DRAFT VERSION JULY 8, 2023
Typeset using LaTeX **modern** style in AASTeX631

Notes on Lightcurve Marginalization

WILL M. FARR[1,2]

[1]Department of Physics and Astronomy, Stony Brook University, Stony Brook NY 11794, USA
[2]Center for Computational Astrophysics, Flatiron Institute, New York NY 10010, USA

ABSTRACT

I review the marginal likelihood for a STARRY/STARRY-process model and how to draw max-likelihood surface map coefficients or sample from the map posterior.

1. FORMALISM

Starry (Luger et al. 2019) and starry-process (Luger et al. 2021).
Definitions:

- Let d be the data vector of observed fluxes of dimension N_d.
- Let Σ_d be the $N_d \times N_d$ covariance matrix for the flux measurements. (Usually Σ_d will be diagonal, with the squared measurement uncertainties for each data point running down the diagonal.)
- Let y be the vector of spherical harmonic coefficients describing the stellar surface map, of dimension N_y.
- Let A be the $N_d \times N_y$ design matrix produced by the STARRY algorithm relating the map spherical harmonic coefficients to the predicted lightcurve.
- Let μ be the dimension N_y mean vector of map spherical harmonics described by the starry process.
- Let Σ_s be the dimension $N_y \times N_y$ covariance matrix for the map spherical harmonics described by the starry process.

The terms in the joint log-likelihood for the data and the starry process that involve the spherical harmonic coefficients are

$$\log L = -\frac{1}{2} \left[(d - Ay)^T \Sigma_d^{-1} (d - Ay) + (\mu - y)^T \Sigma_s^{-1} (\mu - y) \right]. \tag{1}$$

wfarr@flatironinstitute.org

It should be clear that this log-likelihood is quadratic in y; we can "complete the square" and re-write it in the form

$$\log L = \log L_0 - \frac{1}{2}\left(\nu - y\right)^T \Sigma^{-1} \left(\nu - y\right), \tag{2}$$

where Σ is given by

$$\Sigma^{-1} = A^T \Sigma_d^{-1} A + \Sigma_s^{-1}, \tag{3}$$

ν is the maximum likelihood map and is given by

$$\nu = \Sigma\left(A^T \Sigma_d^{-1} d + \Sigma_s^{-1} \mu\right), \tag{4}$$

and

$$\log L_0 = \left.\log L\right|_{y=\nu}. \tag{5}$$

If we wish to marginalize out the map coefficients y, we find

$$\log \bar{L} \equiv \log \int \mathrm{d}y\, e^{\log L} = \log L_0 + \frac{1}{2} \log\det\left(2\pi\Sigma\right). \tag{6}$$

This marginal likelihood can be used to sample over the parameters of the starry process and the non-map parameters from starry that enter the design matrix, A, and therefore influence Σ and L_0.

For each sample of the non-map parameters, we can perform a draw from the conditional posterior over the map parameters by noting that the non-marginalized likelihood is Gaussian for y with mean ν and covariance Σ. So drawing

$$y \sim N\left(\nu, \Sigma\right) \tag{7}$$

or, equivalently,

$$y = \nu + Sn, \tag{8}$$

with

$$n \sim N(0, I) \tag{9}$$

a N_y vector of iid unit normal variables and letting $\Sigma = SS^T$ (i.e. S is the Cholesky factor of Σ, which has already been computed to efficiently implement the log-determinant term in the marginal likelihood).

There's nothing new here; it's just algebra. But note the organization of the calculation above: first Σ, then ν, and then $\log L_0$, which will really simplify your life. All these calculations are reproduced (in more generality) in Hogg et al. (2020).

Note that there are matrix multiplications with one dimension of N_d in the computation, but there are no $N_d \times N_d$ multiplications. Some implementation notes (in a perverted Einstein summation convention where repeated indices are summed, no matter how many times they appear):

1. If the lightcurve observations have independent noise, then Σ_d will be diagonal with entries σ_k^2, $k = 1, \ldots, N_d$, and it is more efficient to exploit this using

$$\left(A^T \Sigma_d^{-1} A\right)_{ij} = A_{ki} \frac{1}{\sigma_k^2} A_{kj}, \tag{10}$$

which involves $N_d \times N_y \times N_y$ computation and also

$$(A^T \Sigma_d^{-1} d)_i = A_{ki} \frac{1}{\sigma_k^2} d_k, \tag{11}$$

which involves $N_y \times N_d$ computation.

2. Often we are given Σ_s and not its inverse. We wish to avoid ever instantiating the components of the inverse matrix, which is a very unstable operation (we will assume, as in the prior point, that Σ_d is diagonal, so this worry does not apply to it). We can factor the expression for Σ^{-1} to obtain

$$\Sigma^{-1} = A^T \Sigma_d^{-1} A + \Sigma_s^{-1} = \Sigma_s^{-1} \left(\Sigma_s A^T \Sigma_d^{-1} A + I\right), \tag{12}$$

whence

$$\nu = \Sigma \left(A^T \Sigma_d^{-1} d + \Sigma_s^{-1} \mu\right) = \left(\Sigma_s A^T \Sigma_d^{-1} A + I\right)^{-1} \left(\Sigma_s A^T \Sigma_d^{-1} d + \mu\right). \tag{13}$$

To compute $\log\det 2\pi\Sigma$ we can use

$$\log\det 2\pi\Sigma = \log\det 2\pi\Sigma_s - \log\det\left(\Sigma_s A^T \Sigma_d^{-1} A + I\right). \tag{14}$$

(Note "$-$" arising from the inverse!) Finally, we want to obtain the Cholesky decomposition of Σ in order to draw from the Gaussian conditional posterior on the map y...TODO.

REFERENCES

Hogg, D. W., Price-Whelan, A. M., & Leistedt, B. 2020, arXiv e-prints, arXiv:2005.14199. https://arxiv.org/abs/2005.14199

Luger, R., Agol, E., Foreman-Mackey, D., et al. 2019, AJ, 157, 64, doi: 10.3847/1538-3881/aae8e5

Luger, R., Foreman-Mackey, D., & Hedges, C. 2021, AJ, 162, 124, doi: 10.3847/1538-3881/abfdb9