The Research Council of Norway

University of Stavanger

Bayesian approaches to

The inverse problem challenge in Lattice QCD

Alexander Rothkopf

@rothkopfAK

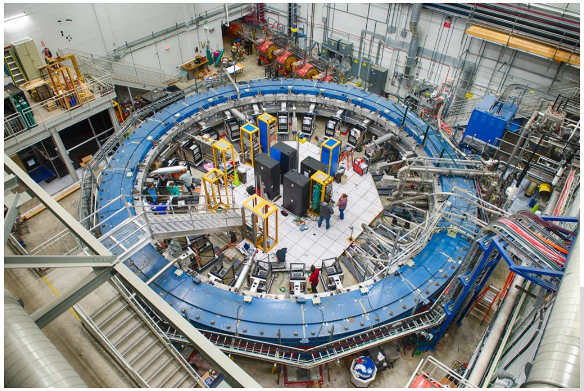
Faculty of Science and Technology
Department of Mathematics and Physics
University of Stavanger

Norwegian Particle, Astroparticle & Cosmology Theory network

Physics beyond weak coupling

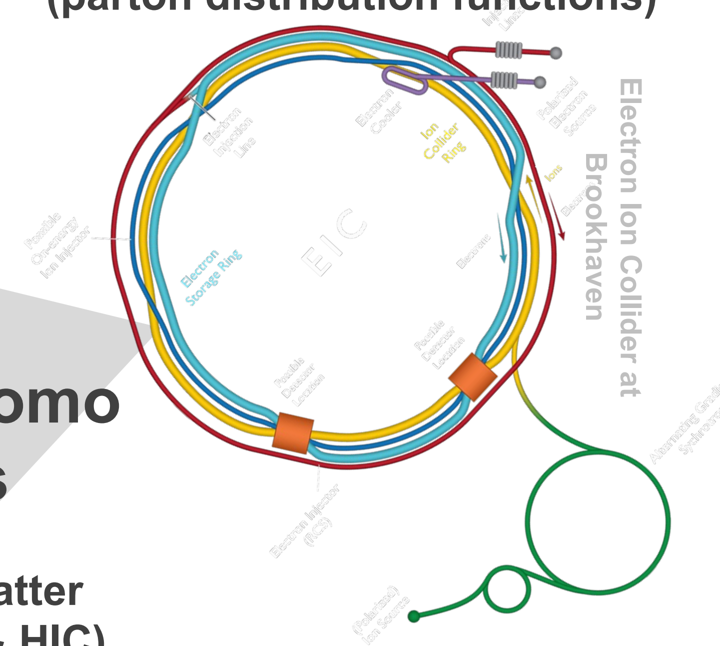
University of Stavanger

Fermilab E989 experiment



Precision particle physics (muon g-2)

Nuclear tomography (parton distribution functions)

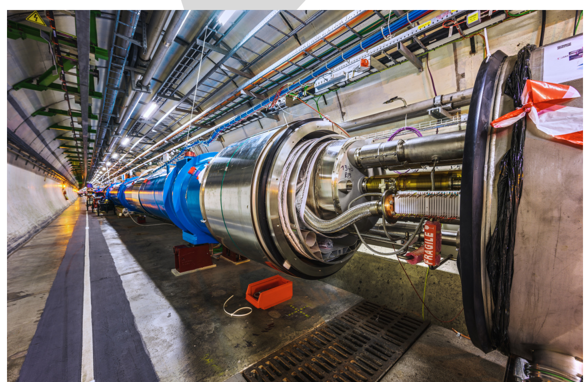


Electron Ion Collider at Brookhaven

Quantum Chromo Dynamics

Primordial nuclear matter (quark-gluon-plasma & HIC)

Large Hadron Collider at CERN

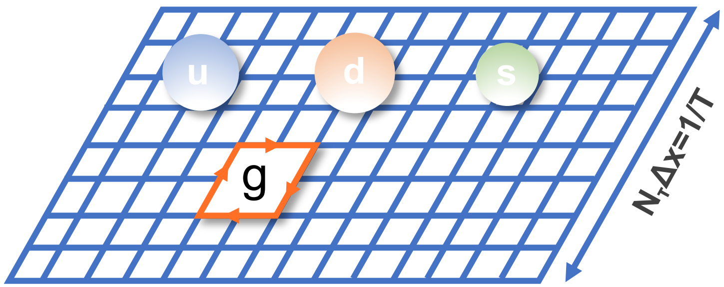


A robust tool: lattice QCD

University of Stavanger

- Non-perturbative 1st principles approach to Quantum Chromo Dynamics

for textbooks see e.g. Montvay and Münster or Gattringer and Lang



- Gluon fields as links: $U_\mu(x) = \exp[ig\Delta x_\mu A_\mu(x)]$
- Quark fields ψ(x) with realistic masses

$$\langle \mathcal{O}(t_1)\mathcal{O}(t_2)\rangle = \int \mathcal{D}U\, \mathcal{O}[U,t_1]\,\mathcal{O}[U,t_2]\, e^{iS_{\rm QCD}[U]}$$

- Euclidean time as key for Monte Carlo & finite temperature
  - Analytic continuation of real-time into imaginary time: statistical interpretation

$$\langle \mathcal{O}[U]\rangle = \int \mathcal{D}U\, \mathcal{O}[U]\, e^{-S_{\rm E}[U]} \Rightarrow \langle \mathcal{O}\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \mathcal{O}[U^{\{k\}}] \quad P[U] \propto e^{-S_{\rm E}[U]}$$

  - Subtle but vital: finite extent in imaginary time = inverse temperature

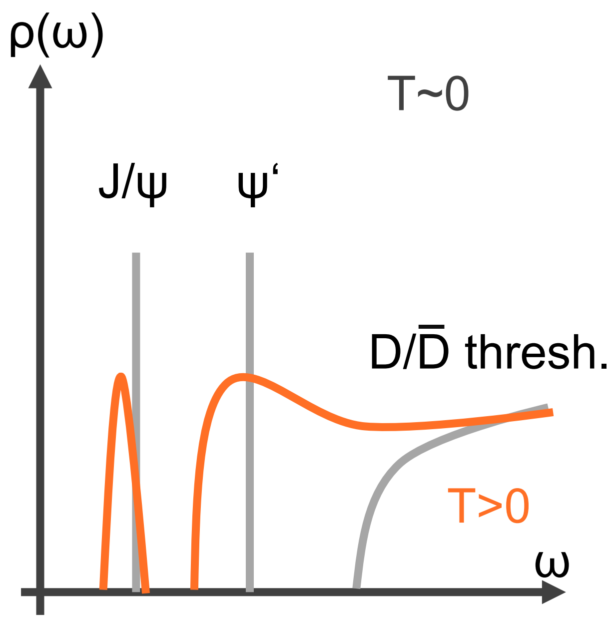
- Only after Δx →0 & N Δx → ∞ physical results ensue (see FLAG report)

Y.Aoki et.al. arXiv:2111.09849

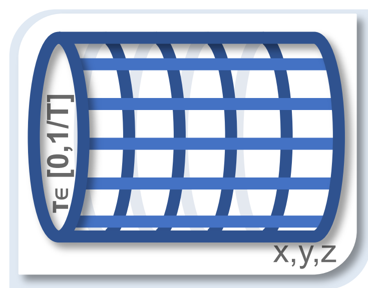
Extracting insight from lattice QCD

University of Stavanger

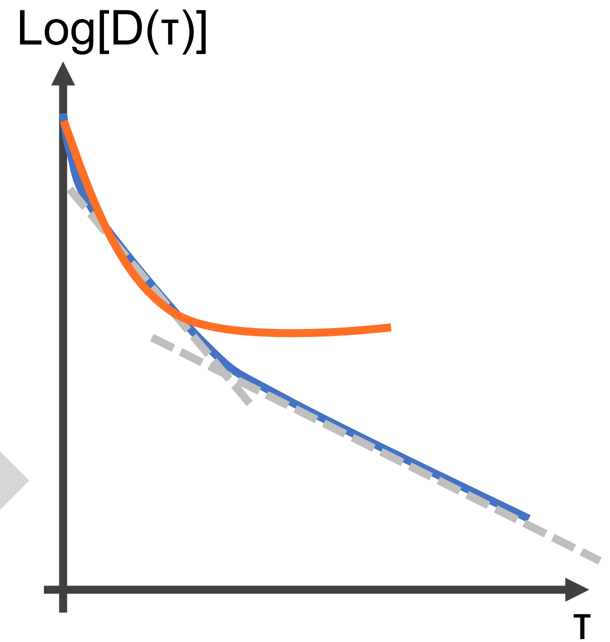
- **Lattice QCD** simulations are similar to a (very) **imperfect detector**



Hadron spectral function



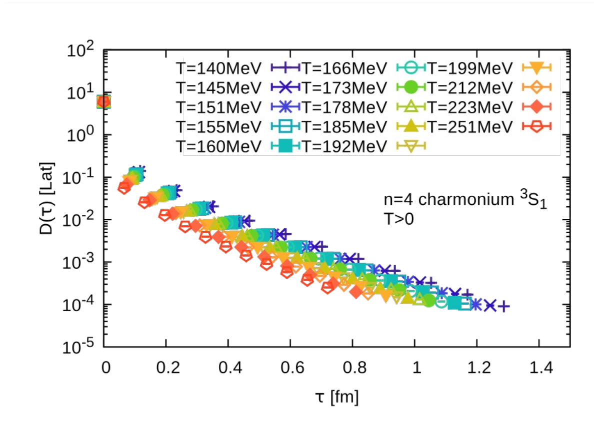
$$D(\tau) = \int_{-2M_Q}^{\infty} d\omega e^{-\tau\omega} \rho(\omega)$$



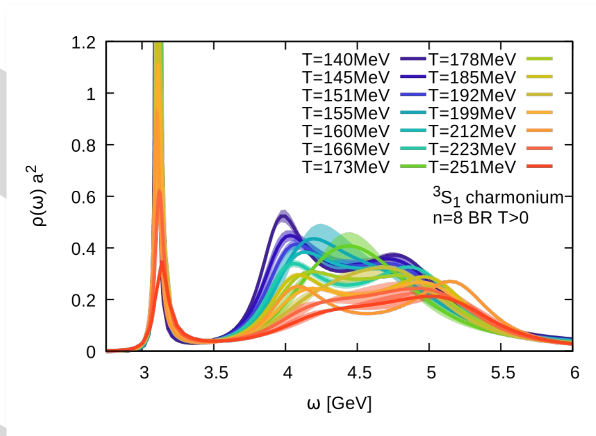
Euclidean time correlation function

- Extraction of spectra ill-posed **unfolding problem**: treated via Bayesian inference

The inverse problem challenge

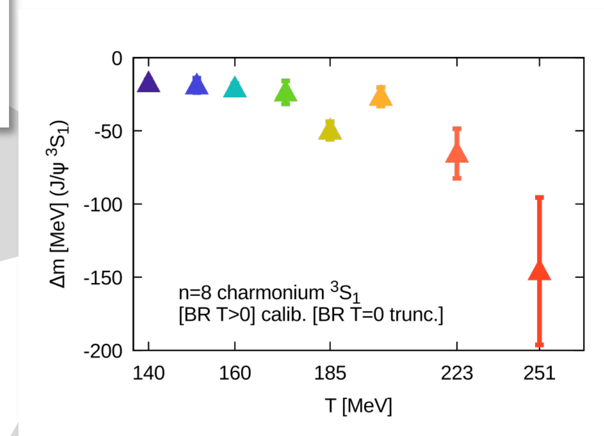


LQCD correlation function
in Euclidean time: **D**



An ill-posed inverse problem

real-time quantity
in momenta: **ρ**

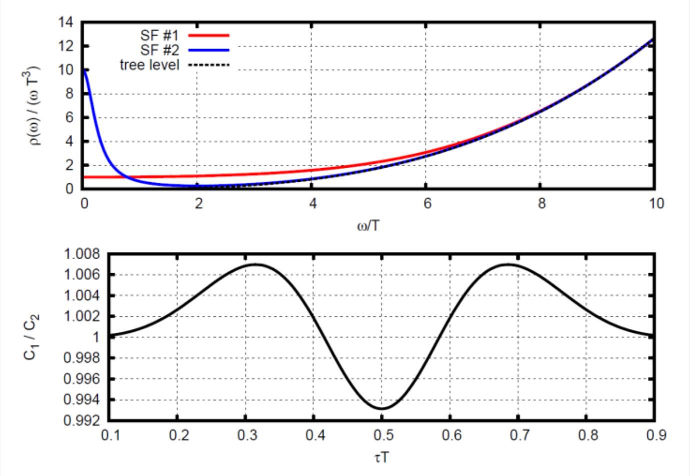


extracted piece of
dynamical information

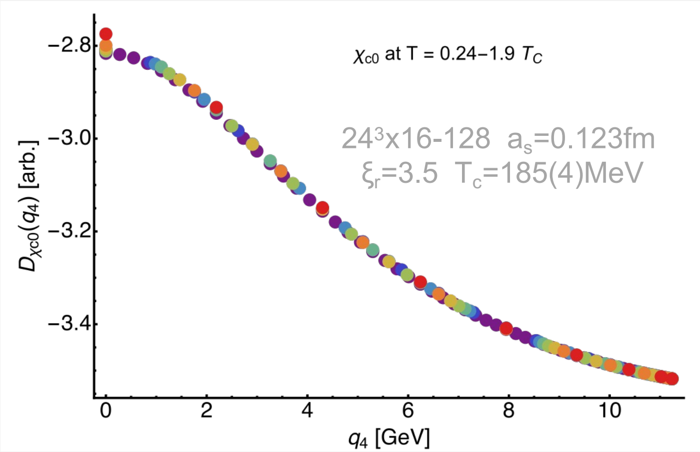
Two intrinsic hurdles

- Invert the spectral representation of the correlator **D**: $\mathbf{D} = K\rho$

$$K^{T=0}_{\omega\tau} = e^{-\omega\tau} \qquad K^{T>0}_{\omega\tau} = \frac{cosh(\omega(\tau-\beta/2))}{sinh(\omega\beta/2)} \qquad K^{PDF}_{x\nu} = cos(x\nu)$$



- Exponential **information loss** due to functional form of the kernel



- **Limited Euclidean range** in T>0 QFT leads to coarse Matsubara frequencies
- Cont. limit: resolve only large w_n behavior

- Goal: extract spectral function accurately using all available knowledge

for the complementary strategy of smeared spectral functions see e.g. Hansen, Meyer, Robaina PRD 96 (2017) 9, 094513

The Bayesian strategy

University of Stavanger

- Extraction of real-time quantities: inversion of ill-conditioned linear transformation

$$D(\tau) = \sum_{l=1}^{N_\omega} \Delta\omega_l \, e^{-\omega_l \tau} \rho_l \qquad D(\tau) = \int_{-2m_Q}^{\infty} d\omega \, e^{-\omega\tau} \rho(\omega)$$

1. N_ω parameters ρ_l >> N_τ datapoints
2. data D_i has finite precision

- Regularize this task using prior information – Bayes introduces prior **P[ρ|I]=exp[S]**

M. Jarrell, J. Gubernatis, Physics Reports 269 (3) (1996)

$$\underbrace{P[\rho|D,I]}_{\text{posterior}} \propto \underbrace{P[D|\rho,I]}_{\text{likelihood}} \underbrace{P[\rho|I]}_{\text{prior}} \xrightarrow[\text{MAP}]{\text{historically}} \left.\frac{\delta P[\rho|D,I]}{\delta\rho}\right|_{\rho=\rho^{\rm BR}} = 0$$

- Prior probability often parametrized by **default model m** and **weight α**

$$\partial_\rho P[\rho|\alpha,m]\big|_{\rho=m} = 0 \qquad \lim_{\alpha\to 0} P[\rho|\alpha,m] = 1$$

α quantifies confidence in the values of m

- Prior information in both choice of **prior functional S** and **default model m**
- **Bayesian continuum limit** $N_\tau \to \infty$ & $\Delta D \to 0$: all version converge to same result

Choice of S influences how "efficiently" one converges to this limit

Implementing the Bayesian strategy

University of Stavanger

- Differences in prior information that is incorporated & how to find extremum.

Tikhonov

$$S_T[m,\alpha] = \alpha \int d\omega (\rho - m)^2$$

Gaussian prior probability, originally hides presence of default model via m=0

A. N. Tikhonov Dokl. Akad. Nauk SSSR 39, 195 (1943)

Maxium Entropy Method (MEM)

Shannon-Jaynes entropy regulator

$$S_{SJ}[m,\alpha] = \alpha \int d\omega \Big(\rho - m - \rho log\Big[\frac{\rho}{m}\Big]\Big)$$

via 4 axioms from 2d image reconstruction

spectral function should behave like a probability density **(coordinate invariance)**

in a two-dimensional setting ρ(x,y) the function can be factorized **(system independence)**

in practice: flat directions & search space is artificially restricted **(Bryan's approach)**

M. Jarrell and J. E. Gubernatis, Phys. Rept. 269, 133 (1996)
M. Asakawa, **T. Hatsuda, Y. Nakahara** P.P.N. Phys. 46, 459 (2001)

Bayesian Reconstruction Method (BR)

Gamma distribution regulator

$$S_{BR} = \alpha \int d\omega \Big(1 - \frac{\rho}{m} + log\Big[\frac{\rho}{m}\Big]\Big)$$

via 4 axioms specifically chosen for 1d reconstruction

end result must not depend on units of spectral function **(scale invariance)**

spectral function shall be smooth (twice differentiable) **(smoothness)**

form of S allows **analytic marginalization of alpha**: true Bayesian treatment for unknown hyperparameter

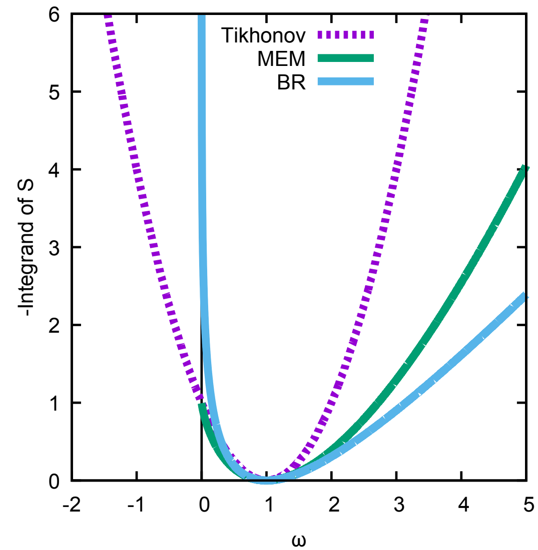
in absence of flat directions improved convergence

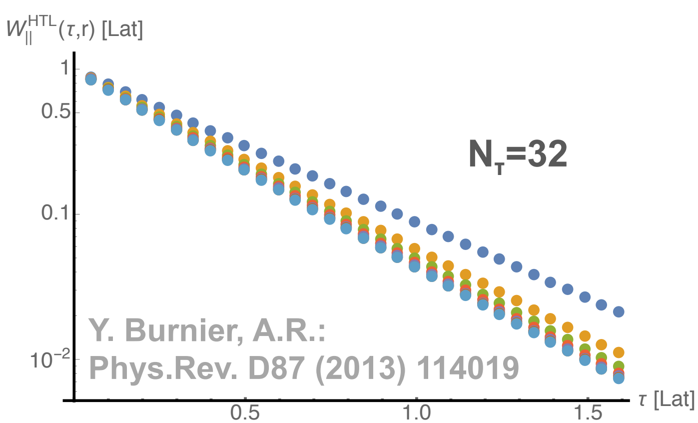
Y.Burnier, A.R. PRL 111 (2013) 18, 182003

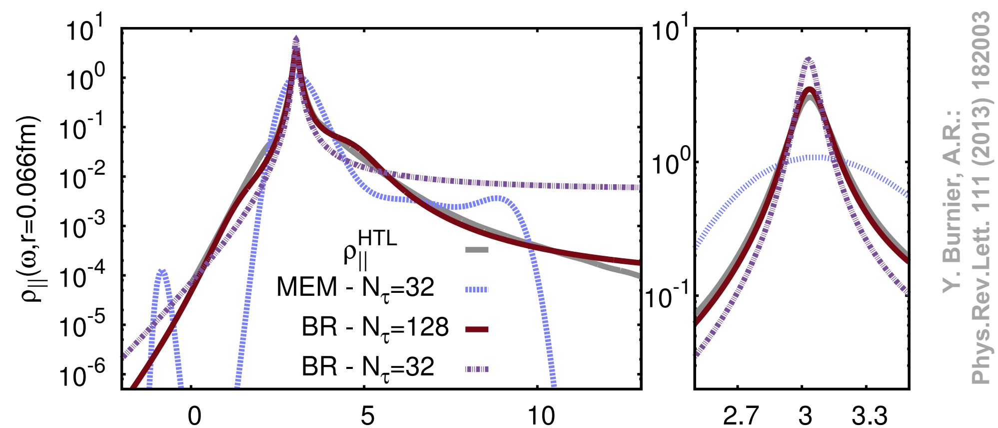
Let the data speak

University of Stavanger

- BR method regulator is the weakest (for ρ>m) on the market, which still provides unique solution
- allows structures encoded in data to manifest more accurately in reconstruction
- weaker in suppressing ringing artefacts, which is relevant when very few input data available
- pertinent example: finite temperature spectral function related to heavy quark bound states



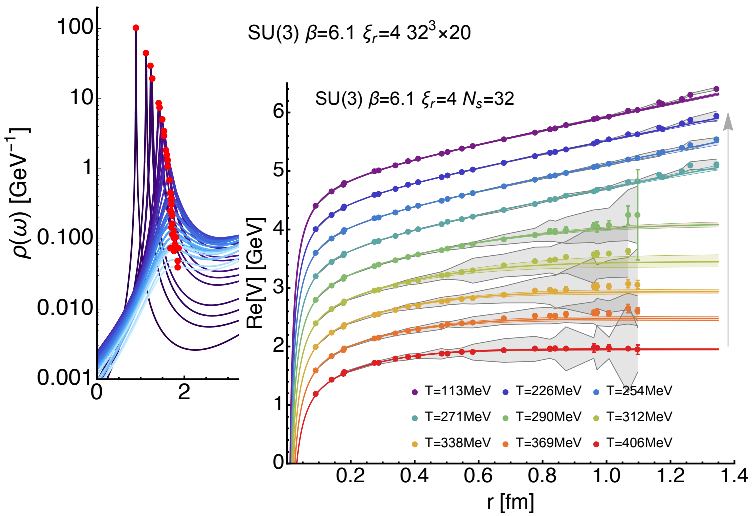




- BR reproduces Lorentzian accurately, MEM always Gaussian-like

Application to strong interactions

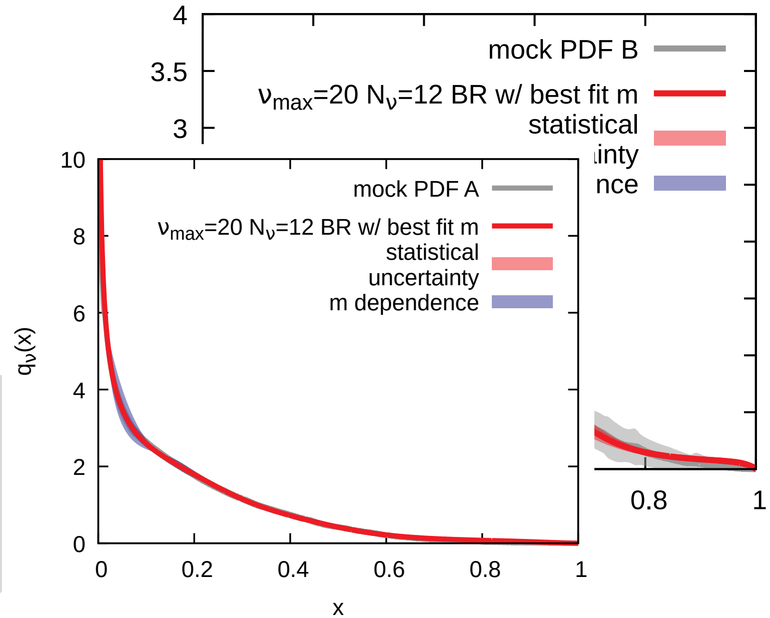
University of Stavanger



T>0 potential between heavy quarks

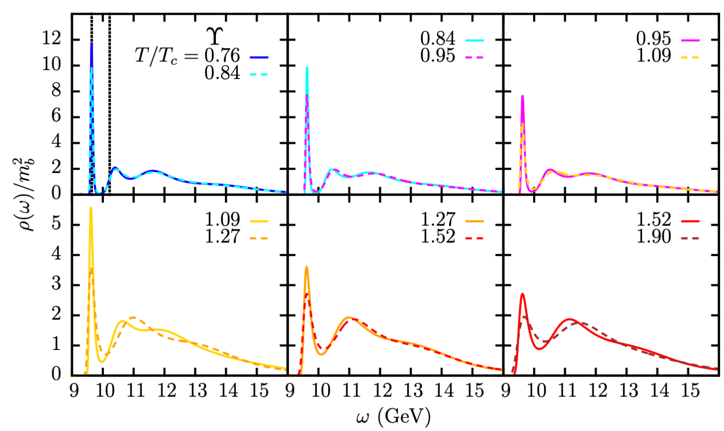
R. Larsen, G. Parkar et.al. PRD 105 (2022) 5, 054513
Y.Burnier, A.R. PRD 95 (2017) 5, 054511
Y. Burnier et.al. JHEP 12 (2015) 101
and work in progress

BR & MEM



Towards PDFs from the lattice

XQCD collaboration PRD 101 (2020) 11, 114503
J. Karpie et.al JHEP 04 (2019) 057
and work in progress x



Spectral functions of T>0 hadrons

FASTSUM collaboration EPJ Web C. 258 (2022) 05011
S. Kim, P. Petreczky, A.R. JHEP 11 (2018) 088
and work in progress

Towards a modern Bayesian approach

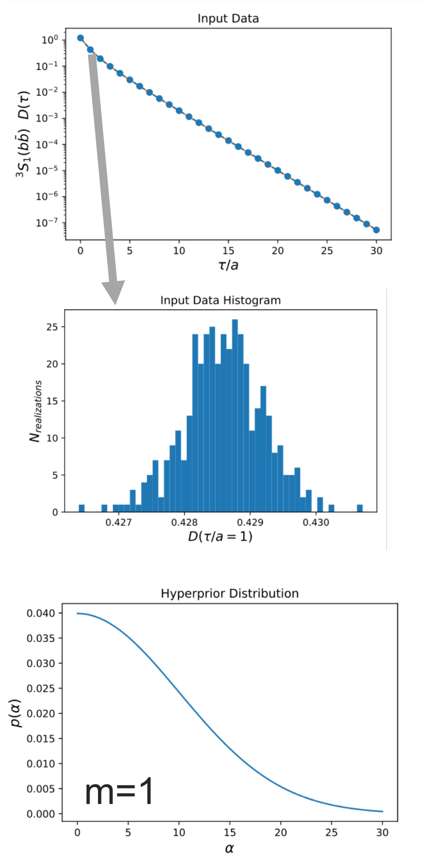
University of Stavanger

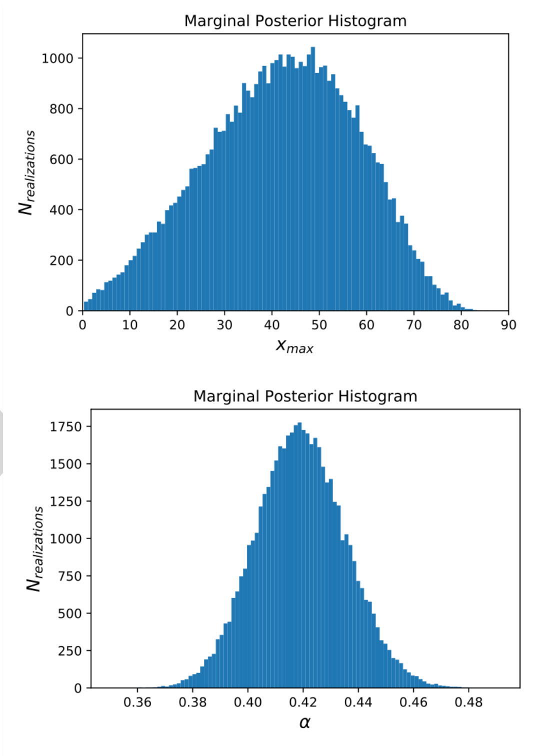
- Elevate to a full Bayesian analysis: sample the posterior via Monte Carlo

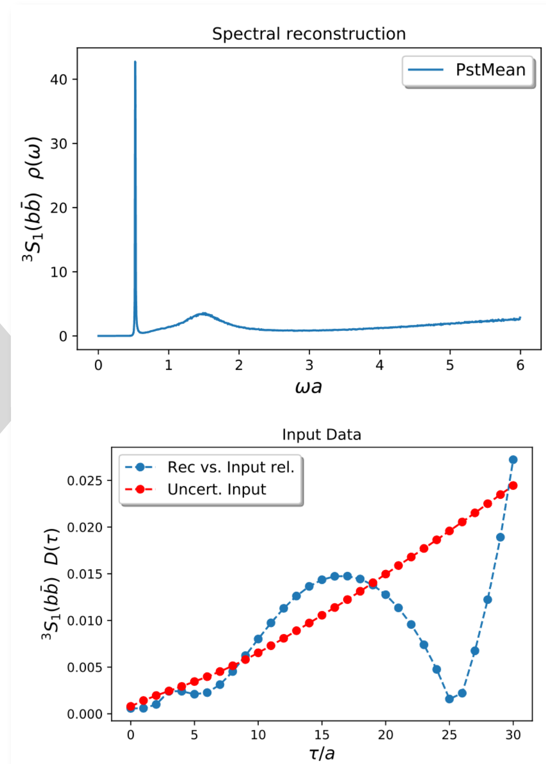
  A.R. PoS Confinement2018 (2018) 026 - 1903.02293

  - Much better control over both statistical and systematic uncertainties
  - Self consistent treatment of hyperparameters α possible

Stan





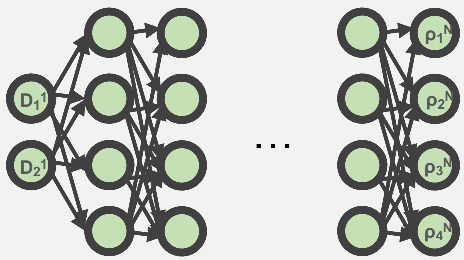


$$P[\rho|D,I] \propto P[D|\rho,I]P[\rho|I]$$

Machine Learning and Bayes

University of Stavanger

Direct Neural Network

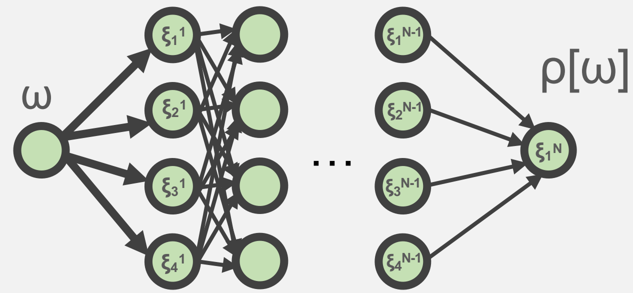


Simpler than in image reconstruction, decoder step known: D=Kρ

supervised Kernel-Ridge or Support Vector Machines regression: differ in loss functional and included regulator

FASTSUM Lattice 2021 2112.02116
L. Kades et. al. PRD 102 (2020) 9, 096001
R. Fournier et.al. PRL 124, 056401 (2020)

NN as basis functions



use as parametrization in supervised or unsupervised setting:
loss includes data + regularization

S. Shi 2201.02564
J. Karpie et.al. JHEP 04 (2019) 057

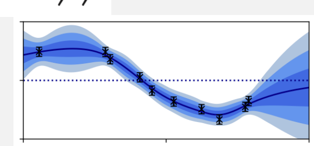
Q: how to encode the regularization in choice of NN geometry?

Gaussian processes

Model observed and predicted data to arise from same Gaussian distribution

$$\begin{pmatrix} f(x) \\ \hat{\mathbf{y}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu(x) \\ \hat{\boldsymbol{\mu}} \end{pmatrix}, \begin{pmatrix} k(x,x) & \hat{\mathbf{k}}^{\mathrm{T}}(x) \\ \hat{\mathbf{k}}(x) & \hat{\mathbf{K}} + \mathbf{C_d} \end{pmatrix} \right)$$

J. Horak et.al. PRD 105, 036014 (2022)



specific choice of correlation matrix, hyperparamters from data

Exploiting Differentiable Programming

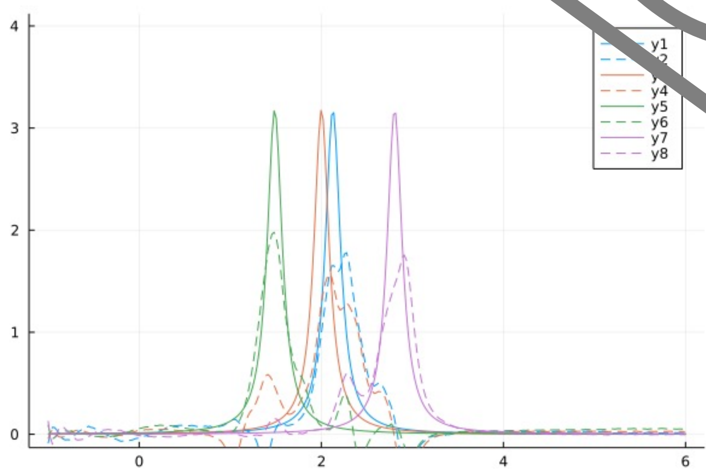
University of Stavanger

- A data driven strategy: identify optimal regulator from mock data training

work in progress with D. Alvestad

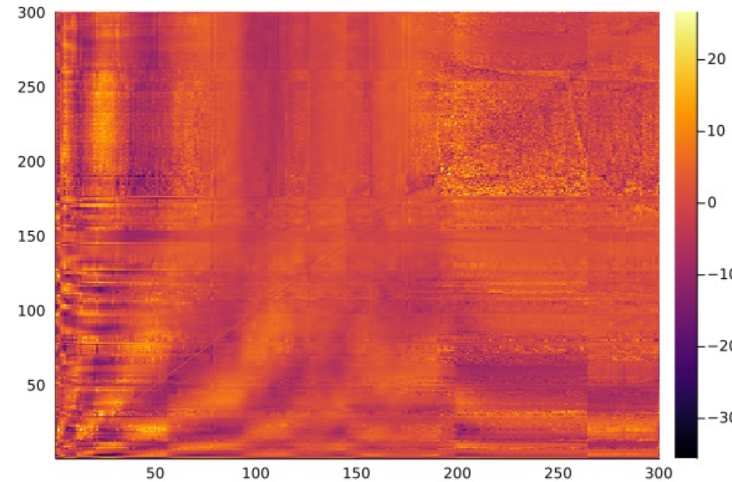
First step: apply to Tikhonov, similar to Gaussian process inference, then BR etc.

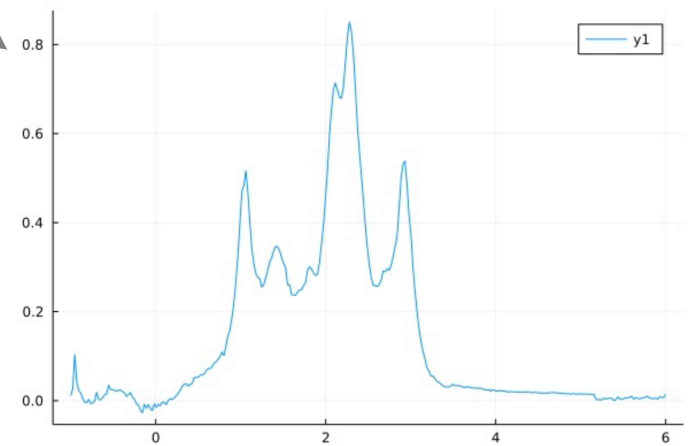
$$P[\rho, I] \sim e^{-(\rho - m)^T \Sigma (\rho - m)}$$



$N_\omega = 300, \omega_{\rm min} = -1, \omega_{\rm max} = 6, N_\tau = 12, \Delta D/D = 10^{-2}$

30 Breit Wigner peaks uniform in ω=[1,3]
with width Γ=0.1





- Does not attempt to solve the inverse problem but reveal structure in regulator.

Summary

- Pressing questions in particle & nuclear physics require nonperturbative QCD input
- Lattice QCD offers a fist-principles approach to QCD beyond weak coupling
- Pertinent information encoded in spectral functions – ill-posed inverse problem
- Active community: application of Bayesian strategies (MEM, BR, Tikhonov) to extract spectral information for T=0 hadron structure or T>0 in-medium physics.
- Bayesian approaches (MEM, BR, Tikhonov) offer access to spectral functions but are limited in specificity of prior knowledge
- Community actively explores both machine learning and differential programming as means to develop more accurate reconstruction strategies

Thank you for your attention

The XVth Quark Confinement and the Hadron Spectrum Conference

XIVth Quark Confinement and the Hadron Spectrum Conference

2022 - University of Stavanger

Track H: Statistical Methods for Physics Analysis in the XXI Century

inviting abstracts for talks and posters (2x500EUR poster prize)

dates: August 1st-6th, 2022

location: University of Stavanger, Norway

web: https://www.ux.uis.no/confxiv/

University of Stavanger

Towards the BR prior probability

- Four axioms for **P[ρ|I]=exp[S]** chosen explicitly for the 1-dim problem at hand

Y.Burnier, A.R. PRL 111 (2013) 18, 182003

Subset independence

$$S \propto \int d\omega\, s[\rho(\omega), m(\omega), \omega]$$

Prior knowledge about different frequency regimes should add in S

Scale invariance

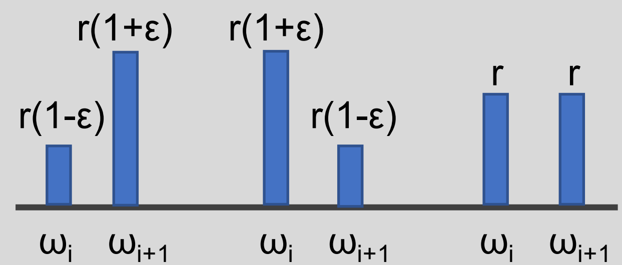
$$S \propto \tilde{\alpha} \int d\omega\, s[\rho(\omega)/m(\omega)]$$

ρ itself is not a probability distribution & ensure that units of ρ do not matter

Smoothness

$$r_i = \rho(\omega_i)/m(\omega_i)$$

Penalize jumps in values between neighboring bins



$$2s[r] - s[r(1+\varepsilon)] - s[r(1-\varepsilon)] \propto \varepsilon^2$$

$$-r^2 s''[r] = C_0$$

Bayesian meaning of m

$$S[r=1]=0,\ S'[r=1]=0,\ S''[r=1]<0$$

$$S[\rho, m, \alpha] = \alpha \int d\omega \Big(1 - \frac{\rho}{m} + ln\Big(\frac{\rho}{m}\Big)\Big)$$

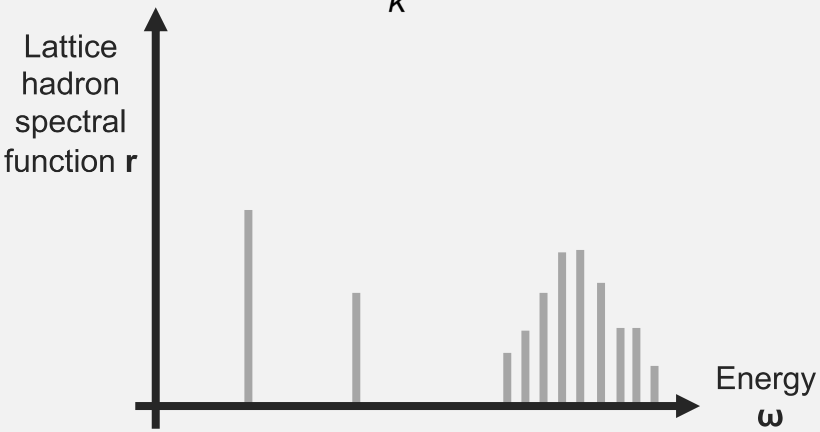
$$P[\rho|m,\alpha] = \Gamma[\alpha+1, \alpha/m]$$

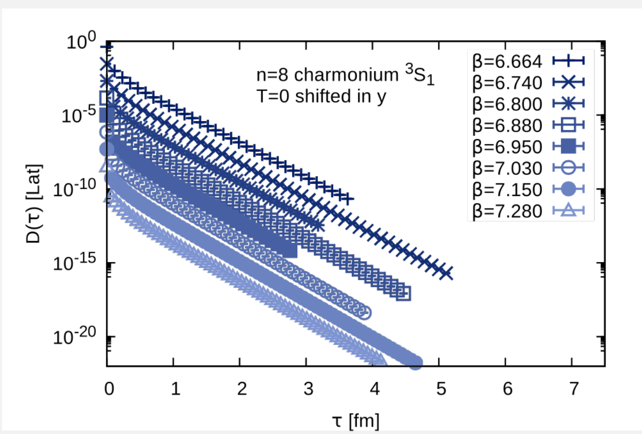
- Important difference to Shannon-Jaynes: S_{BR} diverges around ρ=0 – no flat directions

Beyond exponential fitting

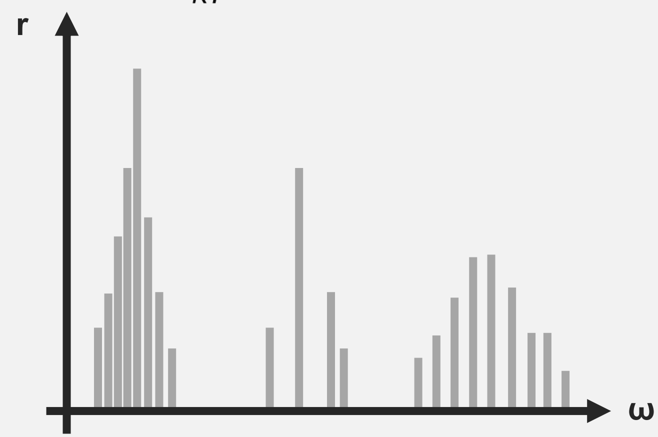
University of Stavanger

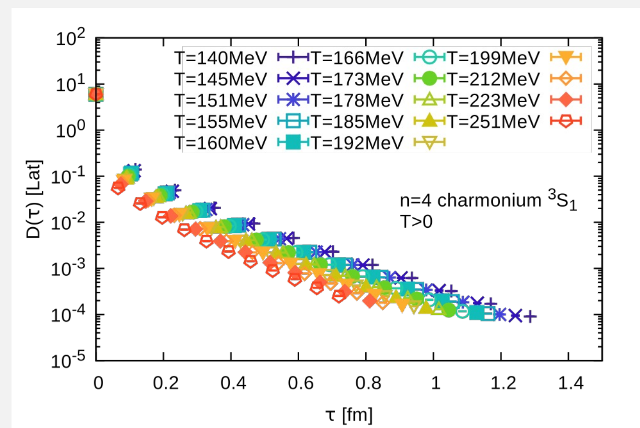
$$\langle MM^\dagger \rangle_{T=0} = \sum_k e^{-E_k \tau} |\langle 0|M|k\rangle|^2$$





$$\langle MM^\dagger \rangle_{T>0} = \sum_{kl} e^{-\beta E_k} e^{-(E_k - E_l)\tau} |\langle k|M|l\rangle|^2$$





- Intricate structure prevents naive use of T=0 strategies (e.g. GEVP)