

Bayesian Approach to Inverse Problems

L Del Debbio

Higgs Centre for Theoretical Physics
University of Edinburgh

work in collaboration with

A Candido, A Chiefa, T Giani, A Lupo, M Panero, G Petrillo, N Tantalo,
M Wilson

inverse problems

ubiquitous in physics, geosciences, engineering...

$$y_I = \int dx C_I(x) f(x)$$

... are known to be ill-defined problems

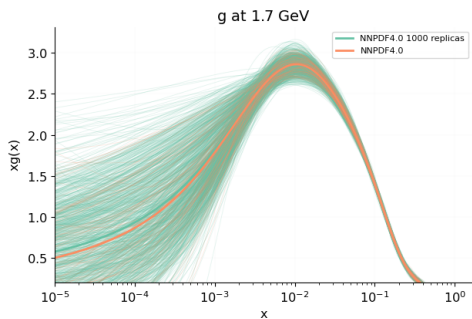
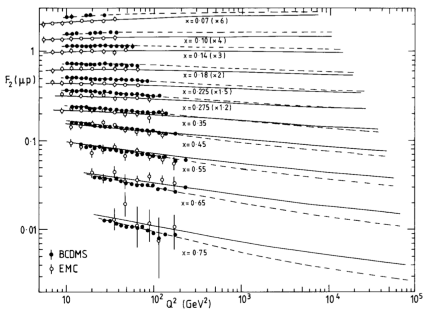
↔ simple parametrization could lead to a biased result for f

examples in particle physics:

- PDFs from DIS/lattice: y_I structure function data, $f(x)$ PDFs
- spectral densities: y_t Euclidean correlators, $f(x)$ spectral function

multiple approaches to this problem: fix functional fits, NN, Backus-Gilbert

example: PDFs from data



bayesian approach

- f is promoted to be a *stochastic process*
- $f(x)$ for $x \in \mathcal{I}$ is a set of stochastic variables
- for any given \mathbf{f} , where $f_i = f(x_i)$, we have a prior $p(\mathbf{f})$
- all a priori knowledge about f is encoded in p (more later)
- posterior distribution obtained from Bayes theorem

$$\tilde{p}(\mathbf{f}) = p(\mathbf{f}|y) = \frac{p(y|\mathbf{f})p(\mathbf{f})}{p(y)}$$

- information about the solution is encoded in the posterior, eg

central value : $E_{\tilde{p}}[\mathbf{f}]$

covariance : $\text{Cov}_{\tilde{p}}[\mathbf{f}, \mathbf{f}']$

gaussian process

we work with GPs, a specific kind of stochastic process

$$f \sim \mathcal{GP}(m, k),$$

where

$$m : \mathcal{I} \rightarrow \mathbb{R}, \quad k : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$$

for a GP, the vector of stochastic variables \mathbf{f}

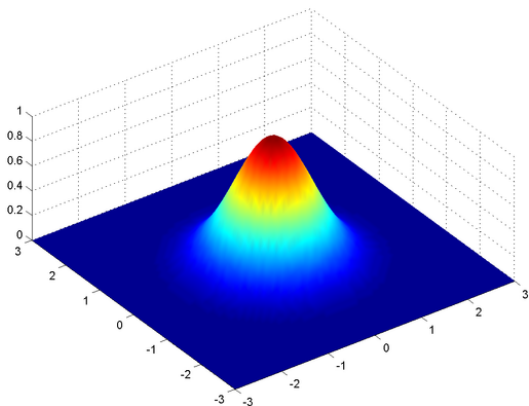
$$\mathbf{x} = \{x_i; i = 1, \dots, N\}, \quad \mathbf{f} = f(\mathbf{x}) = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} \in \mathbb{R}^N, \quad f_i = f(x_i)$$

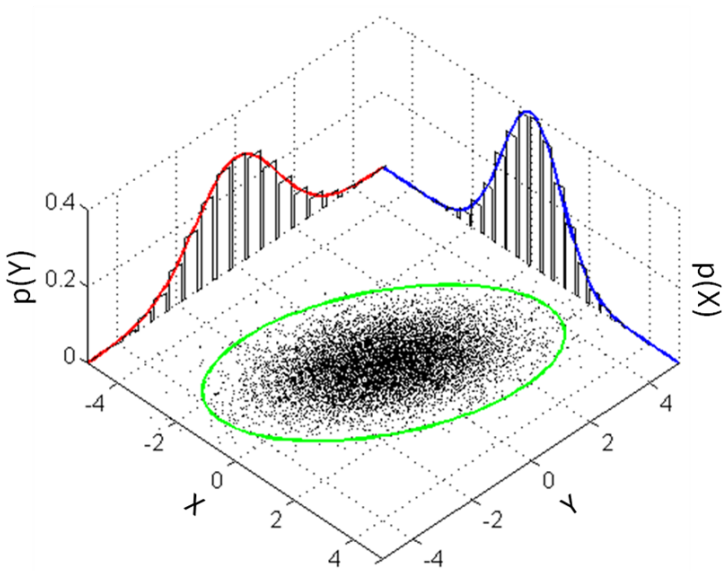
is distributed as a multidimensional Gaussian

$$\mathbf{f} \sim \mathcal{N}(\mathbf{m}, K),$$

for $N = 2$,

$$p(\mathbf{f}) = p(f_1, f_2) = p(f(x_1), f(x_2))$$





prior distribution

mean & covariance

$$\mathbf{m} = m(\mathbf{x}), \quad K = k(\mathbf{x}, \mathbf{x}^T),$$

$$E[f_i] = m_i = m(x_i),$$

$$\text{Cov}[f_i, f_j] = K_{ij} = k(x_i, x_j).$$

specific choices for this work - viz. Gibbs kernel

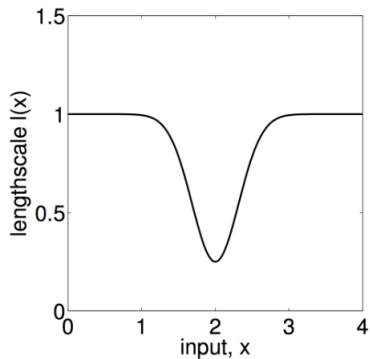
$$m(x) = 0$$

$$k(x, x') = \sigma^2 \sqrt{\frac{2l(x)l(y)}{l^2(x) + l^2(y)}} \exp \left[-\frac{(x - y)^2}{l^2(x) + l^2(y)} \right]$$

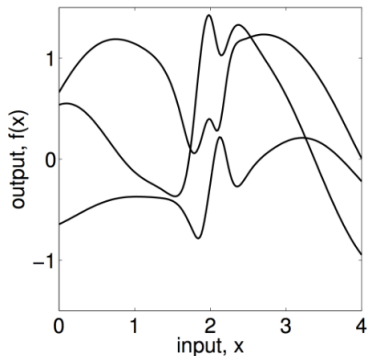
$$l(x) = l_0 \times (x + \delta)$$

hyperparameters : $\theta = (\sigma, l_0)$

Gibbs kernel interpretation



(a)



(b)

setting the problem

sampling f at points $\mathbf{x} = \{x_i; i = 1, \dots, N\}$ and $\mathbf{x}^* = \{x_i^*; i = 1, \dots, M\}$

$$\mathbf{f} \in \mathbb{R}^N, \quad \mathbf{f}^* \in \mathbb{R}^M,$$

the prior probability distribution is a multidimensional Gaussian

$$p(\mathbf{f}, \mathbf{f}^* | \theta) = \frac{1}{\sqrt{\det(2\pi K)}} \times \exp \left\{ -\frac{1}{2} \left((\mathbf{f} - \mathbf{m})^T, (\mathbf{f}^* - \mathbf{m}^*)^T \right) K^{-1} \begin{pmatrix} \mathbf{f} - \mathbf{m} \\ \mathbf{f}^* - \mathbf{m}^* \end{pmatrix} \right\},$$

K is now an $(N + M) \times (N + M)$ matrix

$$K = \begin{pmatrix} k(\mathbf{x}, \mathbf{x}^T) & k(\mathbf{x}, \mathbf{x}^{*T}) \\ k(\mathbf{x}^*, \mathbf{x}^T) & k(\mathbf{x}^*, \mathbf{x}^{*T}) \end{pmatrix} = \begin{pmatrix} K_{\mathbf{x}\mathbf{x}} & K_{\mathbf{x}\mathbf{x}^*} \\ K_{\mathbf{x}^*\mathbf{x}} & K_{\mathbf{x}^*\mathbf{x}^*} \end{pmatrix}.$$

data and theory predictions

dataset central values: $\mathbf{y} = \{y_I, I = 1, \dots, N_{\text{dat}}\}$

dataset fluctuations: $\epsilon \sim \mathcal{N}(0, C_Y)$

linear dependence on f :

$$T_I = \int_{\mathcal{I}} dx C_I(x) f(x) \approx \sum_{i=1}^N (\text{FK})_{Ii} f_i$$

NB: applies to both quasi/pseudo-PDFs and spectral densities

$$E[T_I] = (\text{FK})_{Ij} m_j$$
$$\text{Cov}[T_I, T_J] = (\text{FK})_{Ii} (K_{\mathbf{xx}})_{ij} (\text{FK})_{jJ}^T$$

posterior distribution

we want to determine

$$p(\mathbf{f}, \mathbf{f}^*, \theta | y) = p(\mathbf{f}, \mathbf{f}^* | \theta, y) p(\theta | y)$$

compute each factor independently

$$p(\mathbf{f}, \mathbf{f}^* | \theta, y) \propto \exp \left\{ -\frac{1}{2} \left((\mathbf{f} - \mathbf{m})^T, (\mathbf{f}^* - \mathbf{m}^*)^T \right) K^{-1} \begin{pmatrix} \mathbf{f} - \mathbf{m} \\ \mathbf{f}^* - \mathbf{m}^* \end{pmatrix} \right\} \\ \times \exp \left\{ -\frac{1}{2} \left((\mathbf{FK})\mathbf{f} - y \right)^T C_Y^{-1} \left((\mathbf{FK})\mathbf{f} - y \right) \right\} .$$

posterior distribution

integrating over \mathbf{f}^* yields

$$\int d\mathbf{f}^* p(\mathbf{f}, \mathbf{f}^* | \theta, y) \propto \exp \left\{ -\frac{1}{2} (\mathbf{f} - \mathbf{m})^T K_{\mathbf{xx}}^{-1} (\mathbf{f} - \mathbf{m}) \right\} \\ \times \exp \left\{ -\frac{1}{2} ((\mathbf{FK})\mathbf{f} - y)^T C_Y^{-1} ((\mathbf{FK})\mathbf{f} - y) \right\}$$

introduce $C_{YT} = (\mathbf{FK})K_{\mathbf{xx}}(\mathbf{FK})^T + C_Y$

posterior distribution is Gaussian

$$p(\mathbf{f} | \theta, y) = \mathcal{N}(\mathbf{f}; \tilde{\mathbf{m}}, \tilde{K}_{\mathbf{xx}})$$

$$\tilde{\mathbf{m}} = \mathbf{m} + K_{\mathbf{xx}}(\mathbf{FK})^T C_{YT}^{-1} (y - (\mathbf{FK})\mathbf{m})$$

$$\tilde{K}_{\mathbf{xx}} = K_{\mathbf{xx}} - K_{\mathbf{xx}}(\mathbf{FK})^T C_{YT}^{-1} (\mathbf{FK})K_{\mathbf{xx}}$$

posterior distribution

integrating over \mathbf{f}

$$p(\mathbf{f}^* | \theta, y) = \mathcal{N}(\tilde{\mathbf{m}}^*, \tilde{K}_{\mathbf{x}\mathbf{x}}^*)$$

$$\tilde{\mathbf{m}}^* = \mathbf{m}^* + K_{\mathbf{x}^*\mathbf{x}}(\mathbf{F}\mathbf{K})^T C_{Y^T}^{-1} (\mathbf{y} - (\mathbf{F}\mathbf{K})\mathbf{m}) ,$$

$$\tilde{K}_{\mathbf{x}^*\mathbf{x}^*} = K_{\mathbf{x}^*\mathbf{x}^*} - K_{\mathbf{x}^*\mathbf{x}}(\mathbf{F}\mathbf{K})^T C_{Y^T}^{-1} (\mathbf{F}\mathbf{K})K_{\mathbf{x}\mathbf{x}} .$$

- correction to the mean proportional to $(\mathbf{y} - (\mathbf{F}\mathbf{K})\mathbf{m})$
- correlations in the prior allow to make predictions for \mathbf{f}^*

inference for hyperparameters

using Bayes theorem

$$p(\theta|y) = \frac{p(y|\theta) p_{\theta}(\theta)}{\int d\theta p(y|\theta) p_{\theta}(\theta)},$$

the condition

$$y = (\text{FK}) \mathbf{f} + \epsilon$$

yields

$$p(y|\theta) = \frac{e^{-\frac{1}{2} (y - (\text{FK})\mathbf{m})^T C_{YT}^{-1} (y - (\text{FK})\mathbf{m})}}{\sqrt{\det [2\pi C_{YT}]}}.$$

$p(\theta|y)$ can be sampled by MCMC

PDF from DIS - a toy model

study the triplet PDF T_3

$$T_3 = (u + \bar{u}) - (d + \bar{d})$$

using data for the DIS *triplet* structure function from BCDMS

$$y = F_2^p - F_2^d = C_{T_3} \otimes T_3$$
$$C_Y = \text{Cov} [F_2^p, F_2^p] + \text{Cov} [F_2^d, F_2^d] - 2\text{Cov} [F_2^p, F_2^d]$$

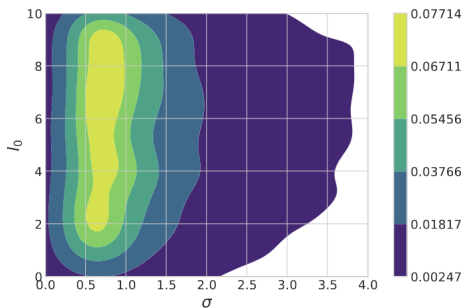
test the methodology using synthetic data

$$y = (\text{FK}) \mathbf{f}_0 + \eta, \quad \text{with} \quad \eta \sim \mathcal{N}(0, C_Y)$$

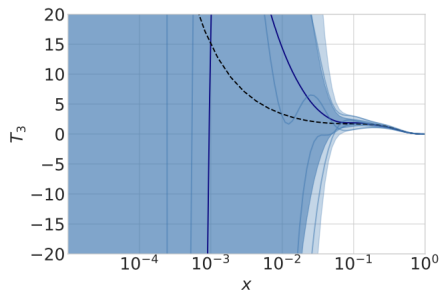
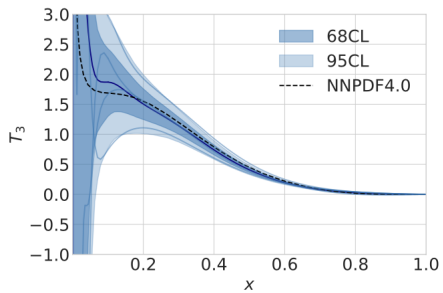
where \mathbf{f}_0 is taken from a known PDF (NNPDF4.0)

inference for the hyperparameters

starting from **flat** priors for the hyperparameters



inference for the PDF



interpretation of the results - closure test

vanishing exp errors

$$y = y_0 = (\text{FK})\mathbf{f}_0, \quad C_Y = 0$$

yields

$$\tilde{\mathbf{m}}^* = R_{\mathbf{x}^*\mathbf{x}}^{(0)} \mathbf{f}_0$$

where we introduced the smearing kernel

$$R_{\mathbf{x}^*\mathbf{x}}^{(0)} = K_{\mathbf{x}^*\mathbf{x}} (\text{FK})^T [(\text{FK})K_{\mathbf{xx}}(\text{FK})^T]^{-1} (\text{FK})$$

the result of Bayesian inference is a smeared version of the 'true' answer

similarly

$$\tilde{\mathbf{m}} = R_{\mathbf{xx}}^{(0)} \mathbf{f}_0, \quad \tilde{K}_{\mathbf{xx}} = \left(\mathbb{1} - R_{\mathbf{xx}}^{(0)} \right) K_{\mathbf{xx}}$$

in data space, consider *bias* and *variance*

$$\mathcal{B} = (\text{FK}) (\tilde{\mathbf{m}} - \mathbf{f}_0) = (\text{FK}) \left(R_{\mathbf{xx}}^{(0)} - \mathbf{1} \right) \mathbf{f}_0 = 0,$$

$$\mathcal{V} = (\text{FK}) \tilde{K} (\text{FK})^T = (\text{FK}) \left(\mathbf{1} - R_{\mathbf{xx}}^{(0)} \right) K_{\mathbf{xx}} (\text{FK})^T = 0.$$

- the methodology reconstructs the input experimental data exactly, independently of the specific values of the hyperparameters
- the model function is not in general reconstructed exactly, *i.e.* $\tilde{\mathbf{m}} \neq \mathbf{f}_0$
- in the functional space, a residual reconstruction error is still present

adding experimental errors

reconstruction kernel

$$R_{\mathbf{x}^*\mathbf{x}} = K_{\mathbf{x}^*\mathbf{x}} (\text{FK})^T \left[(\text{FK}) K_{\mathbf{xx}} (\text{FK})^T + C_Y \right]^{-1} (\text{FK})$$

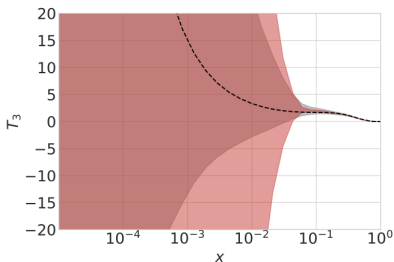
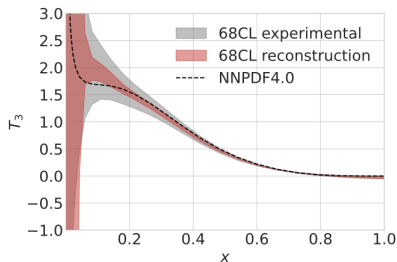
comparison with the 'true' input

$$\begin{aligned} \tilde{\mathbf{m}} - \mathbf{f}_0 &= [R_{\mathbf{xx}} - \mathbb{1}] \mathbf{f}_0 + a_{\mathbf{xx}}^T \eta \\ \tilde{K}_{\mathbf{xx}} &= (\mathbb{1} - R_{\mathbf{xx}}) K_{\mathbf{xx}} (\mathbb{1} - R_{\mathbf{xx}})^T + a_{\mathbf{xx}}^T C_Y a_{\mathbf{xx}} \end{aligned}$$

in data space

$$\begin{aligned} \mathcal{B} &= (\text{FK}) [R_{\mathbf{xx}} - \mathbb{1}] \mathbf{f}_0 + (\text{FK}) a_{\mathbf{xx}}^T \eta \\ \mathcal{V} &= (\text{FK}) (\mathbb{1} - R_{\mathbf{xx}}) K_{\mathbf{xx}} (\mathbb{1} - R_{\mathbf{xx}})^T (\text{FK})^T + (\text{FK}) a_{\mathbf{xx}}^T C_Y a_{\mathbf{xx}} (\text{FK})^T \end{aligned}$$

- limited reconstruction due to smearing, functional uncertainty
- functional uncertainty is not cured by more precise data
- the term proportional to η is the propagation of the experimental error in the reconstructed function, experimental uncertainty



conclusions

- bayesian analysis \longrightarrow rigorous tool to study inverse problems
- all hypotheses are explicitly spelled out in the prior
- for linear data: analytical results useful to build intuition
- Backus-Gilbert methods can be rephrased as GP processes with specific prior [Idd et al 24]
- non trivial connection with NN parametrizations (WIP, next time!)
- being used for PDFs and spectral densities
- crucial for **robust** errors on PDFs and spectral densities