Photometric Data Analysis for GAIA: Mathematical Formulation

Anthony Brown
Leiden Observatory

June 12, 2001

GAIA–AB–01

Abstract. This report describes a linear least squares approach to the creation of algorithms and software for the photometric data analysis for GAIA. The emphasis is on the mathematical formulation of the problem. The ideas described here are essentially a more detailed version of the proposals put forward in section 9.4.2 of the GAIA Concept and Technology Study Report (ESA, 2000). The algorithms will be tested on simulated GAIA data. The assumptions underlying the method described in this report are very restrictive, making it unlikely that it will be useful for the analysis of real GAIA data. Therefore a discussion is included on why this approach is useful and what problems should be studied with it before turning to more complicated data analysis methods.

1 Introduction

All the data collected by the GAIA satellite will be sent down in the form of 'patches', which correspond to small regions of the sky centered on each detected star (point source). The patches containing the multi-colour photometric information will comprise a set of 'samples' (most of these sets being one-dimensional), which in turn consist of a number of CCD pixels binned together electronically before readout. Sometimes additional numerical binning is applied before transmission of the data to the ground. Details are in section 3.3 of the GAIA Study Report (ESA, 2000). Hence there will be no images to which classical photometric data analysis methods (for example, aperture or PSF photometry) can be applied. The task here is to develop algorithms and corresponding software that can work with the photometric data in the form of patches in order to produce the final photometry for the mission.

The real data analysis for the GAIA mission necessarily includes an integrated approach to both the astrometric and photometric aspects of the mission as these affect each other (see GAIA-LL-034; Lindegren, 2001). However for the purpose of this study we will concentrate solely on the photometric data analysis task. The aim is to Assess the quality of the final photometry at the end of mission. Therefore we will work on the following two assumptions; a) all stellar positions down to a certain limiting magnitude are known and b) the GAIA point spread function (PSF) and its dependence on stellar spectral type and position in the field of view are also known.

These assumptions together with the ones listed in the following section are very restrictive. I will further discuss the applicability of the algorithms worked out in this report in Sect. 5.

2 Describing the GAIA 'Image'

In the following a GAIA 'image' refers to the image from one transit of the GAIA field of view across a point on the sky. This image would be obtained if all pixels of the CCD detectors were to be read out and transmitted to the ground. Assuming the CCDs are entirely linear this image consists of a superposition of GAIA PSFs translated to the position of the corresponding stars in the field of view. It is this image that is converted into a set of patches before transmission of the data to the ground. The patches consist of samples which in turn are built up from a set of pixels binned together. This is all schematically illustrated in Fig. 1.

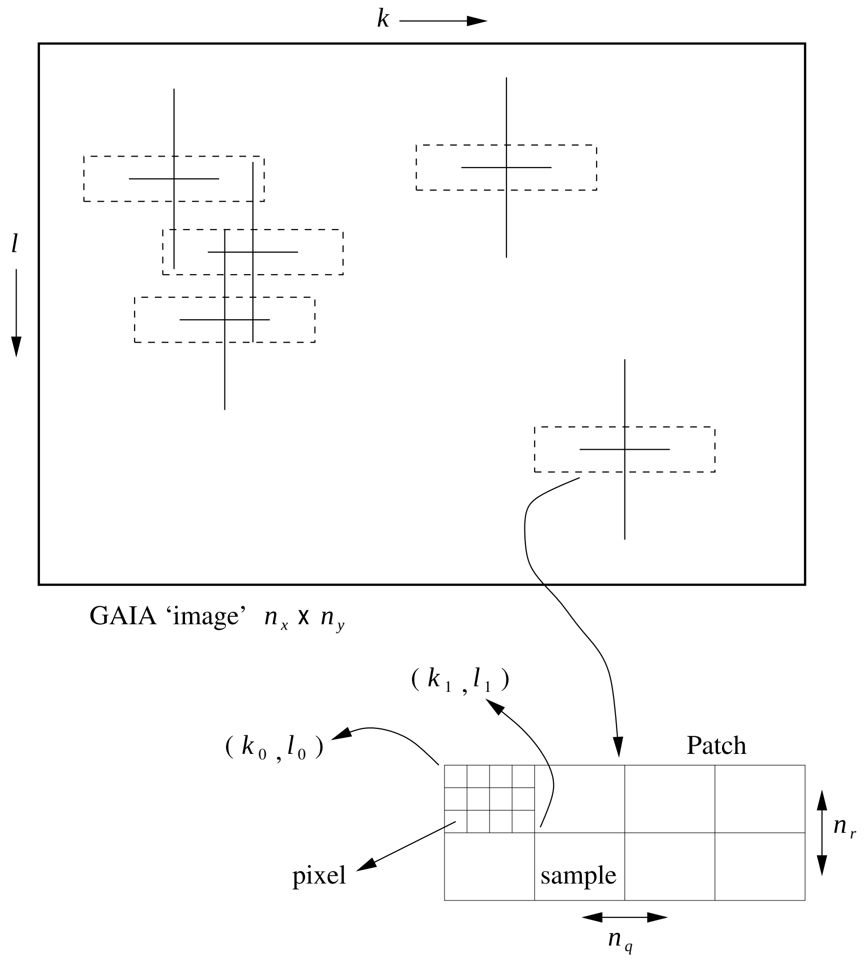

Figure 1: A schematic overview of the GAIA 'image' and how it is turned into patches. Due to the rectangular mirrors the PSF for GAIA consists of a core with four prominent spikes, indicated in the figure by the crosses. Around each stellar position a small region of the sky (a patch) is binned into a set of samples before transmission to the ground. The patch shown here is two-dimensional, consisting of $n_q \times n_r$ samples.

The image consists of $n_{\rm pix} = n_x \times n_y$ pixels labeled (k, l). The coordinates in the GAIA images are indicated by (x, y) with the pixel positions given by $(x(k), y(l))$. [1] For the images to be used in this report the following simplifying assumptions are made ('scan angle' refers to the position angle of a transit across a field on the sky):

1. The positions (x_i, y_i) at zero scan angle are known for all stars i to a certain magnitude limit.

2. The PSF is known and is the same for every star in the image.

3. The scan angle θ is known and is the same for each patch.

[1] Note that the labels (k, l) correspond to the (k, m) mesh of the electronic reference system defined in section 9.2.1 of the Study Report, and (x, y) corresponds to the continuous coordinates (κ, μ) in that system.

4. The sky background b (per pixel) is uniform and constant in time.
5. The CCD detectors are completely linear.
6. The stars are fixed on the sky.

The assumptions of the same PSF for every star and a constant background are not strictly necessary for this study and can be easily relaxed in a later stage. The number of stars in the image is N and the number of scan angles is n_θ. So there will be $N \times n_\theta$ patches of each type.

The assumptions listed above enable a simple description of the image as a superposition of images of the PSF translated to the position of each star and scaled according to the flux, or amplitude, of that star. The positions of the stars in the field of view will depend on the scan angle θ of the given image, hence:

$$(x_i, y_i) = (x_i(\theta), y_i(\theta)) \tag{1}$$

The PSF is indicated by $S(x, y)$ and the stellar amplitude (or flux) by a_i. Hence we can write the image $D_\theta(k, l)$ for a given scan angle θ as:

$$D_\theta(k, l) = b + \sum_{i=1}^{N} a_i S_{\theta,i}(k, l) \;, \tag{2}$$

where:

$$S_{\theta,i}(k, l) = S(x(k) - x_i(\theta), y(l) - y_i(\theta)) \tag{3}$$

The image $D_\theta(k, l)$ can also be indexed with one number $\lambda(k, l)$ and then viewed as a vector of pixels rather than a two dimensional image. This will facilitate writing down Eq. (2) in matrix form. The index λ is calculated as:

$$\lambda(k, l) = n_x k + l + 1 \;, \tag{4}$$

where $k \in [0, n_x - 1]$ and $l \in [0, n_y - 1]$. Equation (2) can thus be written as:

$$D_\theta(\lambda) = b + \sum_{i=1}^{N} a_i S_{\theta,i}(\lambda) \tag{5}$$

In matrix form this becomes:

$$\begin{pmatrix} D_\theta(1) \\ \vdots \\ D_\theta(n_{\text{pix}}) \end{pmatrix} = \begin{pmatrix} 1 & S_{\theta,1}(1) & \cdots & S_{\theta,N}(1) \\ \vdots & \vdots & & \vdots \\ 1 & S_{\theta,1}(n_{\text{pix}}) & \cdots & S_{\theta,N}(n_{\text{pix}}) \end{pmatrix} \begin{pmatrix} b \\ a_1 \\ \vdots \\ a_N \end{pmatrix} \tag{6}$$

The notation can be further simplified if we consider the column vectors $\mathbf{d}_\theta = D_\theta(\lambda)$ and $\mathbf{s}_{\theta,i} = S_{\theta,i}(\lambda)$, where $\lambda \in [1, n_{\text{pix}}]$:

$$\mathbf{d}_\theta = \begin{pmatrix} \mathbf{1} & \mathbf{s}_{\theta,1} & \cdots & \mathbf{s}_{\theta,N} \end{pmatrix} \begin{pmatrix} b \\ \mathbf{a} \end{pmatrix} \;, \tag{7}$$

where $\mathbf{1}$ indicates a column vector consisting of ones, and $\mathbf{a}$ indicates a_i with $i \in [1, N]$.

The matrix on the right hand side of this equation describes with what weight the flux of star i contributes to pixel λ in the GAIA image. Naming this matrix W_θ and writing the vector $(b \;\; \mathbf{a}')$ as $\mathbf{c}'$ (where $\mathbf{v}'$ indicates the transpose of $\mathbf{v}$) we obtain:

$$\mathbf{d}_\theta = \mathsf{W}_\theta \mathbf{c} \tag{8}$$

3 Describing the Patches

As shown in Fig. 1, for a star $i = j$ located at $(x_j(\theta), y_j(\theta))$ the patches consist of a set of $n_{\mathrm{samp}} = n_q \times n_r$ samples which are indexed by $\kappa \in [1, n_{\mathrm{samp}}]$. Each sample in turn consists of a set of pixels given by $k_0 \leq k \leq k_1 \wedge l_0 \leq l \leq l_1$. This set of pixels is a function of $(x_j(\theta), y_j(\theta))$ and κ and is indicated by $\mathcal{U}(j, \theta, \kappa)$. Note that this set of pixels takes into account both the on-chip binning of pixels into patches as well as the subsequent numerical addition of some samples before transmission to the ground (i.e., not every transmitted sample consists of the same amount of pixels, see Fig. 3.7 in the Study Report). The equations for each patch $P_{\theta,j}(\kappa)$ then follow from simply summing the appropriate subsets of Eq. (6):

$$P_{\theta,j}(\kappa) = \sum_{\lambda \in \mathcal{U}(j,\theta,\kappa)} D_\theta(\lambda)\,, \tag{9}$$

where the summation is over the interval of pixels indicated by $\mathcal{U}(j, \theta, \kappa)$. For ease of notation no distinction is made now between the different types of patches shown in Fig. 3.7 of the Study Report (such as BBP, SSM, MBP patches etc.).

Introducing $f(\kappa)$, which is the number of pixels contained in $\mathcal{U}(j, \theta, \kappa)$ (where f only depends on κ for a particular type of patch) and:

$$H_{\theta,i,j}(\kappa) = \sum_{\lambda \in \mathcal{U}(j,\theta,\kappa)} S_{\theta,i}(\lambda)\,, \tag{10}$$

we obtain:

$$\begin{pmatrix} P_{\theta,j}(1) \\ \vdots \\ P_{\theta,j}(n_{\mathrm{samp}}) \end{pmatrix} = \begin{pmatrix} f(1) & H_{\theta,1,j}(1) & \cdots & H_{\theta,N,j}(1) \\ \vdots & \vdots & & \vdots \\ f(n_{\mathrm{samp}}) & H_{\theta,1,j}(n_{\mathrm{samp}}) & \cdots & H_{\theta,N,j}(n_{\mathrm{samp}}) \end{pmatrix} \begin{pmatrix} b \\ a_1 \\ \vdots \\ a_N \end{pmatrix} \tag{11}$$

These equations can again be written in terms of column vectors $\mathbf{p}_{\theta,j} = P_{\theta,j}(\kappa)$ and $\mathbf{h}_{\theta,i,j} = H_{\theta,i,j}(\kappa)$, with $\kappa \in [1, n_{\mathrm{samp}}]$:

$$\mathbf{p}_{\theta,j} = \begin{pmatrix} \mathbf{f} & \mathbf{h}_{\theta,1,j} & \cdots & \mathbf{h}_{\theta,N,j} \end{pmatrix} \begin{pmatrix} b \\ \mathbf{a} \end{pmatrix}\,, \tag{12}$$

where $\mathbf{f}$ indicates a column vector with elements $f(\kappa)$. The matrix on the right hand side in Eq. (11) now indicates the weight with which each star i contributes to the patch of star j at scan angle θ. Hence, writing the matrix on the right hand side as $\mathsf{M}_{\theta,j}$ we obtain:

$$\mathbf{p}_{\theta,j} = \mathsf{M}_{\theta,j}\,\mathbf{c}\,, \tag{13}$$

and as a last step Eq. (13) can be generalised to include patches of different types t:

$$\mathbf{p}_{\theta,j,t} = \mathsf{M}_{\theta,j,t}\,\mathbf{c} \tag{14}$$

In order to solve for the unknowns $\mathbf{c}$ in practice it will not be enough to consider only one patch at a time. The equations for $\mathbf{p}_{\theta,j,t}$ can be gathered together for different values of θ, j, and t into a single set of equations:

$$\mathbf{p} = \mathsf{M}\mathbf{c}\,, \tag{15}$$

where $\mathbf{p}' = (\mathbf{p}'_{\tau_1} \cdots \mathbf{p}'_{\tau_{n_{\mathrm{patch}}}})$ and $\mathsf{M}' = (\mathsf{M}'_{\tau_1} \cdots \mathsf{M}'_{\tau_{n_{\mathrm{patch}}}})$. The variable τ stands for the set (θ, j, t) and n_{patch} is the total number of patches considered.

4 Implementation and Testing

So far the scheme outlined in the previous two sections has been implemented as a set of IDL procedures that work on simulated data (also generated using IDL). This data can be generated for any field on the sky and includes the GAIA PSF as calculated as described in the report SAG-LL-025 (Lindegren, 1998a), as well as the GAIA scanning law as outlined in SAG-LL-026 (Lindegren, 1998b). The patches are generated in accordance with the description in the GAIA Study Report (ESA, 2000). The IDL procedures will be made available and are described in a forthcoming report.

The equations are solved using singular value decomposition, taking into account the errors on the data in the patches. Some very preliminary tests have been done in order to check that the algorithms were correctly implemented but no results on the final photometric accuracies have been obtained yet.

To investigate the performance of this data analysis scheme the following parameters should be varied:

- The number density of stars on the sky
- The magnitude limit out to which stellar positions are known (beyond which fainter stars only contribute to the background)
- The form of the luminosity function (i.e., relative numbers of faint and bright stars)

5 Applicability of the Linear Least Squares Approach

The formulation of the data reduction problem in terms of a linear least squares scheme as outlined in this report relies on very restrictive assumptions. Many of these can be relaxed. For example, the method can be extended to incorporate a changing PSF as a function of position in the field of view and time, and to take into account the parallactic and proper motions of real stars. A varying sky-background can also be incorporated into this approach. However the CCDs themselves will show non-linear behaviour which cannot be easily included in this model of data analysis. Therefore it is unlikely that the data analysis methods presented here will ultimately be useful to tackle the complexities of the real GAIA data.

However, in relation to directly taking, e.g., an iterative approach to the photometric data analysis (where one can take non-linearities into account), the least squares approach has the advantage that it is easy to implement and provides a means of quickly gaining insight into a number of important issues concerning the photometric data analysis.

There are two important questions that can be addressed using the linear least squares approach, the answers to which will provide guidance to the efforts toward constructing the actual data analysis algorithms:

- The algorithms presented in this report rely on information about the positions of stars fainter that the GAIA survey limit. Out to which distance on the sky around a particular star is positional information on other sources necessary, and to what magnitude limit? How accurately do we need to know these positions and what is the effect of the parallactic and proper motion of the background stars? This will help determine how the sky should be partitioned for the overall data analysis task. Part of this information can be obtained by studying the covariance matrix $(\mathbf{M}'\mathbf{M})^{-1}$ associated with Eq. (15).
- Assuming the full astrometric catalogue for the GAIA targets is available as well as knowledge on the background sources; what photometric accuracies can ultimately be achieved as a function of the parameters mentioned in Sect. 4?

A closely related issue that needs to be studied is how the positions and magnitudes of the background stars can be derived from the patches observed in the wide G-band in ASM3, AF17 and SSM. This involves the reconstruction of an image of the piece of sky around each target star covered by these patches. This is not trivial in the case of purely one-dimensional patches. The answers combined with those to the questions above will provide more constraints on the optimal binning scheme for the CCD data.

References

ESA, 2000, GAIA: Composition, Formation and Evolution of the Galaxy – Concept and Technology Study Report, ESA-SCI(2000)4

Lindegren L., 1998a, Point Spread Functions for GAIA Including Aberrations, Report SAG-LL-025

Lindegren L., 1998b, Simulation of GAIA Scanning of Arbitrary Directions, Report SAG-LL-026

Lindegren L., 2001, Proposed Prototype Processes for the GAIA Global Iterative Solution, Report GAIA-LL-034