Optimizing Gaia’s Photometric System:
Thoughts on distance measure and figure of merit

L. Lindegren


Abstract. This note summarizes my considerations on certain aspects of the
design and testing of Gaia’s photometric system: (1) how to measure the dis-
tance between the set of fluxes resulting from two different sets of astrophysical
parameters observed in the same photometric system (a problem relevant for
the blind testing); (2) how to define a figure of merit for an arbitrary pho-
tometric system (relevant for the design of photometric systems); (3) how to
incorporate astrometric information (parallax) into the optimization; (4) as a
consequence of the above, the need to define quantitative goals for the precision
of astrophysical parameters on each scientific target.

1 Notations and general assumptions

\( ST \) = (photometric) scientific targets
\( PS \) = photometric system (MBP or BBP, or both)
\( AP \) = astrophysical parameters (e.g., \( A_V \), \( \log T_{\text{eff}} \), \( \log g \), [Fe/H])
\( i \) = index for the different ST
\( j \) = index for the different filters in a given PS
\( k \) = index for the different AP for which a PS is optimized
\( p_k \) = AP \( k \) (defining the vector \( p \))
\( p_i \) = vector of AP for ST \( i \)
\( \sigma_{k, \text{prior}} \) = prior (if no data) standard error of AP \( k \)
\( \sigma_{ik, \text{goal}} \) = goal (desired) standard error of AP \( k \) for ST \( i \)
\( \sigma_{ik, \text{post}} \) = posterior (achieved) standard error of AP \( k \) for ST \( i \)
\( C_x \) = variance–covariance matrix of the general vector \( x \)
\( \phi_{ij} \) = normalized flux of ST \( i \) in filter \( j \)
\( \epsilon_{ij} \) = standard error on the normalized flux of ST \( i \) in filter \( j \)
\( M(p) \) = \( M_G + A_G \) (absolute magnitude plus extinction) as function of AP
\( \pi_i \) = parallax of ST \( i \)
\( \sigma_{\pi i} \) = standard error in astrometric parallax for ST \( i \)
\( w_i \) = relative weight of ST \( i \)
\( Q_i \) = figure of merit of PS for ST \( i \)
\( Q \) = global figure of merit of PS
\( f(x) \) = a (non-linear) function used in the definition of figure of merit
Remarks:

1. The range of indices is usually implied by the context, e.g., $\sum_i$ means a sum over all relevant ST.

2. Since indices $i$, $j$ and $k$ are always associated with ST, filter and AP, respectively, we can with some abuse of notation write, for instance, $\phi_j(p)$ for the normalized flux in filter $j$ produced by the general AP vector $p$. The vector of normalized fluxes may be denoted $\phi(p)$.

3. The ST are here considered simply as a list of hypothetical stars, each characterized by a given set of AP ($p_i$), parallax ($\pi_i$), etc.

4. The a priori standard errors $\sigma_{k,\text{prior}}$ are large numbers reflecting the initial uncertainty of the AP, of order 2 mag for $A_V$, 0.5 dex for $\log T_{\text{eff}}$, etc.

5. The desired standard errors $\sigma_{ik,\text{goal}}$ are the targeted precisions discussed in Sect. 5.

6. Since in any PS it is only the relative fluxes among the different filters that are relevant for the determination of the AP, it is necessary to somehow normalize the fluxes $\phi_{ij}$ when considering how they are affected by a change in the AP. E.g., a change in $\log g$ ($\sim$ luminosity) may change the flux in all filters by a significant factor, but this is not useful for the determination of $\log g$ unless the factor is different at least for some filters. The normalization could for instance be made in such a way that the flux in the $G$ band is independent of the AP. (The outcome of the optimization should in principle be independent of how this normalization is achieved.)

7. For the mathematical formulation it is assumed that (normalized) fluxes $\phi_{ij}$ can be computed for any vector $p$ of the AP (within the relevant ranges, of course), and that the fluxes are continuous and differentiable functions of $p$. The sensitivity of the normalized fluxes to changes in the AP is described, for each ST, by the sensitivity matrix

$$
S_i = \begin{bmatrix} 
\frac{\partial \phi_{i1}}{\partial p_1} & \cdots & \frac{\partial \phi_{i1}}{\partial p_K} \\
\vdots & \ddots & \vdots \\
\frac{\partial \phi_{iJ}}{\partial p_1} & \cdots & \frac{\partial \phi_{iJ}}{\partial p_K} 
\end{bmatrix} \quad p = p_i
$$

of size $J \times K$, where $J$ is the number of filters considered and $K$ the number of AP. This matrix must in practice be computed by numerical differentiation. Note that this matrix can also be considered a function of the AP: $S(p)$.

8. The present formulation ignores effects of the (in reality very strong) non-linearity of $\phi(p)$, which may lead to uniqueness problems (i.e., two different $p$ mapped to the same $\phi$, even if the mapping is non-degenerate in the vicinity of each parameter point).

---

1. The only context where the non-normalized flux is relevant is for the inclusion of parallax information, which is discussed in Sect. 4.
9. For the inclusion of the parallax information it is necessary to assume that the luminosity or absolute magnitude is a known function of the AP. Since the extinction is included among the AP, we take \( M(p) \) to designate the absolute magnitude plus extinction in the relevant band, e.g., \( M_G + A_G \). According to the distance formula this should equal \( G + 5 \log[\pi/(0.1 \text{ arcsec})] \), where \( \pi \) is the measured parallax.

2 Measuring the flux distance

In parametrization methods based on the nearest neighbour, the AP corresponding to the observed fluxes \( \phi \) are estimated by minimizing the distance in flux space, \( D \equiv |\phi_{\text{obs}} - \phi(p)| \), with respect to \( p \). An important consideration is then how to define the metric in flux space, i.e., how to weight the components of \( \Delta \phi = \phi_{\text{obs}} - \phi(p) \) for the different filters \((j)\). Using a quadratic distance measure, the metric is given by some non-negative definite symmetric \( J \times J \) matrix \( R \) such that \( D^2 = \Delta \phi^T R \Delta \phi \) (where \( \Delta \phi \) is regarded as a column matrix).

An obvious choice of matrix might be \( R = C_\phi^{-1} \), where \( C_\phi = \text{diag}(\epsilon_j^{-2}) \) is the covariance matrix of \( \phi_{\text{obs}} \) (diagonal if the errors on the different filters are assumed uncorrelated). However, this does not take into account that the flux distances along different axes \((j)\) are not equally relevant for the AP determination, and in particular that the fluxes most relevant for a particular parameter (such as metallicity) may depend mainly on the flux in a particular filter, where the flux variance may be relatively large so that the filter contributes little weight to the distance measure. Thus, the flux differences should (also) be weighted according to their relevance for the AP, expressed by their desired covariance matrix \( C_{p,\text{goal}} = \text{diag}(\sigma_{ik,\text{goal}}^{-2}) \) (it is assumed that this matrix is diagonal).

The ‘relevance’ of each flux for the different AP is given by the sensitivity matrix. Thus, a tentative definition of metric is

\[
R = (C_\phi + SC_{p,\text{goal}}S^T)^{-1}
\]

With this metric, a large difference \( \Delta \phi_j \) in one of the filters will be down-weighted (by means of a small value of \( R_{jj} \)) either if the measured flux is very uncertain (large \( \epsilon_j \), expressed through the matrix \( C_\phi \)) or if the flux is irrelevant for the desired precision of AP (expressed through \( C_{p,\text{goal}} \) via the sensitivity matrix \( S \)).

3 Defining a figure of merit

While the flux distance \( D \) defined in Sect. 2 is relevant for finding the AP which most closely reproduces the observed fluxes, it is not directly useful for designing a PS. The reason is that for any PS and any fluxes (consistent with the adopted models) it will always be possible to find AP that reproduce the observed fluxes within the flux errors.

\(^2\)A problem here is that \( C_{p,\text{goal}} \) may depend on which category of target is considered, which is not known until the parametrization has been made; however, we regard this as a second-order problem similar to the problem of the non-linearity of \( \phi(p) \) mentioned under item 8 of Sect. 1.
The figure of merit for a PS must instead be based on how accurately the AP can be determined, irrespective of the actual method used to estimate the AP. Thus we must be able to compute \( \sigma_{ik,\text{post}} \), the standard error of the estimated AP \( k \) for ST \( i \), in an arbitrary PS.

Consider the AP determination as a linearized least-squares estimation of \( \Delta p \) (the improvement to the AP vector), where the observation equation resulting from the flux measured in filter \( j \) reads:

\[
\frac{\partial \phi_j}{\partial p_1} \Delta p_1 + \cdots + \frac{\partial \phi_j}{\partial p_K} \Delta p_K = \Delta \phi_j \pm \epsilon_j
\]  

(3)

Here, \( \Delta \phi_j = \phi_{j,\text{obs}} - \phi_j(p) \) is the \( O - C \) in flux and \( \pm \) indicates the flux uncertainty. Observation equations of unit weight are formed through division by \( \epsilon_j \), whereupon normal equations are formed in the usual manner. It is assumed that the linearization is made around the true parameter vector \( p \), so that the resulting update \( \Delta p \) has zero expectation.

The covariance of the estimated vector, \( p_{\text{post}} \), is then given by the inverse of the normal equations matrix,

\[
C_{p,\text{post}} = \left(S^T C_\phi^{-1} S\right)^{-1}
\]

(4)

and a figure of merit can be computed from the diagonal elements \( \sigma_{ik,\text{post}}^2 \equiv [C_{p,\text{post}}]_{kk} \).

In reality degeneracy among the AP will often make the matrix \( S^T C_\phi^{-1} S \) singular or near-singular, resulting in infinite or very large \( \sigma_{ik,\text{post}} \) as computed from (4). To avoid this we use a standard trick of regularization: adding a suitable positive definite matrix \( B \) makes the whole right-hand side positive definite; thus:

\[
C_{p,\text{post}} = \left(B + S^T C_\phi^{-1} S\right)^{-1}
\]

(5)

\( B \) is the \( a \ priori \) information matrix of the AP.\(^3\) If we take \( a \ priori \) to mean just in the absence of any other data, we have \( B = \text{diag}(\sigma_{k,\text{prior}}^2) \). However, as will be shown in Sect. 4, the information matrix \( B \) can be modified to include the additional information provided by the parallax.

If the PS does not provide any relevant information on the AP (either because the flux variances in \( C_\phi \) are too large or because the elements of the sensitivity matrix \( S \) are too small), then \( C_{p,\text{post}} \sim B^{-1} \), and \( \sigma_{k,\text{post}} \sim \sigma_{k,\text{prior}} \). If the PS manages to determine some AP, but not all of them, then \( \sigma_{k,\text{post}} \) is reduced for the corresponding \( k \).

The success of a PS to achieve the specified astrophysical goals should be measured with respect to the target precisions \( \sigma_{k,\text{goal}} \). This means that the figure of merit must be a function of the ratio \( x \equiv \sigma_{k,\text{post}}/\sigma_{k,\text{goal}} \). If \( x \leq 1 \), then the goal with respect to that AP is completely satisfied. To avoid over-determination of some AP at the expense of others,\(^3\)

\(^3\)Loosely speaking, the information matrix is the inverse of the corresponding covariance matrix. However, the information matrix is more general in the sense that it exists also for degenerate problems, in which case it is non-positive definite, so that the covariance cannot be computed. In a properly normalized least-squares formulation, the information matrix equals the normal equations matrix. The (Fisher) information is generally defined as the negative curvature of the log-likelihood evaluated at the true parameter value.
the figure of merit should not increase (much) if $x$ is less than 1. Thus, the figure of merit should be a non-linear function of $x$ with a break around $x = 1$. Possible simple functions could be

$$f(x) = (1 + x^{2n})^{-1/n}$$

(6)

for $n = 1, 2, \ldots$. They all have the properties that $f \simeq 1$ for $x < 1$ and $f \simeq x^{-2}$ for $x > 1$. (It can be discussed whether the latter property is optimal.)

The figure of merit for a given ST is now computed as

$$Q_i = \sum_k f(\sigma_{ik,\text{post}}/\sigma_{ik,\text{goal}})$$

(7)

and the global figure of merit of the PS, finally, is the weighted sum

$$Q = \sum_i w_i Q_i$$

(8)

4 Including the parallax

Given an astrometric parallax $\pi$ with standard error $\sigma_{\pi}$, we now show how this information can be included in the a priori information matrix $B$. We use that information matrices are additive, and represented by the corresponding normal equations matrices of the normalized (unit weight) observation equations. The computed parallax for given AP is

$$\pi(p) = (0.1 \text{ arcsec}) \times 10^{0.2[M(p)-G]}$$

(9)

where $M(p)$ the absolute magnitude plus extinction and $G$ the observed apparent magnitude. (The choice of the $G$ band for this computation is of course arbitrary.) The linearized observation equation is

$$(0.2 \ln 10)\pi(p) \frac{\partial M}{\partial p} \Delta p = \pi_{\text{obs}} - \pi(p) \pm \sigma_{\pi}$$

(10)

where we have neglected the uncertainty in the measured $G$ compared with that in the astrometric parallax. With

$$b_k = (0.2 \ln 10) \frac{\pi(p)}{\sigma_{\pi}} \frac{\partial M}{\partial p_k}$$

(11)

we find that $b_kb_{k'}$ should be added to the element $B_{kk'}$ in the previously defined (diagonal) information matrix $B$. It is easy to see how the inclusion of parallax information works in the above equations: if $M(p)$ depends significantly on the particular AP $k$ (e.g., log $g$), and if $\pi/\sigma_{\pi}$ is large, then $b_k$ becomes numerically large, and a large positive number is added to the diagonal element $B_{kk}$. In equation (5), this causes a corresponding reduction in $\sigma_{k,\text{post}}$. But it may also lift a possible degeneracy among the AP reflected in a near-singularity in the second term of the equation. Therefore, the added parallax information reduces not only the uncertainty of that particular AP $k$, but in general all the AP are improved.
5 The need to define target precisions

It should be clear from the preceding discussion that quantification of the target precisions $\sigma_{ik, \text{goal}}$ of the AP is necessary both in order to define a physically meaningful flux distance [via the metric in (2)] and to compute the figure of merit for a given PS [via (7)]. This must be done even if the same target precisions are assigned to all ST, since it also defines the relative weight assigned to the different AP (e.g., log $g$ versus [Fe/H]).