Noname manuscript No.
(will be inserted by the editor)

Full Bayes with general prior

Do you have a subtitle? If so, write it here

Fred J Hickernell · Jagadeeswaran Rathinavel

Received: date / Accepted: date

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords First keyword · Second keyword · More

1 Introduction

Your text comes here. Separate text sections with

2 Bayesian Cubature

The Bayesian approach for numerical analysis was popularized by Diaconis [**?**]. The earliest reference for such kind of approach dates back to Poincaré, where, the theory of interpolation was discussed. Diaconis motivates the reader by interpreting the most well known numerical methods, 1) trapezoidal rule and 2) splines, from the statistical point of view with whatever is known about the integrand as prior information. For example, the trapezoidal rule can be interpreted as a Bayesian method with prior information being modeled as a Brownian motion in the sample space $\mathcal{C}[0,1)$, the space of continuous functions.

This research is focused on the Bayesian approach for numerical integration that is known as Bayesian cubature as introduced by O'Hagan [**?**]. Bayesian cubature returns a probability distribution, that expresses belief about the

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

true value of integral, $\mu(f)$. This posterior probability distribution is based on a prior that depends on f, which is computed via Bayes' rule using the *data* contained in the function evaluations [**?**]. The distribution in general captures numerical uncertainty due to the fact that we have only used a finite number of function values to evaluate the integral.

2.1 Bayesian Posterior Error

We assume the integrand, f, is an instance of a stochastic Gaussian process, i.e., $f \sim \mathcal{GP}(m, s^2 C_{\boldsymbol{\theta}})$. Specifically, f is a real-valued random function with constant mean m and covariance function $s^2 C_{\boldsymbol{\theta}}$, where s is a positive scale factor, and $C_{\boldsymbol{\theta}} : [0,1]^d \times [0,1]^d \to \mathbb{R}$ is a symmetric, positive-definite function and, parameterized by $\boldsymbol{\theta}$:

$$\mathsf{C}^T = \mathsf{C}, \quad \boldsymbol{a}^T \mathsf{C} \boldsymbol{a} > 0, \quad \text{where } \mathsf{C} = \left(C_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)_{i,j=1}^n, \\ \text{for all } \boldsymbol{a} \neq 0,\ n \in \mathbb{N},\ \text{distinct } \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in [0,1]^d. \tag{1}$$

The covariance function, C, and the Gram matrix, C, depend implicitly on $\boldsymbol{\theta}$, but the notation may omit this for simplicity's sake. Procedures for estimating or integrating out the hyperparameters m, s, and $\boldsymbol{\theta}$ are explained later in this section.

For a Gaussian process, all vectors of linear functionals of f have a multivariate Gaussian distribution. For any deterministic sampling scheme with distinct nodes, $\{\boldsymbol{x}_i\}_{i=1}^n$, and defining $\boldsymbol{f} := (f(\boldsymbol{x}_i))_{i=1}^n$ as the multivariate Gaussian vector of function values, it follows from the definition of a Gaussian process that

$$\boldsymbol{f} \sim \mathcal{N}(m\mathbf{1}, s^2 \mathsf{C}), \tag{2a}$$

$$\mu \sim \mathcal{N}(m, s^2 c_0), \tag{2b}$$

$$\text{where } c_0 := \int_{[0,1]^d \times [0,1]^d} C_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{t}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{t}, \tag{2c}$$

$$\mathrm{cov}(\boldsymbol{f}, \mu) = \left(\int_{[0,1]^d} C(\boldsymbol{t}, \boldsymbol{x}_i) \, \mathrm{d}\boldsymbol{t} \right)_{i=1}^n =: \boldsymbol{c}. \tag{2d}$$

Here, c_0 and $\boldsymbol{c}$ depend implicitly on $\boldsymbol{\theta}$. We assume the covariance function C is simple enough that the integrals in these definitions can be computed analytically. We need the following lemma to derive the posterior error of our cubature.

Lemma 1 *[?, (A.6), (A.11–13)] If $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)^T \sim \mathcal{N}(\boldsymbol{m}, \mathsf{C})$, where $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are random vectors of arbitrary length, and*

$$\boldsymbol{m} = \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}(\boldsymbol{Y}_1) \\ \mathbb{E}(\boldsymbol{Y}_2) \end{pmatrix},$$

$$\mathsf{C} = \begin{pmatrix} \mathsf{C}_{11} & \mathsf{C}_{21}^T \\ \mathsf{C}_{21} & \mathsf{C}_{22} \end{pmatrix} = \begin{pmatrix} \mathrm{var}(\boldsymbol{Y}_1) & \mathrm{cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2) \\ \mathrm{cov}(\boldsymbol{Y}_2, \boldsymbol{Y}_1) & \mathrm{var}(\boldsymbol{Y}_2) \end{pmatrix}$$

then

$$\boldsymbol{Y}_1|\boldsymbol{Y}_2 \sim \mathcal{N}\big(\boldsymbol{m}_1 + \mathsf{C}_{21}^T\mathsf{C}_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{m}_2), \quad \mathsf{C}_{11} - \mathsf{C}_{21}^T\mathsf{C}_{22}^{-1}\mathsf{C}_{21}\big).$$

Moreover, the inverse of the matrix C *may be partitioned as*

$$\mathsf{C}^{-1} = \begin{pmatrix} \mathsf{A}_{11} & \mathsf{A}_{21}^T \\ \mathsf{A}_{21} & \mathsf{A}_{22} \end{pmatrix},$$
$$\mathsf{A}_{11} = (\mathsf{C}_{11} - \mathsf{C}_{12}\mathsf{C}_{22}^{-1}\mathsf{C}_{21})^{-1}, \qquad \mathsf{A}_{21} = -\mathsf{C}_{22}^{-1}\mathsf{C}_{21}\mathsf{A}_{11},$$
$$\mathsf{A}_{22} = \mathsf{C}_{22}^{-1} + \mathsf{C}_{22}^{-1}\mathsf{C}_{21}\mathsf{A}_{11}\mathsf{C}_{21}^T\mathsf{C}_{22}^{-1}.$$

It follows from Lemma **??** that the *conditional* distribution of the integral given observed function values, $\boldsymbol{f} = \boldsymbol{y}$ is also Gaussian:

$$\mu|(\boldsymbol{f} = \boldsymbol{y}) \sim \mathcal{N}\big(m(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1}) + \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}, \;\; s^2(c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})\big). \tag{3}$$

The natural choice for the cubature is the posterior mean of the integral, namely,

$$\widehat{\mu}|(\boldsymbol{f} = \boldsymbol{y}) = m(1 - \mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{c}) + \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}, \tag{4}$$

which takes the form of (**??**). Under this definition, the cubature error has zero mean and a variance depending on the choice of nodes:

$$(\mu - \widehat{\mu})|(\boldsymbol{f} = \boldsymbol{y}) \sim \mathcal{N}\left(0, \quad s^2(c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})\right).$$

A credible interval for the integral is given by

$$\mathbb{P}_f\left[|\mu - \widehat{\mu}| \le \mathrm{err}_{\mathrm{CI}}\right] = 99\%, \tag{5a}$$
$$\mathrm{err}_{\mathrm{CI}} = 2.58s\sqrt{c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}}. \tag{5b}$$

Naturally, 2.58 and 99% can be replaced by other quantiles and credible levels.

3 Full Bayes

Rather than using maximum likelihood to determine m and s, one can treat them as hyper-parameters with a non-informative, conjugate prior, namely $\boldsymbol{\rho}_{m,s^2}(\xi, \lambda) \propto 1/\lambda$. Then the posterior density for the integral given the data using Bayes theorem is,

$$
\begin{aligned}
&\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) \\
&\quad \propto \int_0^\infty \int_{-\infty}^\infty \rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}, m=\xi, s^2=\lambda)\rho_{\boldsymbol{f}}(\boldsymbol{y}|\xi,\lambda)\rho_{m,s^2}(\xi,\lambda)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad\qquad\qquad \text{by the properties of conditional probability} \\
&\quad \propto \int_0^\infty \int_{-\infty}^\infty \rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}, m=\xi, s^2=\lambda)\rho_{\boldsymbol{f}}(\boldsymbol{y}|\xi,\lambda)\rho_{m,s^2}(\xi,\lambda)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad\qquad\qquad \text{by Bayes' Theorem} \\
&\quad \propto \int_0^\infty \frac{1}{\lambda^{(n+3)/2}} \int_{-\infty}^\infty \exp\Biggl(-\frac{1}{2\lambda}\biggl\{\frac{[z-\xi(1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}]^2}{c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}} \\
&\qquad\qquad + (\boldsymbol{y}-\xi\boldsymbol{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\boldsymbol{1})\biggr\}\Biggr)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad \text{by (\textbf{??}), (\textbf{??}) and } \rho_{m,s^2}(\xi,\lambda)\propto 1/\lambda \\
&\quad \propto \int_0^\infty \frac{1}{\lambda^{(n+3)/2}} \int_{-\infty}^\infty \exp\left(-\frac{\alpha\xi^2-2\beta\xi+\gamma}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\xi\mathrm{d}\lambda,
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha &= (1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})^2 + \boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{1}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}), \\
\beta &= (1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})(z-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}) + \boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}), \\
\gamma &= (z-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y})^2 + \boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}).
\end{aligned}
$$

In the derivation above and below, factors that are independent of ξ, λ, or z can be discarded since we only need to preserve the proportion. But, factors that depend on ξ, λ, or z must be kept. Completing the square $\alpha\xi^2-2\beta\xi+\gamma = \alpha(\xi-\beta/\alpha)^2-(\beta^2/\alpha)+\gamma$, allows us to evaluate the integrals with respect to ξ and λ:

$$
\begin{aligned}
&\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) \\
&\quad \propto \int_0^\infty \frac{1}{\lambda^{(n+3)/2}} \exp\left(-\frac{\gamma-\beta^2/\alpha}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\cdots \\
&\qquad\qquad \cdots\int_{-\infty}^\infty \exp\left(-\frac{\alpha(\xi-\beta/\alpha)^2}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\quad \propto \int_0^\infty \frac{1}{\lambda^{(n+2)/2}} \exp\left(-\frac{\gamma-\beta^2/\alpha}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\lambda \\
&\quad \propto \left(\gamma-\frac{\beta^2}{\alpha}\right)^{-n/2} \propto \left(\alpha\gamma-\beta^2\right)^{-n/2}.
\end{aligned}
$$

Finally, we simplify the key term:

$$
\begin{aligned}
\alpha\gamma - \beta^2 = {} & \mathbf{1}^T \mathsf{C}^{-1}\mathbf{1}(c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})(z - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y})^2 \\
& - 2\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})(z - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}) \\
& + (1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})^2\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}) \\
& + [\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y} - (\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y})^2](c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})^2 \\
\propto {} & \mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}\left(z - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y} - \frac{(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}\right)^2 \\
& - \frac{[(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y}]^2}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}} + (1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})^2\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y} \\
& (c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})[\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y} - (\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y})^2] \\
\propto {} & \left(z - \left[\frac{(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})\mathbf{1}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}} + \boldsymbol{c}\right]^T \mathsf{C}^{-1}\boldsymbol{y}\right)^2 \\
& + \left[\frac{(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})^2}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}} + (c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})\right] \times \boldsymbol{y}^T\left[\mathsf{C}^{-1} - \frac{\mathsf{C}^{-1}\mathbf{1}\mathbf{1}^T\mathsf{C}^{-1}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}\right]\boldsymbol{y} \\
\propto {} & (z - \widehat{\mu}_{\text{full}})^2 + (n-1)\sigma^2_{\text{full}} \\
\propto {} & \left(1 + \frac{1}{n-1}\frac{(z - \mu_{\text{full}})^2}{\widehat{\sigma}^2_{\text{full}}}\right),
\end{aligned}
$$

i.e.,

$$
\alpha\gamma - \beta^2 \propto \left(1 + \frac{(z - \widehat{\mu}_{\text{full}})^2}{(n-1)\widehat{\sigma}^2_{\text{full}}}\right), \tag{6}
$$

where $\widehat{\mu}_{\text{full}} = \widehat{\mu}_{\text{EB}}$ and

$$
\widehat{\sigma}^2_{\text{full}} := \frac{1}{n-1}\boldsymbol{y}^T\left[\mathsf{C}^{-1} - \frac{\mathsf{C}^{-1}\mathbf{1}\mathbf{1}^T\mathsf{C}^{-1}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}\right]\boldsymbol{y}\times\left[\frac{(1 - \boldsymbol{c}^T\mathsf{C}^{-1}\mathbf{1})^2}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}} + (c_0 - \boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})\right].
$$

The confidence interval is:

$$
\mathbb{P}_f\left[|\mu - \widehat{\mu}_{\text{EB}}| \le \text{err}_{\text{full}}\right] = 99\%, \tag{7}
$$

where

$$
\text{err}_{\text{full}} := t_{n-1,0.995}\widehat{\sigma}_{\text{full}} > \text{err}_{\text{EB}}.
$$

Here $t_{n-1,0.995}$ denotes the 99.5 percentile of a standard Student's t-distribution with $n-1$ degrees of freedom. This means that $\mu|(\boldsymbol{f} = \boldsymbol{y})$, properly centered and scaled, has a Student's t-distribution with $n-1$ degrees of freedom. The estimated integral is the same as in the empirical Bayes case, $\widehat{\mu}_{\text{full}} = \widehat{\mu}_{\text{EB}}$, but the credible interval is wider. In other words, the stopping criterion for the full Bayes case is more conservative than that in the empirical Bayes case, (**??**) [**?**].

Because the shape parameter, $\boldsymbol{\theta}$, enters the definition of the covariance kernel in a non-trivial way, the only way to treat it as a hyperparameter and assign a tractable prior would be for the prior to be discrete. We believe in practice that choosing such a prior involves more guesswork than using the empirical Bayes estimate of $\boldsymbol{\theta}$ in (**??**) [**?**] or the cross-validation approach described next.

4 Full Bayes with general prior

Rather than using non-informative, conjugate prior one can use general prior, namely $\boldsymbol{\rho}_{m,s^2}(\xi,\lambda) \propto g(\lambda)$, which can generalize to any general function. One would be curious if the posterior function can be obtained from the data, i.e, the integrand values. The posterior density for the integral given the data using Bayes theorem is,

$$
\begin{aligned}
&\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) \\
&\quad \propto \int_0^\infty \int_{-\infty}^\infty \rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}, m=\xi, s^2=\lambda)\rho_{\boldsymbol{f}}(\boldsymbol{y}|\xi,\lambda)\rho_{m,s^2}(\xi,\lambda)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad\qquad\qquad\qquad \text{by the properties of conditional probability} \\
&\quad \propto \int_0^\infty \int_{-\infty}^\infty \rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}, m=\xi, s^2=\lambda)\rho_{\boldsymbol{f}}(\boldsymbol{y}|\xi,\lambda)\rho_{m,s^2}(\xi,\lambda)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad\qquad\qquad\qquad \text{by Bayes' Theorem} \\
&\quad \propto \int_0^\infty \frac{g(\lambda)}{\lambda^{(n+1)/2}} \int_{-\infty}^\infty \exp\Bigg(-\frac{1}{2\lambda}\Bigg\{\frac{[z-\xi(1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}]^2}{c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}} \\
&\qquad\qquad + (\boldsymbol{y}-\xi\boldsymbol{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\boldsymbol{1})\Bigg\}\Bigg)\,\mathrm{d}\xi\mathrm{d}\lambda \\
&\qquad\qquad \text{by (??), (??) and } \rho_{m,s^2}(\xi,\lambda) \propto g(\lambda) \\
&\quad \propto \int_0^\infty \frac{g(\lambda)}{\lambda^{(n+1)/2}} \int_{-\infty}^\infty \exp\left(-\frac{\alpha\xi^2-2\beta\xi+\gamma}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\,\mathrm{d}\xi\mathrm{d}\lambda,
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha &= (1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})^2 + \boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{1}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}), \\
\beta &= (1-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{1})(z-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y}) + \boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}), \\
\gamma &= (z-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{y})^2 + \boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y}(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c}).
\end{aligned}
$$

In the derivation above and below, factors that are independent of ξ, λ, or z can be discarded since we only need to preserve the proportion. But, factors that depend on ξ, λ, or z must be kept. Completing the square $\alpha\xi^2-2\beta\xi+\gamma = \alpha(\xi-\beta/\alpha)^2-(\beta^2/\alpha)+\gamma$, allows us to evaluate the integrals with respect to

ξ and λ:

$$\begin{aligned}
&\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y})\\
&\quad\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{(n+1)/2}} \exp\left(-\frac{\gamma-\beta^2/\alpha}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\cdots\\
&\qquad\qquad\cdots\int_{-\infty}^\infty \exp\left(-\frac{\alpha(\xi-\beta/\alpha)^2}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\xi\mathrm{d}\lambda\\
&\quad\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{n/2}} \exp\left(-\frac{\gamma-\beta^2/\alpha}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\lambda.
\end{aligned}$$

This can be interpreted as Laplace transform of $g(\lambda)$,

$$\begin{aligned}
\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) &\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{n/2}} \exp\left(-\frac{\gamma-\beta^2/\alpha}{2\lambda(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\right)\mathrm{d}\lambda\\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{n/2}} \exp\left(-\frac{1}{\lambda}\chi\right)\mathrm{d}\lambda,\\
&\quad \text{where}\quad \chi=\frac{\gamma-\beta^2/\alpha}{2(c_0-\boldsymbol{c}^T\mathsf{C}^{-1}\boldsymbol{c})}\propto 1+\frac{(z-\widehat{\mu}_{\text{full}})^2}{(n-1)\widehat{\sigma}^2_{\text{full}}}.
\end{aligned}$$

Let $\lambda=\dfrac{1}{w}$, $\quad \mathrm{d}\lambda=-w^{-2}\mathrm{d}w$ then,

$$\begin{aligned}
\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) &\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{n/2}} \exp\left(-\frac{1}{\lambda}\chi\right)\mathrm{d}\lambda\\
&= \int_0^\infty \frac{g(1/w)}{w^{-n/2}} \exp\left(-w\chi\right)(-w^{-2})\mathrm{d}w\\
&= \int_\infty^0 -g(1/w)w^{\frac{n}{2}-2}\exp\left(-w\chi\right)\mathrm{d}w\\
&= \int_0^\infty g(1/w)w^{\frac{n-4}{2}}\exp\left(-w\chi\right)\mathrm{d}w\\
&= \mathcal{LT}\{g(1/\cdot)\}^{\left(\frac{n-4}{2}\right)}(\chi),
\end{aligned}$$

where $\mathcal{LT}(\cdot)$ denotes the Laplace transform and $(\frac{n-4}{2})$ indicates the $\frac{n-4}{2}$th derivative taken after the transform. Here we used frequency domain derivative property of the Laplace transform. The above result can be further simplified by replacing $\gamma-\beta^2/\alpha$ from (**??**),

$$\begin{aligned}
\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) &\propto \mathcal{LT}\{g(1/\cdot)\}^{\left(\frac{n-4}{2}\right)}(\chi)\\
&\propto \mathcal{LT}\left\{g(1/\cdot)\right\}^{\left(\frac{n-4}{2}\right)}\left(1+\frac{(z-\widehat{\mu}_{\text{full}})^2}{(n-1)\widehat{\sigma}^2_{\text{full}}}\right)\quad \text{by}\quad (\textbf{??}).
\end{aligned}$$

Thus, $\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y})$ is proportional to $(\frac{n-4}{2})$th derivative of the Laplace transform of $g(1/\cdot)$ evaluated at χ, where $\chi\propto 1+\frac{(z-\widehat{\mu}_{\text{full}})^2}{(n-1)\widehat{\sigma}^2_{\text{full}}}$.

We demonstrate the general prior with the non-informative conjugate that we used above, i.e., if $g(1/\lambda) = \lambda$ then,

$$\begin{aligned}
\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) &= \int_0^\infty g(1/w) w^{\frac{n}{2}-2} \exp\left(-w\chi\right) \mathrm{d}w \\
&= \left(\mathcal{LT}(g(1/t))\right)^{(\frac{n}{2}-2)}|_{t=\chi} = \ \left(\mathcal{LT}(t)\right)^{(\frac{n}{2}-2)}|_{t=\chi} \\
&= \left(1/u^2\right)^{(\frac{n}{2}-2)}|_{u=\chi} \\
&\propto \chi^{-n/2} = \left(\frac{\gamma - \beta^2/\alpha}{2(c_0 - \boldsymbol{c}^T \mathsf{C}^{-1}\boldsymbol{c})}\right)^{-n/2} \\
&\propto \left(\gamma - \frac{\beta^2}{\alpha}\right)^{-n/2} \\
&\propto \left(\alpha\gamma - \beta^2\right)^{-n/2},
\end{aligned}$$

where we used the fact that the Laplace transform of $g(1/t) = t$ is $1/u^2$. After the transform, taking $(\frac{n}{2}-2)$th derivative gives us the result. This shows when using a generic prior, it leads to a posterior of the form $\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y}) \propto \mathcal{LT}\{g(1/\cdot)\}^{(\frac{n-4}{2})}(\chi)$ with full Bayes approach, i.e, the posterior $\rho_\mu(z|\boldsymbol{f}=\boldsymbol{y})$ is a function of $1 + \frac{(z-\widehat{\mu}_{\text{full}})^2}{(n-1)\widehat{\sigma}^2_{\text{full}}}$.

Our motivation to experiment with the general prior was to show that it may be possible to infer the prior from the integrand samples. We demonstrated it with the non-informative prior, which shows the possibility to compute the prior from function values. Obtaining an arbitrary prior from the integrand samples is the topic of future work.

5 Diagnostics for the Gaussian Process Assumption

The starting point for our Bayesian cubature is the assumption that the integrand arises from a Gaussian process. This means that the function data, $\boldsymbol{f}$, shall satisfy a multivariate Gaussian distribution, as in (**??**). We propose a transform, $\boldsymbol{Z} = \frac{1}{n}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H(\boldsymbol{f} - m\boldsymbol{1})$, to the given function data $\boldsymbol{f}$ to verify the assumption. We can verify that the transformed data, $\boldsymbol{Z}$, has zero mean and is also uncorrelated because

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{Z}\right] &= \frac{1}{n}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H(\mathbb{E}\left[\boldsymbol{f}\right] - m\boldsymbol{1}) \\
&= 0
\end{aligned}$$

$$
\begin{aligned}
\operatorname{cov}(\boldsymbol{Z}) &= \frac{1}{n^2}\mathbb{E}\left[\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H(\boldsymbol{f}-m\mathbf{1})(\boldsymbol{f}-m\mathbf{1})^T\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H\right] \\
&= \frac{1}{n^2}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H\mathbb{E}\left[(\boldsymbol{f}-m\mathbf{1})(\boldsymbol{f}-m\mathbf{1})^T\right]\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H \\
&= \frac{1}{n^2}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H\frac{1}{n}\mathsf{V}\mathsf{\Lambda}\mathsf{V}^H\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H \\
&= \frac{1}{n^3}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}(n)\mathsf{\Lambda}(n)\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H = \mathsf{I}
\end{aligned}
$$

Thus, the elements of $\boldsymbol{Z}$ are IID standard Gaussian random variables. In practice, using the estimated m_{EB} further simplifies,

$$
\begin{aligned}
\boldsymbol{Z} &= \frac{1}{n}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\mathsf{V}^H(\boldsymbol{f}-m\mathbf{1}) \\
&= \frac{1}{n}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}(\mathsf{V}^H\boldsymbol{f}-\frac{\tilde{f}_1}{n}\mathsf{V}^H\mathbf{1}) \\
&= \frac{1}{n}\mathsf{V}\mathsf{\Lambda}^{-\frac{1}{2}}\left(\mathsf{V}^H\boldsymbol{f}-\tilde{f}_1\left(1,0,\ldots,0\right)^T\right)
\end{aligned}
$$

To verify this hypothesis, we build a toy function using the known Bernoulli polynomial series,

$$
f_{\theta,r}(x) = \hat{f}_0 + \sqrt{\theta}\sum_{k\neq 0}^{N}\frac{f_{c,k}\cos(2\pi kx)+f_{s,k}\sin(2\pi kx)}{k^r}
$$

where $f_0 \sim \mathrm{IID}\mathcal{N}(c,b^2)$, $f_{c,k}, f_{s,k} \sim \mathrm{IID}\mathcal{N}(0,a^2)$ and a,b,c are constants. The series is truncated to an arbitrarily chosen length N. It has zero mean and covariance

$$
\begin{aligned}
\operatorname{cov}(f) &= \mathbb{E}[f(x)f(t)] \\
&= \operatorname{cov}(f_0)+\theta\sum_{k=1}^{N}\left(\mathbb{E}(f_{c,k}^2)\frac{\cos(2\pi kx)\cos(2\pi kt)}{k^{2r}}+\mathbb{E}(f_{s,k}^2)\frac{\sin(2\pi kx)\sin(2\pi kt)}{k^{2r}}\right) \\
&= b^2 + a^2\theta\sum_{k=1}^{N}\frac{\cos(2\pi k(x-t))}{k^{2r}}
\end{aligned}
$$

Figures **??** and **??** are normal probability plots of the Z_i using empirical Bayes estimates of m and $\boldsymbol{\theta}$. **more goes here**.

5.1 Posterior

We compute posterior distribution to verify the Gaussian assumption.

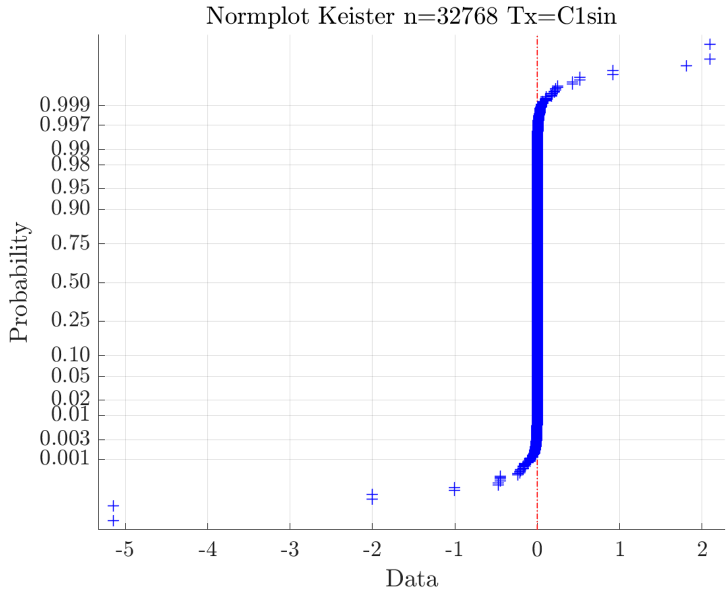

Fig. 1 Normal plot : Keister function with arbMean assumption

5.1.1 Non-informative prior

Assume the prior is of the form $\rho_{m,s^2}(\xi, \lambda) = \frac{1}{\lambda}$

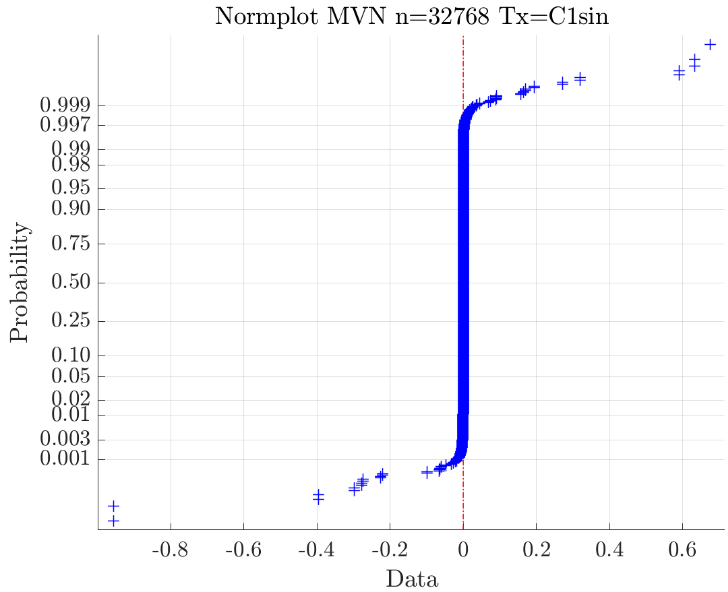

Fig. 2 Normal plot : MVN with arbMean assumption

$$
\begin{aligned}
\rho_{f(\boldsymbol{x})}(\boldsymbol{y}) &= \int_0^\infty \int_{-\infty}^\infty \rho_f(\boldsymbol{y}|m=\xi, s^2=\lambda)\rho_{m,s^2}(\xi,\lambda)\mathrm{d}\xi\mathrm{d}\lambda \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{1}{(\sqrt{2\pi})^n\sqrt{\det(\lambda\mathsf{C})}} \exp\left(-\frac{1}{2\lambda}(\boldsymbol{y}-\xi\mathbf{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\mathbf{1})\right)\frac{1}{\lambda}\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \frac{1}{(\sqrt{2\pi})^n}\int_0^\infty \frac{1}{\lambda^{\frac{n}{2}}\lambda}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}(\boldsymbol{y}-\xi\mathbf{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\mathbf{1})\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{1}{\lambda^{\frac{n}{2}+1}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}\underbrace{(\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1})}_{D}\left[\xi^2 - 2\xi\underbrace{\frac{\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}}_{A} + \underbrace{\frac{\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y}}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}}_{B}\right]\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{1}{\lambda^{\frac{n}{2}+1}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi^2-2\xi A + A^2 - A^2 + B\right]\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{1}{\lambda^{\frac{n}{2}+1}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi^2-2\xi A + A^2\right]\right)\mathrm{d}\xi \exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{1}{\lambda^{\frac{(n+1)}{2}}}\frac{1}{\sqrt{2\pi\lambda/D}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi-A\right]^2\right)\mathrm{d}\xi \exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{1}{\lambda^{\frac{(n+1)}{2}}}\exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&\propto (D[B-A^2])^{-\frac{(n-1)}{2}} \\
&\propto \left(\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y} - \frac{(\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y})^2}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}\right)^{-\frac{(n-1)}{2}} \\
&= \left(\boldsymbol{y}^T\left[\mathsf{C}^{-1} - \frac{(\mathsf{C}^{-1}\mathbf{1}\mathbf{1}^T\mathsf{C}^{-1})}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}\right]\boldsymbol{y}\right)^{-\frac{(n-1)}{2}} \\
&= \left((\boldsymbol{y}-m_{\mathrm{EB}})^T\mathsf{C}^{-1}(\boldsymbol{y}-m_{\mathrm{EB}})\right)^{-\frac{(n-1)}{2}}, \quad \text{where} \quad m_{\mathrm{EB}} = \frac{(\mathbf{1}^T\mathsf{C}^{-1}\boldsymbol{y})}{\mathbf{1}^T\mathsf{C}^{-1}\mathbf{1}}
\end{aligned}
$$

5.1.2 TODO: How to interpret as Student's t-distribution?

5.1.3 Generic prior

Assume the prior is of the form $\rho_{m,s^2}(\xi,\lambda) = g(\lambda)$

$$
\begin{aligned}
\rho_{f(\boldsymbol{x})}(\boldsymbol{y}) &= \int_0^\infty \int_{-\infty}^\infty \rho_f(\boldsymbol{y}|m=\xi, s^2=\lambda)\rho_{m,s^2}(\xi,\lambda)\mathrm{d}\xi\mathrm{d}\lambda \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{1}{(\sqrt{2\pi})^n\sqrt{\det(\lambda\mathsf{C})}} \exp\left(-\frac{1}{2\lambda}(\boldsymbol{y}-\xi\boldsymbol{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\boldsymbol{1})\right) g(\lambda)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \frac{1}{(\sqrt{2\pi})^n}\int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{n}{2}}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}(\boldsymbol{y}-\xi\boldsymbol{1})^T\mathsf{C}^{-1}(\boldsymbol{y}-\xi\boldsymbol{1})\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{n}{2}}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}\underbrace{(\boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{1})}_{D}\left[\xi^2 - 2\xi\underbrace{\frac{\boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{y}}{\boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{1}}}_{A} + \underbrace{\frac{\boldsymbol{y}^T\mathsf{C}^{-1}\boldsymbol{y}}{\boldsymbol{1}^T\mathsf{C}^{-1}\boldsymbol{1}}}_{B}\right]\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{n}{2}}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi^2-2\xi A + A^2 - A^2 + B\right]\right)\mathrm{d}\xi\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{n}{2}}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi^2-2\xi A + A^2\right]\right)\mathrm{d}\xi \exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{(n-1)}{2}}}\frac{1}{\sqrt{2\pi\lambda/D}}\int_{-\infty}^\infty \exp\left(-\frac{1}{2\lambda}D\left[\xi-A\right]^2\right)\mathrm{d}\xi \exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{(n-1)}{2}}}\exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda
\end{aligned}
$$

This can be interpreted as Laplace transform of $g(\lambda)$. Let $\eta = \frac{1}{2}D[B-A^2]$ and $\lambda = \frac{1}{w}, \mathrm{d}\lambda = -w^{-2}\mathrm{d}w$

$$
\begin{aligned}
\rho_{f(\boldsymbol{x})}(\boldsymbol{y}) &\propto \int_0^\infty \frac{g(\lambda)}{\lambda^{\frac{(n-1)}{2}}}\exp\left(-\frac{1}{2\lambda}D[B-A^2]\right)\mathrm{d}\lambda \\
&= \int_0^\infty \frac{g(1/w)}{w^{-\frac{(n-1)}{2}}}\exp\left(-w\eta\right)(-w^{-2})\mathrm{d}w \\
&= \int_\infty^0 -g(1/w)w^{\frac{(n-1)}{2}-2}\exp\left(-w\eta\right)\mathrm{d}w \\
&= \int_0^\infty g(1/w)w^{\frac{(n-5)}{2}}\exp\left(-w\eta\right)\mathrm{d}w \\
&= \mathcal{LT}\{g(1/\eta)\}^{(\frac{n-5}{2})'}
\end{aligned}
$$

where $\mathcal{LT}(.)$ denotes the Laplace transform and $(\frac{n-5}{2})'$ indicates the $(\frac{n-5}{2})$th derivative taken after the transform. Here we used frequency domain derivative property of the Laplace transform. $\rho_{f(\boldsymbol{x})}(\boldsymbol{y})$ is proportional to $(\frac{n-5}{2})$th derivative of the Laplace transform of $g(1/\eta)$.

$$
\begin{aligned}
D[B - A^2] &= \left(\boldsymbol{y}^T \mathsf{C}^{-1} \boldsymbol{y} - \frac{(\mathbf{1}^T \mathsf{C}^{-1} \boldsymbol{y})^2}{\mathbf{1}^T \mathsf{C}^{-1} \mathbf{1}} \right) \\
&= \left((\boldsymbol{y} - m_{\mathrm{EB}})^T \mathsf{C}^{-1} (\boldsymbol{y} - m_{\mathrm{EB}}) \right), \quad \text{where} \quad m_{\mathrm{EB}} = \frac{(\mathbf{1}^T \mathsf{C}^{-1} \boldsymbol{y})}{\mathbf{1}^T \mathsf{C}^{-1} \mathbf{1}}
\end{aligned}
$$

5.1.4 TODO: How to integrate further?