Stochastic Framework for Parton Distributions Fits

L Del Debbio

Higgs Centre for Theoretical Physics
The University of Edinburgh

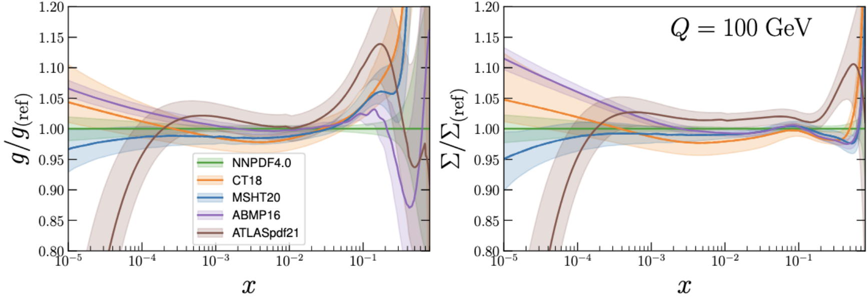
work in collaboration with

A Chiefa, R Kenway
T Giani, A Candido, G Petrillo, M Wilson
A Lupo, N Tantalo, M Provero

see also

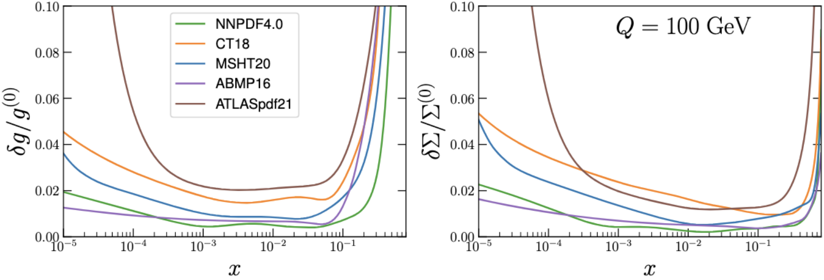
- [Idd et al (2022)] Gaussian Processes and PDF fits
- [Candido et al (2024)] more on GPs and PDF fits
- [Idd et al (2024)] relation between GPs and Backus-Gilbert
- [Chiefa et al (2025)] NN training and PDF uncertainties

state of the art - central values



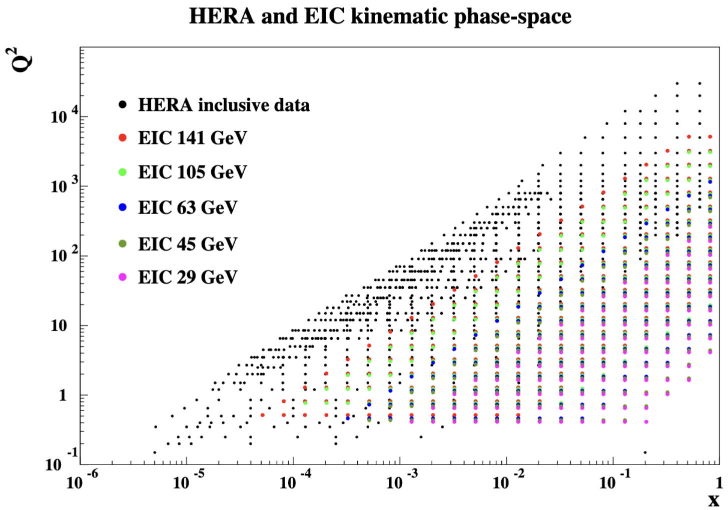
[Acta Phys Polon. B (2022) 53]

state of the art - error bars



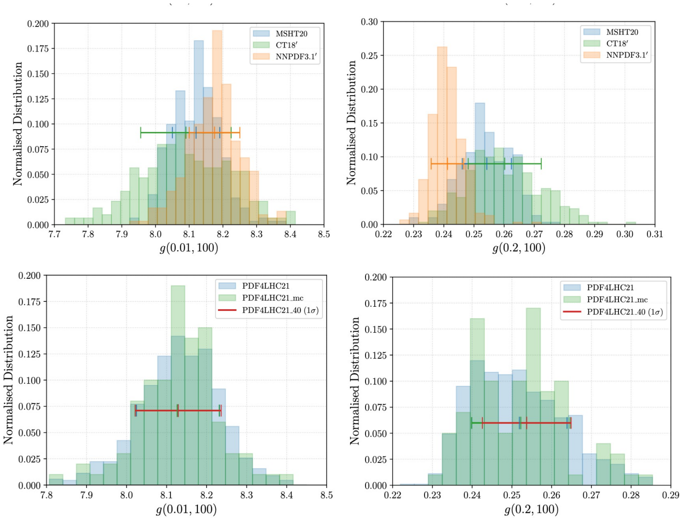
[Acta Phys Polon. B (2022) 53]

impact of EIC data



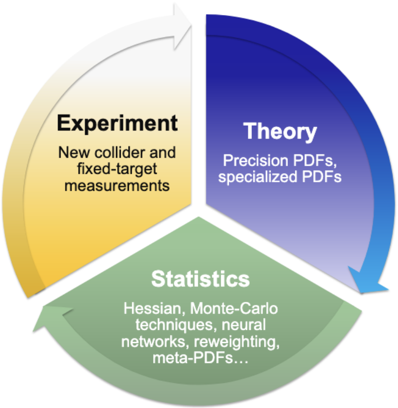
[EPJC (2023) 83:1011]

state of the art - combined PDF sets



[PDF4LHC21 (2022)]

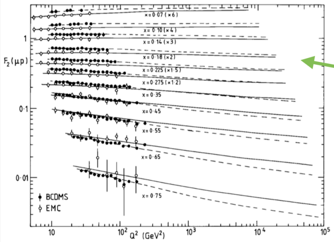
focus on methodology



plan of the talk

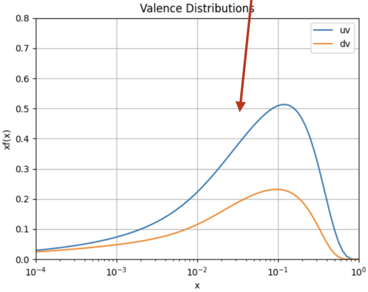
- origin of the problem
- stochastic processes for inverse problems
- how to rephrase various methodologies in a common framework

inverse problems



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

[NNPDF4.0]



inverse problems

focus on **linear** forward maps

$$T_I = \int dx\, C_I(x) f(x) = (\mathrm{FK})_{I\alpha} f_\alpha\,, \quad f_\alpha = f(x_\alpha)$$

$$I = 1, \ldots, N_{\mathrm{dat}}\,, \quad \alpha = 1, \ldots, N_{\mathrm{grid}}$$

- DIS data: $f(x)$ are PDFs
- quasi-PDFs/pseudo-PDFs from lattice QCD: $f(x)$ are PDFs
- ill-defined: results depend on the regularization
- central value/error bars to be determined by data, not by the methodology

minimizing the χ^2

$$
\begin{aligned}
\chi^2 &= \sum_{I,J=1}^{N_{\rm dat}} \left(Y_I - T_I\right) \left(C_Y^{-1}\right)_{IJ} \left(Y_J - T_J\right) \\
&= \left(f - f_{\rm min}\right)^T M \left(f - f_{\rm min}\right) + \chi^2_{\rm min}
\end{aligned}
$$

where

$$
M = (\mathrm{FK})^T C_Y^{-1} (\mathrm{FK})
$$

the solution is not unique/unstable:

$$
\begin{aligned}
f_{\rm min} &= M^{+} (\mathrm{FK})^T C_Y^{-1} Y + f_{\parallel} \\
f_{\parallel} &\in \mathrm{Ker}(M)
\end{aligned}
$$

fitting to a functional form is **one** way to regularize the problem (bias)

stochastic processes for inverse problems

- solution represented as a stochastic process [Idd et al (2022)]
- we need to assign a prior probability distribution $p(f)$

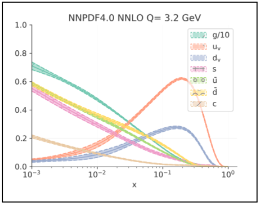
$$p(f) = \int dh\, p(h)\, p(f|h)$$

- data used to update the prior to the posterior $\tilde{p} = p(f|Y)$

$$\tilde{p}(f) = p(f|Y) = \int dh\, p(h|Y)\, p(f|h, Y)$$

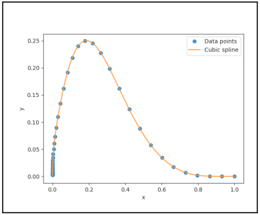
- observables computed as averages over the posterior

stochastic processes in practice



- we want to characterize the solution $f(x)$
- usually, focus on central value & error
- **here**: complete description from the **joint probability distribution**

$$\tilde{p}(f_1, \ldots, f_{N_{\mathrm{grid}}})\,, \quad f_\alpha = f(x_\alpha)\,, \quad \alpha = 1, \ldots, N_{\mathrm{grid}}$$



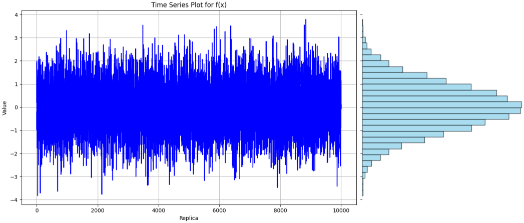
- and then compute

$$\bar{f}(x) \pm \delta f(x) = \mathbb{E}_{\tilde{p}}\left[f(x)\right] \pm \sqrt{C(x,x)}\,,$$
$$C(x,x') = \mathbb{E}_{\tilde{p}}\left[\left(f(x) - \bar{f}(x)\right)\left(f(x') - \bar{f}(x')\right)\right]$$

- and study any other observable $O(f)$

probability distributions as ensembles of replicas

- probability distributions are represented by ensembles of *replicas*



- observables are computed from averages over replicas

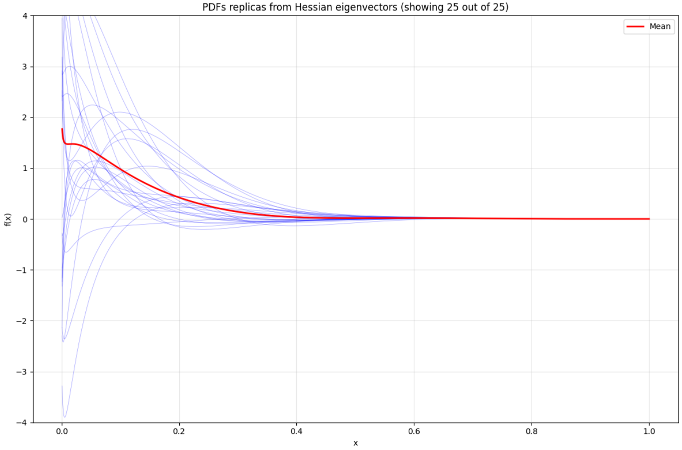
$$\bar{O} = \mathbb{E}_p\left[O(f)\right] = \int df\, p(f)\, O(f) = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} O\left(f^{(k)}\right)$$

fit to functional form & Hessian errors

- **choose** a functional form and minimize the χ^2 to determine the best fit parameters
- **assume** the probability distribution for the parameters is a Gaussian
- the covariance is given by the inverse of the Hessian matrix at the minimum of the loss
- Hessian errors are given in terms of eigenvectors $S_i^{\pm}$ in parameter space
- the uncertainty is then propagated to any other observable

$$\sigma_F = \frac{1}{2}\sqrt{1 \sum_{i=1}^{N_{\rm par}} \left[F(S_i^+) - F(S_i^-)\right]^2}$$

replicas from Hessian eigenvectors



GP methodology

- choose a Gaussian Process prior $p(f|h) = \mathcal{GP}\left(0, K(\sigma, \ell)\right)$
- hyperparameters $h = (\sigma, \ell)$ control the shape of the kernel
- update prior to posterior using Bayes theorem

$$p(f|h, Y) \propto p(Y|f, h)\, p(f|h) = \mathcal{GP}\left(\tilde{m}, \tilde{K}\right)$$

- posterior parameters

$$\tilde{m} = \left[K \frac{1}{1 + MK}\right] (\mathrm{FK})^T C_Y^{-1} Y$$
$$\tilde{K}^{-1} = K^{-1} + M$$

- MCMC sampling of hyperparameters, $p(h|Y)$ [Candido et al (2024)]

Tikhonov regularization

choosing

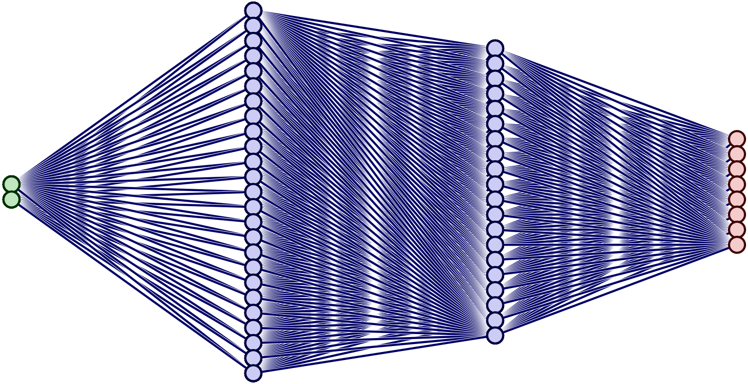
$$K_{\mathrm{Tik}}(\sigma, \ell) = \sigma^2 I$$

we recover Tikhonov regularization with parameter $\lambda = 1/\sigma^2$

$$\tilde{m} = \frac{1}{M + \lambda} (\mathrm{FK})^T C_Y^{-1} Y$$

Tikhonov regularization is equivalent to a GP with a diagonal kernel

NNPDF methodology



$$\rho_i^{(\ell)} = \rho\left(\phi_i^{(\ell)}\right)\,, \quad \phi_i^{(\ell)} = \sum_{j=1}^{n_{\ell-1}} w_{ij}^{(\ell)} \rho_j^{(\ell-1)} + b_i^{(\ell)}$$

$$\boxed{f(x) = \phi^{(L)}(x;\theta)}$$

NN prior distribution

parameters θ are initialized using a Glorot-Normal distribution

initialize weights and biases using Gaussians

$$\langle b_i^{(\ell)} \rangle = 0 \,, \quad \langle b_{i_1}^{(\ell)} b_{i_2}^{(\ell)} \rangle = \delta_{i_1 i_2} C_b^{(\ell)}$$

$$\langle w_{ij}^{(\ell)} \rangle = 0 \,, \quad \langle w_{i_1 j_1}^{(\ell)} w_{i_2 j_2}^{(\ell)} \rangle = \delta_{i_1 i_2} \delta_{j_1 j_2} \frac{C_w^{(\ell)}}{n_{\ell-1}}$$

parameters/functions duality

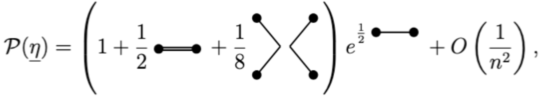
$$p(\phi^{(\ell)}) = \int \underbrace{[dw\, p(w)]\, [db\, p(b)]}_{dh\, p(h)} \underbrace{\prod_{i,\alpha} \delta \left(\phi_{i\alpha}^{(\ell)} - \sum_j w_{ij}^{(\ell)} \rho \left(\phi_{j\alpha}^{(\ell-1)} \right) - b_i^{(\ell)} \right)}_{p(f|h)}$$

EFT approach

symmetry and $1/n$ counting yield the probability distribution for ϕ

$$
\begin{aligned}
p(\phi^{(\ell)}) &= \frac{1}{Z}\,\exp\left[-S(\phi^{(\ell)})\right] \\
&= \frac{1}{Z}\,\exp\left[-\frac{1}{2}\gamma^{(\ell)}_{\alpha_1\alpha_2}\phi^{(\ell)}_{\alpha_1}\cdot\phi^{(\ell)}_{\alpha_2}\right. \\
&\qquad\qquad \left. -\frac{1}{8n_{\ell-1}}\gamma^{(\ell)}_{\alpha_1\alpha_2,\alpha_3\alpha_4}\phi^{(\ell)}_{\alpha_1}\cdot\phi^{(\ell)}_{\alpha_2}\,\phi^{(\ell)}_{\alpha_3}\cdot\phi^{(\ell)}_{\alpha_4} + O(1/n^2_{\ell-1})\right]
\end{aligned}
$$

correlators can be computed using Feynman diagrams



going deep – recursion relations

two-pt function at leading order

$$K^{(\ell+1)}_{\alpha_1\alpha_2} = C_b^{(\ell+1)} + C_w^{(\ell+1)} \frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle \Big|_{O(1)}$$

$$= C_b^{(\ell+1)} + C_w^{(\ell+1)} \frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle_{K^{(\ell)}}$$

where

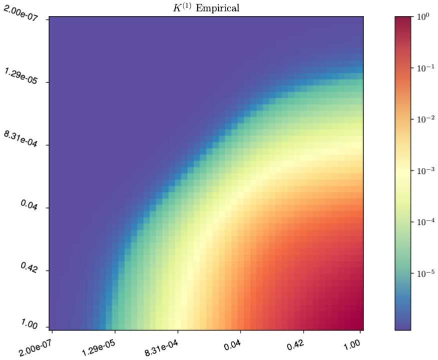
$$\frac{1}{n_\ell} \langle \vec{\rho}^{(\ell)}_{\alpha_1} \cdot \vec{\rho}^{(\ell)}_{\alpha_2} \rangle_{K^{(\ell)}} = \int \prod_\alpha d\phi_\alpha \, \frac{e^{-\frac{1}{2}\left(K^{(\ell)}\right)^{-1}_{\beta_1\beta_2} \phi_{\beta_1}\phi_{\beta_2}}}{\left|2\pi K^{(\ell)}\right|^{1/2}} \, \rho(\phi_{\alpha_1})\rho(\phi_{\alpha_2})$$

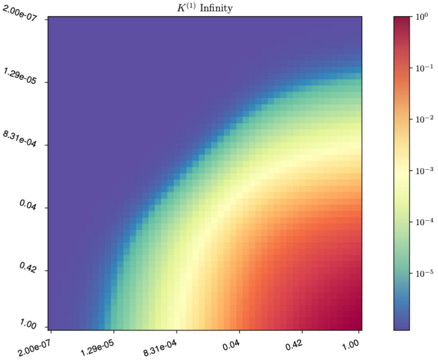
increasing ℓ, the couplings *evolve*, exactly like an RG evolution

$$\text{prior:} \quad p(f) = \int dh \, p(f|h) \, p(h) = p(\phi^{(L)})$$

covariance matrices at initialization

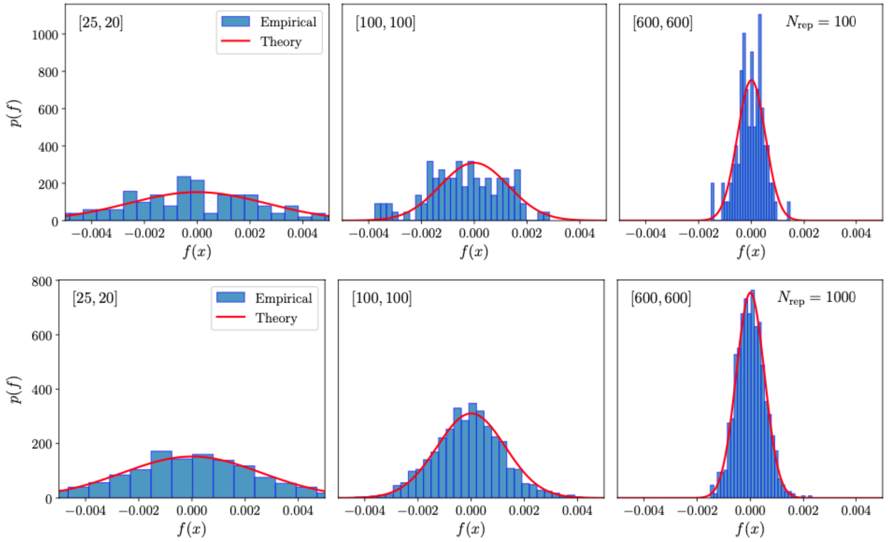
covariances in the first layer – consistency check





NNPDF initialization - 1

$xT_3(x), x = 0.0065$ replica distribution at initialization



training

for all parametrizations and for a **quadratic loss**

$$\mathcal{L}_t = \frac{1}{2} \left(Y - T[f_t]\right)^T C_Y^{-1} \underbrace{\left(Y - T[f_t]\right)}_{\epsilon_t}$$

gradient descent

$$\begin{aligned}
\frac{d}{dt}\theta_{t,\mu} &= -\nabla_\mu \mathcal{L}_t \\
\frac{d}{dt} f_t &= (\nabla_\mu f_t) \frac{d}{dt}\theta_{t,\mu} = \Theta_t \underbrace{\left(\frac{\partial T[f]}{\partial f}\right)_t}_{(\mathrm{FK})^T} C_Y^{-1} \epsilon_t
\end{aligned}$$

where

$$\Theta_t = (\nabla_\mu f_t)(\nabla_\mu f_t)^T$$

is the Neural Tangent Kernel (NTK)

solution of the training flow

integrating the flow equation

$$f_t = \mathcal{F}_t(f_0, Y)$$

solution of the inverse problem is the trained function at time T

f_T is a stochastic process, with its distribution $\tilde{p}(f_T)$

introduce an auxiliary field L_t

$$p(f, L|Y) = \frac{1}{Z} \exp\Big\{-S(\phi_0) - \int dt\, L_t \left[\frac{d}{dt} f_t + H_t f_t\right]\Big\}$$

extra-dimensional field theory encoding training dynamics

outlook

- inverse problems are ill-defined and need regularization
- stochastic processes provide a rigorous framework
- all hypotheses are clearly spelled in the prior
- explore different priors: GPs, NNs, fixed functional forms
- NNs provide flexible priors with controlled complexity
- compute and compare posterior distributions for different methodologies
- improved precision from the latest lattice determination of α_s