A posteriori reduction of the set of basis functions for LSF modelling

prepared by: L. Lindegren
affiliation : Lund Observatory
approved by:
reference: GAIA-C3-TN-LU-LL-090-01
issue: 1
revision: 0
date: 2010-10-31
status: Issued

Abstract

This note describes how to reduce the size of the set of basis functions needed for LSF modelling, once a comprehensive library of real LSFs has been compiled. The method uses Principal Component Analysis (PCA) of the library coefficients.

Document History

Issue	Revision	Date	Author	Comment
D	0	2010-10-07	LL	Creation and 1st draft
1	0	2010-10-31	LL	Minor corrections

1 The generic LSF model

The generic LSF model described in Lindegren (2010) has the following form:

$$L(u) = H_0(u - h_0) + \sum_{n=1}^{N-1} h_n H_n(u - h_0)\,, \tag{1}$$

where h_n ($n = 0 \ldots N-1$) are model parameters and $H_n(x)$ are basis functions satisfying

$$\int_{-\infty}^{+\infty} H_n(x)\,\mathrm{d}x = \delta_{0n}\,. \tag{2}$$

The property (2) guarantees that $L(u)$ is properly normalized for any choice of model parameters.

The current implementation of the generic model is described in Lindegren (LL-088) and is based on the basis functions derived in Lindegren (LL-084) by PCA analysis of a large set of physically possible LSFs, assuming arbitrary wavefront errors (within reasonable bounds) and spectral energy distributions.

In the real Gaia, only a small subset of these wavefront errors will actually occur, and the range of stellar spectra may also be more restrictive.[1] As a consequence, the LSF library applicable to the real Gaia will only use a (small) subset of all possible parameter combinations in (1). Although any particular LSF may require a large number of parameters ($N \sim 40$) in order to be accurately represented in the current (Lindegren, LL-088) implementation of the generic model, the subspace spanned by the LSF library may effectively have a much smaller dimensionality (say $\tilde{N} \sim 10$). It is then possible to redefine the basis functions is such a way that only $\tilde{N}$ free parameters are needed to represent the library to sufficient accuracy. The reduced set of basis functions can be found by PCA of the coefficients in the LSF library.

2 Transformation of the basis functions and parameters

Let us assume that we have defined a transformed set $\{\tilde{H}_n(x)\}_{n=0}^{N-1}$ of basis functions also satisfying

$$\int_{-\infty}^{+\infty} \tilde{H}_n(x)\,\mathrm{d}x = \delta_{0n}\,, \tag{3}$$

[1] The real range of spectra will be more restrictive in the sense that the vast majority of stars fall within a relatively narrow span of SEDs, compared to what was simulated in Lindegren (LL-084). On the other hand, the real data will undoubtedly contain many extreme SEDs – including almost monochromatic emission-line spectra – that surely fall outside the simulated range. It may be necessary to accept some degradation of the LSF calibration for such objects.

such that any LSF that can be represented by the original set $\{H_n(x)\}_{n=0}^{N-1}$ can be exactly represented also by the transformed set. That is, for any parameter vector $\boldsymbol{h} = [h_1, h_2, \ldots h_{N-1}]'$ there exists a transformed vector $\tilde{\boldsymbol{h}} = [\tilde{h}_1, \tilde{h}_2, \ldots \tilde{h}_{N-1}]'$ such that

$$H_0(x) + \sum_{n=1}^{N-1} h_n H_n(x) = \tilde{H}_0(x) + \sum_{n=1}^{N-1} \tilde{h}_n \tilde{H}_n(x)\,. \tag{4}$$

Moreover, we want the transformation to be such that if the representation is truncated after $\tilde{N} < N$ terms, we make a minimal error in the LSF approximation. Because of the very different nature of the zeroth-order terms (H_0 and $\tilde{H}_0$) with respect to the rest of the representation, it is necessary to discuss them separately. The transformation of H_0 is discussed below, while that of H_n ($n > 0$) is deferred to the next section.

The optimization of the transformation will be made on the given LSF library, e.g., by minimizing the RMS error as calculated over the library. (It is in principle possible to assign weights to the LSFs, but for the present analysis we assume equal weights.) To minimize the expected variance of the sum over terms with $n > 0$, we first of all want to make $\tilde{H}_0(x)$ equal to the mean LSF in the library. Let $\langle\,\rangle$ denote averaging over all the LSFs in the library. Then

$$\tilde{H}_0(x) = H_0(x) + \sum_{n=1}^{N-1} \langle h_n \rangle H_n(x)\,, \tag{5}$$

which gives the required transformation of H_0. Subtracting this from (4) we see that

$$\sum_{n=1}^{N-1} \left(h_n - \langle h_n \rangle\right) H_n(x) = \sum_{n=1}^{N-1} \tilde{h}_n \tilde{H}_n(x) \tag{6}$$

must now be satisfied for arbitrary $\boldsymbol{h}$. We write this in matrix form as

$$\boldsymbol{e}'\boldsymbol{H} = \tilde{\boldsymbol{h}}'\tilde{\boldsymbol{H}} \tag{7}$$

where $\boldsymbol{e} = \boldsymbol{h} - \langle \boldsymbol{h} \rangle$ and

$$\boldsymbol{H} = \begin{bmatrix} H_1(x) \\ H_2(x) \\ \vdots \\ H_{N-1}(x) \end{bmatrix}, \qquad \tilde{\boldsymbol{H}} = \begin{bmatrix} \tilde{H}_1(x) \\ \tilde{H}_2(x) \\ \vdots \\ \tilde{H}_{N-1}(x) \end{bmatrix}. \tag{8}$$

It is clear that each new basis function $\tilde{H}_n(x)$ ($n > 0$) must be a linear combination of the old functions, with a non-singular coefficient matrix $\boldsymbol{F}$ of dimension $(N-1) \times (N-1)$; thus,

$$\tilde{\boldsymbol{H}} = \boldsymbol{F}\boldsymbol{H}\,. \tag{9}$$

If now the parameters are correspondingly transformed,

$$\tilde{\boldsymbol{h}} = (\boldsymbol{F}')^{-1}\boldsymbol{e}\,, \tag{10}$$

then (7) obviously holds for arbitrary parameters.

Equations (5), (9) and (10) thus provide a consistent way of transforming the basis functions and parameters, such that (3) is satisfied. The shift parameter clearly needs no transformation ($\tilde{h}_0 = h_0$).

3 Choice of transformation

We now want to choose $\boldsymbol{F}$ such that, for any $\tilde{N} < N$, the truncated expansion

$$L(x) \simeq \tilde{H}_0(x) + \sum_{n=1}^{\tilde{N}-1} \tilde{h}_n \tilde{H}_n(x) \tag{11}$$

gives the smallest possible approximation error. (For simplicity the shift parameter h_0 is assumed to be zero from here on.) The approximation error is

$$\Delta L(x) = \sum_{n=\tilde{N}}^{N-1} \tilde{h}_n \tilde{H}_n(x) = \tilde{\boldsymbol{h}}' \boldsymbol{T} \tilde{\boldsymbol{H}} \tag{12}$$

where $\boldsymbol{T} = \mathrm{diag}(0, 0, \ldots, 0, 1, \ldots, 1)$ is a diagonal matrix with zeroes in the first $\tilde{N}-1$ positions along the diagonal, and ones in the remaining $N - \tilde{N}$ positions.

Without loss of generality, we can write the coefficient matrix in (9) as

$$\boldsymbol{F} = \boldsymbol{U}'(\boldsymbol{R}')^{-1}\,, \tag{13}$$

where $\boldsymbol{U}$ is a unitary (orthonormal) matrix and $\boldsymbol{R}$ an upper-triangular positive-definite matrix.[2] The inverse transposed coefficient matrix that appears in (10) is

$$(\boldsymbol{F}')^{-1} = \boldsymbol{R}'\boldsymbol{U}\,. \tag{14}$$

So the question is how to compute a unitary matrix $\boldsymbol{U}$ and a triangular matrix $\boldsymbol{R}$ to achieve our goal.

Let $\boldsymbol{A}$ be the symmetric, positive-definite $(N-1) \times (N-1)$ matrix

$$\boldsymbol{A} = \int_{-\infty}^{+\infty} \boldsymbol{H}\boldsymbol{H}'\,\mathrm{d}x\,. \tag{15}$$

Using the Cholesky algorithm this can be uniquely factorized as

$$\boldsymbol{A} = \boldsymbol{R}'\boldsymbol{R}\,, \tag{16}$$

[2] The existence of such matrices follows from the QR factorization theorem. The reason for the peculiar choice of notation for these matrices (rather than, say, $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{R}$) will become apparent later.

where $\boldsymbol{R}$ is upper-triangular; we make this our choice of $\boldsymbol{R}$. Furthermore, let

$$\boldsymbol{C} = \boldsymbol{R}\langle \boldsymbol{e}\boldsymbol{e}' \rangle \boldsymbol{R}' \tag{17}$$

be the covariance of $\boldsymbol{R}\boldsymbol{e}$ calculated over the LSF library. Using the SVD, this can be factorized as

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}' \,, \tag{18}$$

where $\boldsymbol{U}$ is a unitary matrix and $\boldsymbol{D}$ is a diagonal matrix containing the singular values in non-increasing sequence.

The square of the approximation error is

$$|\Delta L(x)|^2 = \tilde{\boldsymbol{h}}'\boldsymbol{T}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}'\boldsymbol{T}'\tilde{\boldsymbol{h}} = \tilde{\boldsymbol{h}}'\boldsymbol{T}\boldsymbol{U}'(\boldsymbol{R}')^{-1}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{U}\boldsymbol{T}'\tilde{\boldsymbol{h}} \,. \tag{19}$$

Integrating over x and using (15), (16) and the orthonormality of $\boldsymbol{U}$, we find the total squared error

$$\int_{-\infty}^{+\infty} |\Delta L(x)|^2 \, \mathrm{d}x = \tilde{\boldsymbol{h}}'\boldsymbol{T}\tilde{\boldsymbol{h}} = \sum_{n=\tilde{N}}^{N-1} \tilde{h}_n^2 \,. \tag{20}$$

On the other hand, since $\tilde{\boldsymbol{h}} = \boldsymbol{U}'\boldsymbol{R}\boldsymbol{e}$, we have

$$\langle \tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}' \rangle = \boldsymbol{U}'\boldsymbol{R}\langle \boldsymbol{e}\boldsymbol{e}' \rangle \boldsymbol{R}'\boldsymbol{U} = \boldsymbol{D} \,, \tag{21}$$

from which

$$D_{nn} = \langle \tilde{h}_n^2 \rangle \,. \tag{22}$$

The expected value of the total squared error is, therefore,

$$\left\langle \int_{-\infty}^{+\infty} |\Delta L(x)|^2 \, \mathrm{d}x \right\rangle = \sum_{n=\tilde{N}}^{N-1} D_{nn} \,, \tag{23}$$

that is the sum of last $N - \tilde{N}$ singular values. Since the SVD is such that the singular values in $\boldsymbol{D}$ are arranged in non-increasing sequence ($D_{11} \geq D_{22} \geq \cdots$), it is seen that the total squared error is indeed minimized by leaving out the $N - \tilde{N}$ terms with the highest n, as was our objective.

4 Summary of procedure

Given the LSF library in the form of one coefficient vector $\boldsymbol{h}$ for each LSF (excluding the 0th element h_0), and a set of basis functions H_0, $\boldsymbol{H}$, the new coefficients and basis functions are constructed by the following steps:

1. Calculate $\boldsymbol{A} = \int_{-\infty}^{+\infty} \boldsymbol{H}\boldsymbol{H}' \, \mathrm{d}x$

2. Calculate the upper-triangular factor $\boldsymbol{R}$ by means of a Cholesky factorization of $\boldsymbol{A}$
3. Calculate the mean parameter vector $\langle \boldsymbol{h} \rangle$
4. Calculate $\tilde{H}_0 = H_0 + \langle \boldsymbol{h} \rangle' \boldsymbol{H}$
5. Calculate $\boldsymbol{e} = \boldsymbol{h} - \langle \boldsymbol{h} \rangle$
6. Calculate $\boldsymbol{C} = \boldsymbol{R} \langle \boldsymbol{e}\boldsymbol{e}' \rangle \boldsymbol{R}'$
7. Make the SVD $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$ to obtain $\boldsymbol{U}$ and $\boldsymbol{D}$
8. Calculate $\boldsymbol{F} = \boldsymbol{U}'(\boldsymbol{R}')^{-1}$
9. Calculate $\tilde{\boldsymbol{H}} = \boldsymbol{F}\boldsymbol{H}$
10. Calculate $\tilde{\boldsymbol{h}} = (\boldsymbol{F}')^{-1}\boldsymbol{e}$

The singular values D_{nn} should also be saved, as they contain information about the expected approximation error as function of the number of basis functions.

A few remarks concerning the practical calculations:

(a) In Step 1, the integral can be limited to a finite interval (say, $-10 \leq x \leq 10$). The total squared error then refers to this interval as well.

(b) The elements of $\boldsymbol{A}$ are $A_{mn} = \int H_m(x) H_n(x)\, \mathrm{d}x$. Since the basis functions are nearly orthogonal, $\boldsymbol{A}$ is expected to be nearly diagonal. Due to the normalization of $H_n(x)$ introduced in Lindegren (LL-088), the diagonal elements can however be quite different.

(c) The expressions in Step 4 and 9 involve linear operations on the set of basis functions. Since each basis function in turn is defined as a linear combination of more elementary functions (biquartic spline, left and right tails), one must in practice perform the corresponding operations on the coefficients representing each basis function. For example, in Step 9 the right-tail coefficients for $\tilde{\boldsymbol{H}}$ are obtained by a linear combination of the right-tail coefficients for $\boldsymbol{H}$, using the coefficient matrix $\boldsymbol{F}$.

5 References

[LL-084], Lindegren, L., 2009, *Minimum-dimension LSF modelling*,
GAIA-C3-TN-LU-LL-084,
URL http://www.rssd.esa.int/llink/livelink/open/2915742

[LL-088], Lindegren, L., 2010, *LSF modelling with zero to many parameters*,
GAIA-C3-TN-LU-LL-088,
URL http://www.rssd.esa.int/llink/livelink/open/3042524

Lindegren, L., 2010, *A generic LSF/PSF model and its role in IDT, IDU and AGIS*, Tech. rep.,
Lund Observatory,
GAIA-C3-TN-LU-LL-089-01