, the Gaia epoch, which is 2016.0 for Gaia DR3. In these equations time is in tropical years⁷, and $\Delta\alpha'$ and $\Delta\delta$ and parallax and proper motion are in common angular coordinates, typically mas.

Equations 33 and 34 have the same time, parallax, and proper motion dependence as equations 2 and 3, so the probabilistic model developed in sections 2 to 4 applies as before. The role of the constants $g_x, g_y, \phi_x,$ and ϕ_y is now taken by $\alpha, \delta, \varepsilon,$ and t_γ . P_\oplus , the tropical year, is implicit in the measurement of time. Again we specify that our first measurement of the source (at t_1) is at $(\Delta\alpha_1 \cos \delta, \Delta\delta_1) = (\Delta\alpha'_1 \cos \delta, \Delta\delta'_1)$, which we can set to $(0, 0)$ as the centre of the tangent-plane projection for convenience.

⁷Proper motions are usually given in angle per Julian year (365.25 days), so strictly we need to make a small correction here.

7 References

Bailer-Jones, C.A.L., 2017, *Practical Bayesian Inference*, Cambridge University Press

Green, R.M., 1985, *Spherical Astronomy*, Cambridge University Press

Kovalevsky, J. & Seidelmann, P.K., 2004, *Fundamentals of Astronomy*, Cambridge University Press