Inverse Problems and Applications to Lattice.

1. Introduction & Backus - Gilbert
2. Gaussian Processes
3. Neural Networks
4. Training NNs.

1.1 Introduction

Numerous examples of inverse problems in lattice QCD.

Starting from a dataset

central values : $\{ y_I ,\ I = 1 \ldots N_{dat} \}$.

covariance matrix : $C_Y \in \mathbb{R}^{N_{dat} \times N_{dat}}$.

We want to determine $f : M \to \mathbb{R}$

$$y_I = \int_M dx \ C_I(x) \, f(x) .$$

↑ known

Examples: PDF
spectral densities ↗ later!
geoscience, other fields.

Ill-defined problem

↳ not enough info in $\{ y_I \}$ to fully reconstruct f

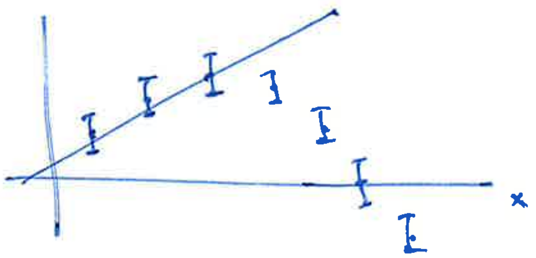
Example

$$C_I(x) = \delta(x - x_I) \quad \Rightarrow \quad y_I = f(x_I) .$$

choose a fixed functional form, e.g. polynomials.

There are infinitely many polynomials going exactly through the data pts.

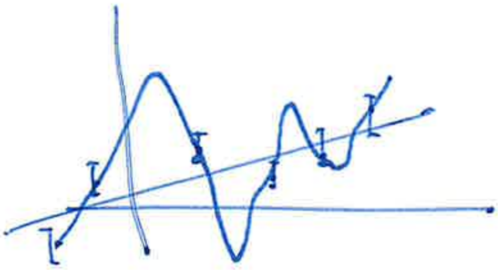
Underfitting



cannot fit the features in the data.

"too few" parameters.

Overfitting



'too many parameters, does not generalize well.
(cfr. plots f. overfitting in Wikipedia, my own plot in the notes).

We want to move away f. the idea of fully determining the solution $f(x)$. f could be a distribution! (later!)

↳ more robust results.

In particular: unbiased extrapolation, clear statements about assumptions. estimate of error!

1.2 Backus-Gilbert

AP Valentine & M Sambridge
Geophys. J. Int. (2020) 220, 1632

$$y_I = \int_M dx \; G_I(x) \, f(x)$$

remember f. later!

$$\Rightarrow \sum_I a_I y_I = \int_M dx \left[\sum_I a_I G_I(x) \right] f(x)$$

$$= \int_M dx \; \Phi_{\hat{a}}(x) \, f(x) \qquad (2.1)$$

Linear combinations of data yield some average property of f.

we require: $\int_M dx \, \Phi_{\hat{a}}(x) = 1$

$\Phi_{\hat{a}}(x)$ is a prob. distribution.

We can choose $\vec{a}$ so that $\Phi_{\vec{a}}$ is "close" to some target averaging function. In order to define "close", we need a distance in the space of fun.

$$\mathcal{L}[u] = \left[\int_M dx\, dx'\, u(x)\, k(x,x')\, u(x')\right]^{1/2}$$

$k(x,x')$ is symmetric, so def. fn.

Then: $d[u,v] = \mathcal{L}[u-v]$.

Estimate $f(x_0)$ using B-G

we need to choose $\vec{a}$ s.t.

$d[\Phi_{\vec{a}}(x), \delta(x-x_0)]^2$ is minimized.

i.e. $\frac{\partial}{\partial a_k}\left(d[\Phi_{\vec{a}}(x), \delta(x-x_0)]^2 \right) = 0$

$$\Rightarrow \frac{\partial}{\partial a_k}\left(\int dx\, dx' \left[\sum_I a_I C_I(x) - \delta(x-x_0)\right] k(x,x') \left[\sum_J a_J C_J(x') - \delta(x'-x_0)\right] \right) = 0$$

$$\Rightarrow \int dx\, dx'\, C_k(x)\, k(x,x') \left[\sum_J a_J C_J(x') - \delta(x'-x_0)\right] = 0$$

$$\Rightarrow \sum_J \underbrace{\left[\int dx\, dx'\, C_k(x)\, k(x,x')\, C_J(x')\right]}_{\hat{W}_{kJ}} a_J - \underbrace{\left[\int dx\, C_k(x)\, k(x,x_0)\right]}_{w_k(x_0)} = 0$$

$$\hat{W}_{kJ}\, a_J - w_k(x_0) = 0.$$

$$\Rightarrow \hat{a}_0 = \hat{W}^{-1} w(x_0)$$

$$\Rightarrow f(x_0) \approx \sum_I a_{0,I}\, y_I = w(x_0)^T \hat{W}^{-1} y$$

assuming $\hat{W}$ is invertible

ie. there exists a unique $\vec{a}_0$.

Regularize the problem

$$\chi^2(x_0,\hat{a}) = d\left[\Phi_{\hat{a}}(x), \delta(x-x_0)\right]^2 + a^T C_y\, a \qquad (4.1)$$

↳ favours "small" a_I

$C_y = \mathrm{diag}\left(\sigma_I^2\right)$

the larger σ_I^2, the smaller a_I

i.e. favour fits w. small errors in the linear combination.

Exercise

Minimize the χ^2 in Eq. (4.1) wrt to a_k.

Deduce that $f(x_0) \approx w(x_0)^T\left(\hat{W} + C_y\right)^{-1} y$

and est. $C_{yT} = \hat{W} + C_y$

NB: $\hat{W}$ & w both depend on the choice of $k(x,x')$.

Freedom in the choice of $k(x,x')$.

B-G method is not limited to point estimators, i.e. $f(x_0)$.
It allows any average property of $f(x)$ to be computed, often with greater accuracy than $f(x_0)$.

1.3 Lattice Spectral Densities & HLT   1903.06476

$$c(t) = \frac{1}{L^3} \sum_{\vec{x}} \langle O(t,\vec{x}) \bar{O}(0) \rangle_L \quad , \quad t>0$$

$$= \int_0^\infty dE \, e^{-tE} \rho_L(E) \qquad \text{in Euclidean time.}$$

where $\rho_L(E) = \frac{1}{L^3} \sum_{\vec{x}} \langle O(0,\vec{x}) \, \delta(E - H_L) \, \bar{O}(0) \rangle_L$

H_L is the Hamiltonian in finite volume.

↳ discrete spectrum.

$$\rho_L(E) = \sum_n \rho_n(L) \, \delta\left(E - E_n(L)\right)$$

is a sum of Dirac's deltas, i.e. a distribution.

Focus on smeared spectral densities.

$$\hat{\rho}_L(\sigma, E_*) = \int dE \, \Delta_\sigma(E_*, E) \rho_L(E).$$

↑ smooth fn. of E_*.

Take the infinite volume limit:

$$\rho(E_*) = \lim_{\sigma \to 0} \lim_{L \to \infty} \hat{\rho}_L(\sigma, E_*)$$

↑ order of the limits!

For practical purposes the limit $\sigma \to 0$ is not strictly needed. Exp. results can be smeared w. the same kernel Δ_σ and compared.

Note that in the B-G procedure discussed above, the point estimator for $f(x)$ is obtained as a smeared version of the unknown fn., with the smearing kernel $\Phi_{\vec{a}}(x)$ determined by the values of a_I after minimization.

Formulation of the problem.

$$y_I = c(t_I) = \int_0^\infty dE \; b_T(t_I, E)\, \rho_L(E)$$

$$T \to \infty, \quad b_\infty(t_I, E) = e^{-t_I E}$$

$$\left.\begin{aligned} I &\to t_I \,;\, t \\ x &\to E \\ C_I(x) &\to b_T(t_I, E) \end{aligned}\right\} \text{dictionary}$$

$$\hookrightarrow \hat{\rho}_L^{BG}(E_*) = \sum_{t=0}^{t_{max}} g_t(E_*) \times c(t+a).$$

$$\eta = e^{-Ea}, \quad b_\infty(t_I, E) = \eta^{n_I}, \quad \text{where } n_I = \frac{t_I}{a}$$

HLT : optimal approx. of a chosen smearing kernel.

e.g.
$$\Delta_\sigma(E_*, E) = \frac{1}{\mathcal{N}} \exp\left\{ -\frac{1}{2} \frac{(E-E_*)^2}{\sigma^2} \right\}$$

$$\mathcal{N} = \int_0^{+\infty} dE \; \exp\left\{ -\frac{(E-E_*)^2}{2\sigma^2} \right\}$$

$$\bar{\Delta}_\sigma(E_*, E) = \sum_{t=0}^{t_{max}} g_t(\lambda, E_*, \sigma)\, b_T(t+a, E).$$

cfr. with Eq. (2.1) : $\sum_I a_I \, c_I(x)$

$$\left| \begin{aligned} I &\to t \\ x &\to E \end{aligned} \right.$$

Find $g_t(\lambda, E_*, \sigma)$ that yield the best approx. of $\Delta_\sigma(E_*, E)$ by minimizing

$$W[\lambda, g] = (1-\lambda)\, A[g] + \lambda \frac{B[g]}{C(0)^2}$$

$$+ \int_0^\infty dE \; \bar{\Delta}_\sigma(E_*, E) = 1 \text{ . constraint.}$$

$$A[g] = \int_{E_0}^\infty dE \left| \bar{\Delta}_\sigma(E_*, E) - \Delta_\sigma(E_*, E) \right|^2$$

$$B[g] = \sum_{t t'} g_t(\lambda, E_*, \sigma) \, (C_Y)_{tt'} \, g_{t'}(\lambda, E_*, \sigma) \, .$$

$$(C_Y)_{tt'} = \mathrm{Cov}\left[C(t), C(t') \right] .$$

Exercise : Show that the solution of the minimization yields

$$g_t(\lambda, E_*, \sigma) = W^{-1}(\lambda)_{tt'} \Big[f_{t'}(\lambda, E_*, \sigma) + R_{t'} \frac{1 - R^T W^{-1}(\lambda) f(\lambda, E_*, \sigma)}{R^T W^{-1}(\lambda) R} \Big] .$$

$$R_t = \int_0^\infty dE \; b_T(t+a, E)$$

$$f_t(\lambda, E_*, \sigma) = (1-\lambda) \int_{E_0}^\infty dE \; b_T(t+a, E) \, \Delta_\sigma(E_*, E) \, .$$

$$W(\lambda)_{tt'} = (1-\lambda) M_{tt'} + \lambda \frac{(C_Y)_{tt'}}{C(0)^2}$$

$$M(E_0)_{tt'} = \int_{E_0}^\infty dE \; b_T(t+a, E) \, b_T(t'+a, E) \, .$$

↑ ill-conditioned matrix

* A comment on A for $T \to \infty$.

$$A[g] = \int_{E_0}^\infty dE \; \Big| \sum_{t=0}^{t_{max}} g_t(\lambda, E_*, \sigma) \, e^{-(t+a)E} - \Delta_\sigma(E^*, E) \Big|^2$$

$$x = e^{-Ea} \Rightarrow dx = -ax \, dE$$

$$= \frac{1}{a} \int_0^{e^{-E_0 a}} \frac{dx}{x} \Big| \sum_{t=0}^{t_{max}} g_t(\lambda, E_*, \sigma) \, x^{-(t+1)} - \Delta_\sigma(E^*, E) \Big|^2$$

↑ increase t_{max} to improve the approximation.

HLT: focus on approximating the smearing kernel $\Delta_\sigma(E^*, E)$.

Monitor the quality of the approximation.

$$\delta_\sigma(E_*, E) = 1 - \frac{\bar{\Delta}_\sigma(E^*, E)}{\Delta_\sigma(E^*, E)}$$

e.g. $\delta_\sigma(E_*, E_*) = 0.05$ as a function of t_{max}.

$\Rightarrow$ value of σ_{min} v. t_{max}.

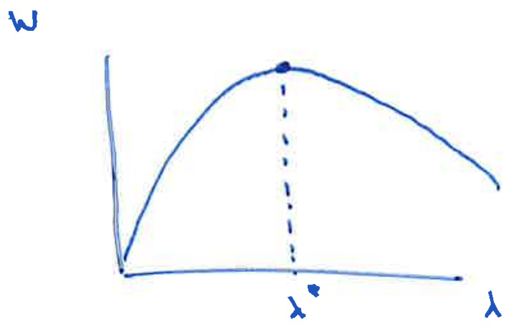
Two regimes of the minimization procedure.

$\lambda \to 0$, $(1-\lambda) M[g]$ dominates

good approximation of $\Delta_\sigma(E_*, E)$.

$\lambda \to 1$, $\lambda B[g]$ dominates

minimize the errors of the reconstructed fn.



$$\lambda^* = \arg\max_{\lambda} W[\lambda, g_t(\lambda, E_*, \sigma)].$$