Bayesian Interpretation of Backus Gilbert methods

with L Del Debbio, M Panero, N Tantalo

Alessandro Lupo

Lattice@CERN 2024

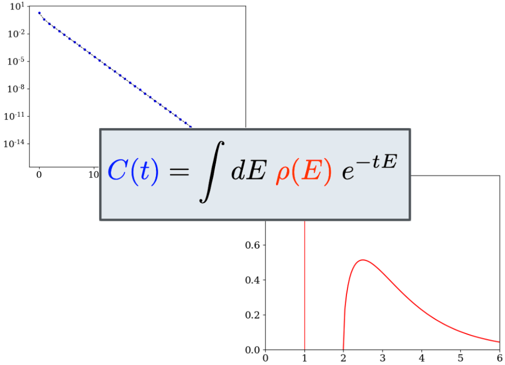
July 10, 2024

The inverse problem

- We are concerned with computing the spectral density $\rho(E)$ associated to a lattice correlator $C(t)$
- Ill-posed in presence of a finite set of noisy data
- There are ways to regularise the problem. Different assumptions, but one way to express the result

$$\rho_\sigma(E) = \sum_t g_t(\sigma; E)\, C(t)$$

$$\rho(E) = \lim_{\sigma \to 0} \rho_\sigma(E)$$



$$C(t) = \int dE\ \rho(E)\ e^{-tE}$$

Difficulties

- Finite set of measurements vs function with potentially continuous support

Difficulties

- Finite set of measurements vs function with potentially continuous support

- Target function is a distribution

Difficulties

- Finite set of measurements vs function with potentially continuous support
- Target function is a distribution
- Information is suppressed by $\exp(-tE)$

Difficulties

- Finite set of measurements vs function with potentially continuous support
- Target function is a distribution
- Information is suppressed by $\exp(-tE)$
- 💀 We work we data that is affected by errors

Smearing

- Smearing must be introduced to have a function that is smooth even in a finite volume

$$\rho_\sigma(\omega) = \int dE\, \mathcal{S}_\sigma(E,\omega)\, \rho(E)$$

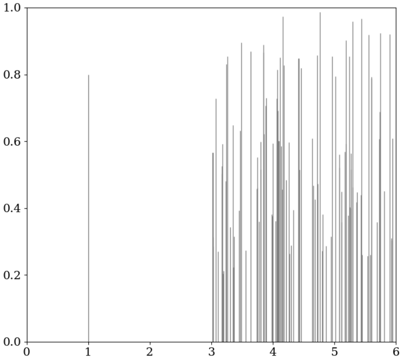
- Linear combinations of correlators automatically produce a smeared SD

$$\begin{aligned}\rho_\sigma(\omega) &= \sum_t g_t(\sigma;\omega)\; C(t) \\ &= \sum_t g_t(\sigma;\omega) \int dE\; e^{-tE} \rho(E)\end{aligned}$$

- We can now take the infinite volume limit

$$\lim_{L\to\infty} \rho_L(E) = 🚫$$

$$\lim_{\sigma\to 0} \lim_{L\to\infty} \rho_L(\sigma;E) = \rho(E)$$



Bayesian Inference with Gaussian Processes

- Aim for a probability distribution over a functional space of possible spectral densities

- Consider the stochastic field $\mathcal{R}(E)$ Gaussian-distributed around the prior value $\rho^{\text{prior}}(E)$ with covariance $\mathcal{K}^{\text{prior}}(E, E')$.

$$\mathcal{GP}\left(\rho^{\text{prior}}(E), \mathcal{K}^{\text{prior}}(E, E')\right)$$

- Similarly, assume that observational noise is Gaussian: $\eta(t)$

$$\mathbb{G}\left(\eta, \text{Cov}_d\right) = \exp\left(-\frac{1}{2}\vec{\eta}^T \, \text{Cov}_d^{-1} \, \vec{\eta}\right)$$

- The stochastic variable associated to the correlator, $\mathcal{C}$, is related to $\mathcal{R}$ and η via

$$\mathcal{C}(t) = \int dE \, e^{-tE} \mathcal{R}(E) + \eta(t)$$

- Incomplete list of references:
  FASTSUM collab. , Valentine, Sambridge 19 , Horak, Pawlowski, Rodríguez-Quintero, Turnwald, Urban 21
  Del Debbio, Giani, Wilson 21

- The joint, posterior distribution is again Gaussian, centred around ρ^{post} centre and variance:

$$\rho^{\mathrm{post}}(\omega) = \rho^{\mathrm{prior}}(\omega) + \sum_{t=1}^{t_{\max}} g_t^{\mathrm{GP}}(\omega) \left(C(t) - \int_0^\infty dE\, e^{-tE} \rho^{\mathrm{prior}}(E) \right)$$

$$\mathcal{K}^{\mathrm{post}}(\omega,\omega) = \left(\mathcal{K}^{\mathrm{prior}}(\omega,\omega) - \sum_{t=1}^{t_{\max}} g_t^{\mathrm{GP}}(\omega)\, f_t^{\ \mathrm{GP}}(\omega) \right)$$

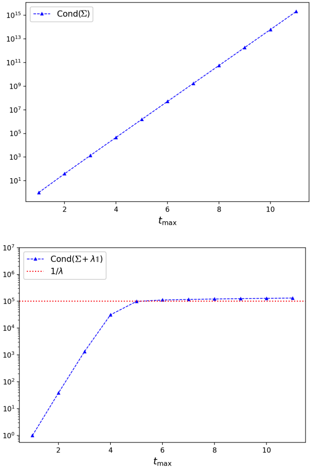
- The coefficients can be written as

$$\vec{g}^{\,\mathrm{GP}}(\omega) = (\Sigma^{GP} + \lambda \mathrm{Cov_d})^{-1}\, \vec{f}^{\ \mathrm{GP}}$$

- With the following ingredients:

$$\Sigma^{\mathrm{GP}}{}_{tr} = \int dE_1 \int dE_2\, e^{-tE_1}\, \mathcal{K}^{\mathrm{prior}}(E_1, E_2)\, e^{-rE_2} \quad \text{ill cond}$$

$$f_t^{\ \mathrm{GP}}(\omega) = \int dE\, \mathcal{K}^{\mathrm{prior}}(\omega, E)\, e^{-tE}$$



Backus-Gilbert methods: ideal world

- (HLT) Fix and target an appropriate smearing kernel such that when $\sigma \to 0$ we recover $S_\sigma(E,\omega) \to \delta(E-\omega)$

- We need to find the set of coefficients spanning $S_\sigma(E,\omega)$:

$$\sum_{\tau=1}^{\infty} g_\tau^{\mathrm{true}}(\sigma, E)\, e^{-a\tau\omega} = S_\sigma(E,\omega)$$

- We can find the coefficients by minimising

$$A[g] = \int_{E_0}^{\infty} dE\; e^{\alpha E} \left| \sum_{\tau=1}^{\infty} g_\tau(\sigma, E)\, e^{-a\tau\omega} - S_\sigma(E,\omega) \right|^2$$

- Without errors on $C(t)$ and infinitely many points, this is the solution.

- In reality, the correlator is known at a finite number of points. This translates into a systematic error in the reconstructed kernel and therefore in the reconstructed SD

$$\sum_{\tau=1}^{\tau_{\max}} g_\tau(\sigma, E)\, C(a\tau) = \rho_\sigma(E) + r(\tau_{\max}, \sigma; E)$$

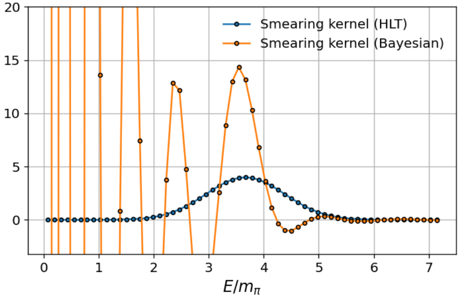
- The sum truncated to $\tau_{\max}$ is however well-defined and define unambiguously a given smearing kernel

- In fact, let us look at an example for both HLT and GP. For the latter, we shall choose a prior:

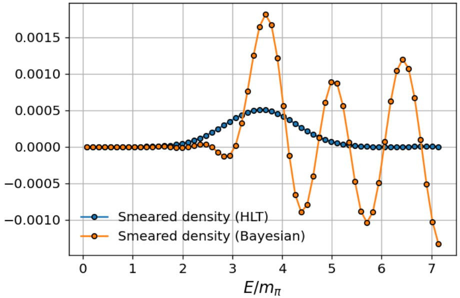
$$\mathcal{K}_\epsilon^{\text{prior}}(E, E') = \frac{e^{-(E-E')^2/2\epsilon^2}}{\lambda} \,, \qquad \rho^{\text{prior}} = 0$$

Backus-Gilbert methods: less ideal world

- Blue should be a Gaussian
- Orange should be what it should be



- Similarly for the reconstructed smeared density:



Backus-Gilbert methods: real world

- The main complication is that noisy data severely hinder this approach. Minimising $A[g]$ amounts to solve a massively ill-conditioned linear system

$$\vec{g} = \Sigma^{-1}\vec{f}$$

$$\Sigma_{tr} = \int dE_1 \; e^{-tE_1} \; e^{-rE_1}$$

- Backus-Gilbert regularisation:

$$\int_0^\infty dE \; e^{\alpha E} \; \left| \sum_{t=1}^{t_{\max}} g_t e^{-tE} - \mathcal{S}_\sigma(\omega, E) \right|^2 + \lambda \; \vec{g} \cdot \mathrm{Cov}_d \cdot \vec{g}$$

- The linear system is now

$$\vec{g} = (\Sigma + \lambda \mathrm{Cov_d})^{-1} \; \vec{f}$$

- In both cases the coefficients that generate the solution are written as:

$$\vec{g}(\omega) = \left(\Sigma + \lambda \mathrm{Cov_d}\right)^{-1} \vec{f}$$

$$\Sigma^{\mathrm{GP}}{}_{tr} = \int dE_1 \int dE_2 \, e^{-tE_1} \, \mathcal{K}^{\mathrm{prior}}(E_1, E_2) \, e^{-rE_2}$$

$$f_t{}^{\mathrm{GP}}(\omega) = \int dE \, \mathcal{K}^{\mathrm{prior}}(\omega, E) \, e^{-tE}$$

$$\Sigma^{\mathrm{HLT}}_{tr} = \int dE_1 \, e^{-tE_1} \, e^{-rE_1}$$

$$f^{\mathrm{HLT}}_t(\omega) = \int dE \, S_\sigma(\omega, E) \, e^{-tE}$$

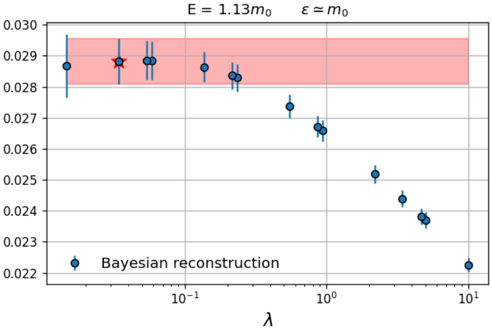
- They can be mapped into one another only at $\sigma = 0$.
- They regularise the problem in the very same way.
- What about λ?

Unphysical parameters & physical results

- λ introduces a bias. Recent application of BG methods perform a "stability analysis" {Bulava et al. 21 }

- We could do the same with the Bayesian reconstruction. Let us pick a prior:

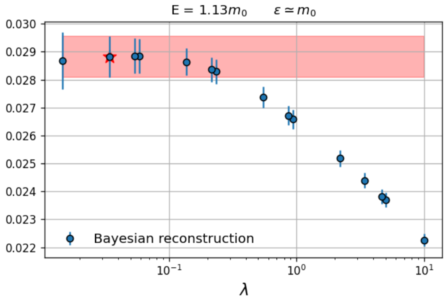
$$\mathcal{K}_{\epsilon}^{\text{prior}}(E, E') = \frac{e^{-(E-E')^2/2\epsilon^2}}{\lambda} , \qquad \rho^{\text{prior}} = 0$$

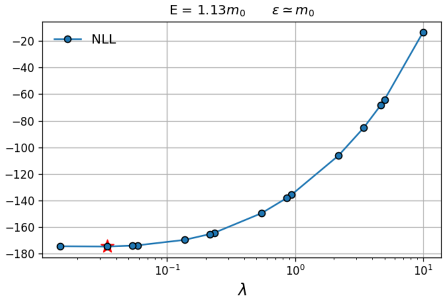


- In the Bayesian literature, hyperparameters are determined by minimising the negative log likelihood (NLL)

$$-\log P(\text{data}|\text{parameters})$$

- The methods seem compatible





Bayesian formulation of BG

- Compute the posterior probability distribution for a spectral density smeared with a **fixed kernel** $G_\sigma(E, E') = \exp^{-(E-E')^2/2\sigma^2}$

Bayesian formulation of BG

- Compute the posterior probability distribution for a spectral density smeared with a **fixed kernel** $G_\sigma(E,E') = \exp^{-(E-E')^2/2\sigma^2}$

❗ **Diagonal model covariance**

$$\mathcal{K}(E,E') = \frac{\delta(E-E')}{\lambda} \ ,$$

Bayesian formulation of BG

- Compute the posterior probability distribution for a spectral density smeared with a **fixed kernel** $G_\sigma(E, E') = \exp^{-(E-E')^2/2\sigma^2}$

! **Diagonal model covariance**

$$\mathcal{K}(E, E') = \frac{\delta(E - E')}{\lambda} \ ,$$

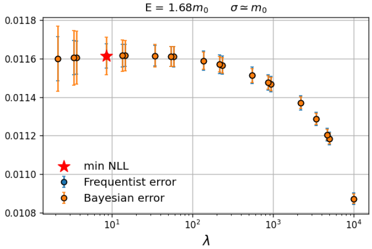
- The solution is now given by the same coefficients as HLT
$$g^{\text{GP}}(\sigma;\omega) = g(\sigma;\omega) \quad \text{even at finite } \sigma$$

Bayesian formulation of BG

- Compute the posterior probability distribution for a spectral density smeared with a **fixed kernel** $G_\sigma(E, E') = \exp^{-(E-E')^2/2\sigma^2}$

❗ **Diagonal model covariance**

$$\mathcal{K}(E, E') = \frac{\delta(E - E')}{\lambda} \quad ,$$



- The solution is now given by the same coefficients as HLT

$$g^{\mathrm{GP}}(\sigma;\omega) = g(\sigma;\omega) \quad \text{even at finite } \sigma$$

- The only difference is in the error (bootstrap for Backus-Gilbert methods)

$$\mathcal{K}^{\sigma}_{\mathrm{post}}(\omega,\omega)^2 = \frac{1}{2}\int dE \left(\sum_t g_t(\sigma,\omega)e^{-tE} - G_\sigma(E,\omega)\right) G_\sigma(E,\omega)$$

More on the bias

- Generate toys for spectral densities / correlators

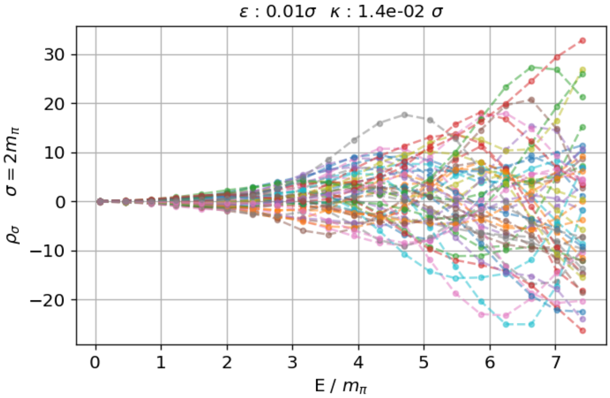
$$C(t) = \sum_{n=0}^{n_{\max}-1} w_n e^{-|t|E_n} , \qquad E_0 < E_1 \leq \dots ,$$

- We are generating instances of w_n with a GP, centred around zero, and covariance

$$K_{\text{weights}}(n, n') = \kappa \ \exp\left(-\frac{(E_n - E_{n'})^2}{2\epsilon^2} \right) ,$$

- with ϵ smaller than the spacing between states

- For the corresponding correlators, we inject noise from a covariance matrix measured on the lattice.

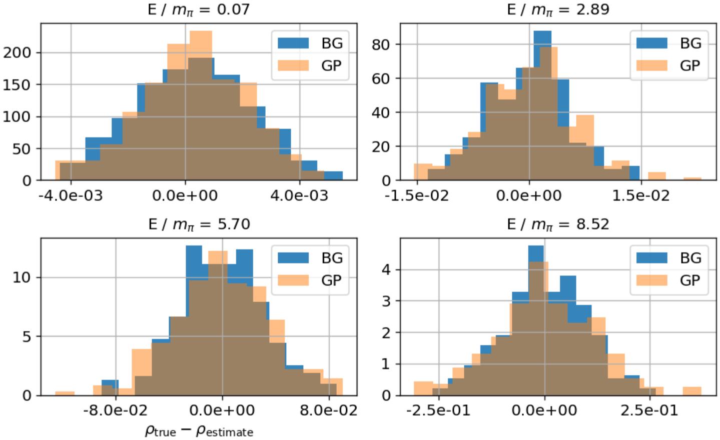


ε : 0.01σ κ : 1.4e-02 σ
ρ_σ $\sigma = 2m_\pi$
30
20
10
0
−10
−20
0
1
2
3
4
5
6
7
E / m_π

Preliminary results

PRELIMINARY

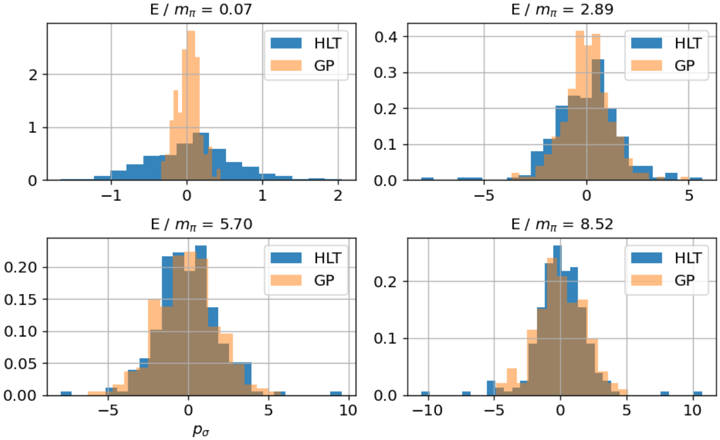
- Results for $\rho_\sigma(E)$ (true) - $\rho_\sigma(E)$ (estimate)



Preliminary results

PRELIMINARY

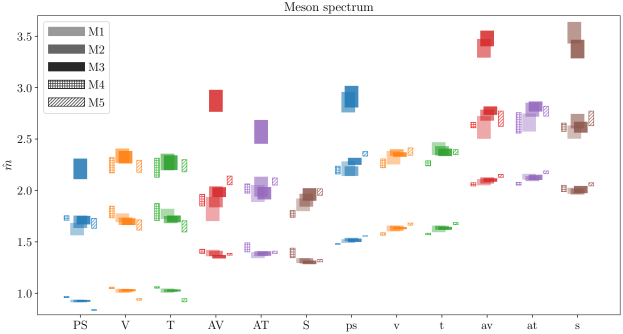
- Results for $p_\sigma(E) = \frac{\rho^{\text{true}}(E)_\sigma - \rho_\sigma^{\text{estimate}}(E)}{\Delta_\sigma^{\text{tot}}(E)}$



- Do I have more time?

Abusing this talk: spectroscopy

- In a previous paper [2211.09581] we explored the possibility to perform finite-volume spectroscopy using smeared spectral densities
- Recent developments in [2405.01388]



work with:

E. Bennett, L. Del Debbio, **N. Forzano**, R.C. Hill, D. K. Hong, H. Hsiao, J.-W. Lee, C.-J. D. Lin, B. Lucini, AL, M. Piai, D. Vadacchino, F. Zierler

- Do I have even more time?

Neural Networks

$$\begin{pmatrix} \text{A. Barone} \\ \text{A. De Santis} \\ \text{A. Lupo} \end{pmatrix}$$

Neural Networks

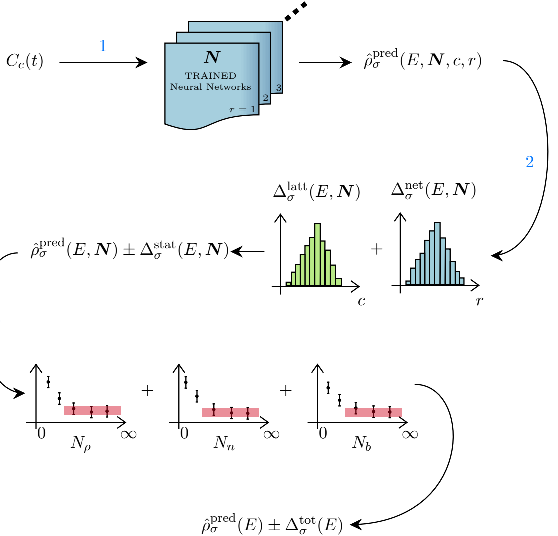
$$\begin{pmatrix} \text{A. Barone} \\ \text{A. De Santis} \\ \text{A. Lupo} \end{pmatrix}$$

- Non-linear algorithms: supervised deep-learning techniques
- The authors addressed the tasks of setting up a model independent training strategy and providing a reliable estimate of systematic errors

✔ A functional-basis (Chebyshev) to parametrize the correlators and the smeared spectral densities of the training sets in a model independent way

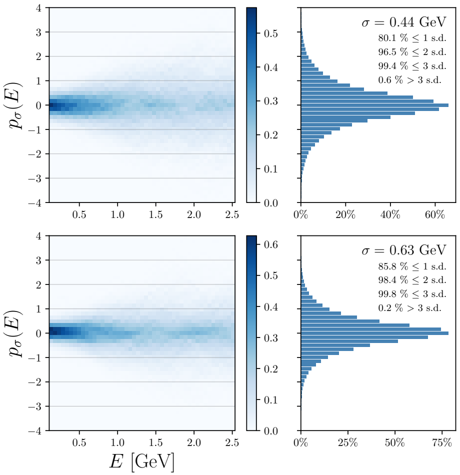
✔ the introduction of the ensemble of machines to estimate over the systematic error

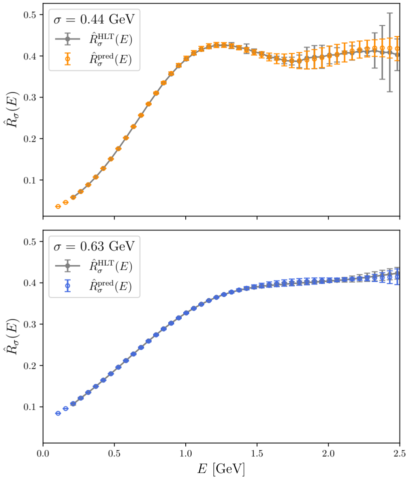
Neural Networks



$$p_\sigma(E) = \frac{\hat{\rho}_\sigma^{\text{pred}}(E) - \hat{\rho}_\sigma^{\text{true}}(E)}{\Delta_\sigma^{\text{tot}}(E)}$$

[Credits: Buzzicotti, De Santis, Tantalo]





[Credits: De Santis, Tantalo, et al.]

Backup

- Consider Gaussian probability density for the vector $\psi \in \mathbb{R}^{p+d}$ with covariance Σ,

$$\mathbb{G}[\psi;\Sigma] = \frac{1}{\sqrt{\det\left(2\pi\Sigma^{-1}\right)}} \exp\left(-\frac{1}{2}\,\psi^T \Sigma\,\psi\right) \;.$$

- Block diagonal:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \;, \qquad \psi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \,,$$

- $\phi_1 \in \mathbb{R}^p, \phi_2 \in \mathbb{R}^d$. Σ_{11} is a $p \times p$ matrix, Σ_{22} is $d \times d$, and Σ_{12} and Σ_{21} are $p \times d$ and $d \times p$ respectively

- We perform a LDU decomposition of the total covariance Σ. To this end, we introduce the matrices L and R,

$$L = \begin{pmatrix} 1_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & 1_d \end{pmatrix} \ , \qquad R = \begin{pmatrix} 1_p & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & 1_d \end{pmatrix} \ ,$$

- such that $W \equiv L\Sigma R$ is diagonal,

$$W = \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \ ,$$

- where the Schur complement of Σ_{11} appears in the top-left block. The inverse of the covariance can be now written as

$$\Sigma^{-1} = R\, W^{-1}\, L \ . \tag{1}$$

- The previous equations can be used to evaluate the scalar product $\psi^T \Sigma \psi$:

$$\begin{aligned} \psi^T \Sigma\, \psi &= \left(\phi_1 - \phi_{(1|2)}\right)^T \Sigma^{-1}_{(11|2)} \left(\phi_1 - \phi_{(1|2)}\right) + \phi_2^T\, \Sigma_{22}\, \phi_2 \ , \\ \phi_{(1|2)} &\equiv \Sigma_{12} \Sigma_{22}^{-1} \phi_2 \ , \\ \Sigma_{(11|2)} &\equiv \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \ , \end{aligned} \tag{2}$$

- as well as the determinant of Σ^{-1},

$$\det \Sigma = \det \left(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right) \det \left(\Sigma_{22} \right) \ . \tag{3}$$

- then the joint probability can be rewritten as

$$\mathbb{G}\left[\psi ; \Sigma\right] = \mathbb{G}\left[\phi_1 - \phi_{(1|2)} ; \Sigma_{(11|2)}\right] \mathbb{G}\left[\phi_2 ; \Sigma_{22}\right] \ . \tag{4}$$

- Factorising the conditional probability $\mathbb{G}\left[\phi_1 - \phi_{(1|2)} ; \Sigma_{(11|2)}\right]$ as in the Eq in the main text.

- In the context of GPs the hyperparameters (including λ) are selected so that the resulting probability of observing the data,

$$\mathbb{G}\left[\vec{C}^{\,\text{obs}} - \vec{C}^{\,\text{prior}};\, \Sigma + \text{Cov}_d\right] \ , \tag{5}$$

is maximised.

- Equivalently, one minimises the "negative logarithmic likelihood" (NLL)

$$\frac{1}{2}\,\text{Log}\det\left(\Sigma + \text{Cov}_d\right) + \frac{1}{2}\left(\vec{C}^{\,\text{obs}} - \vec{C}^{\,\text{prior}}\right)\frac{1}{\Sigma + \text{Cov}_d}\left(\vec{C}^{\,\text{obs}} - \vec{C}^{\,\text{prior}}\right) \tag{6}$$