

Renormalization Group for Signal Analysis

A Different Take on Distances Between Distributions for Uncertainty Quantification

Riccardo Finotello

Digital Twins for Nuclear and Particle Physics 2025

Università degli Studi di Messina, Italy

7 October 2025

The People Behind This Work



Riccardo Finotello

CEA Paris-Saclay 



Vincent Lahoche

CEA Paris-Saclay 



Dine Ousmane Samary

University of Abomey-Calavi 



Harold Erbin

IPhT 



Parham Radpay

CEA Paris-Saclay 

Table of Contents

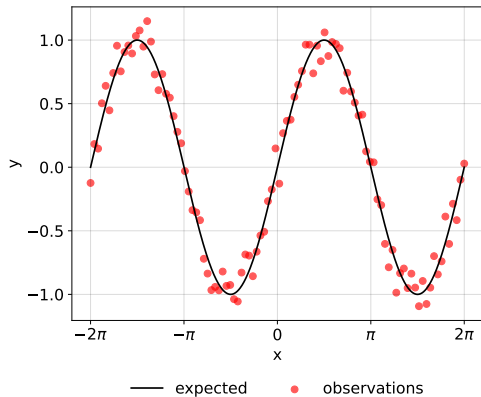
- 1. Data Analysis**
- 2. Effective Field Theory Approach**
- 3. Renormalization Group Approach**
- 4. Anomaly and Signal Detection**
- 5. Conclusions and a Look at the Future**



1 ■ Data Analysis

Two Sides of the Same Coin

Data Distributions and Uncertainty Quantification



Data Analyst Approach

“Physical” process:

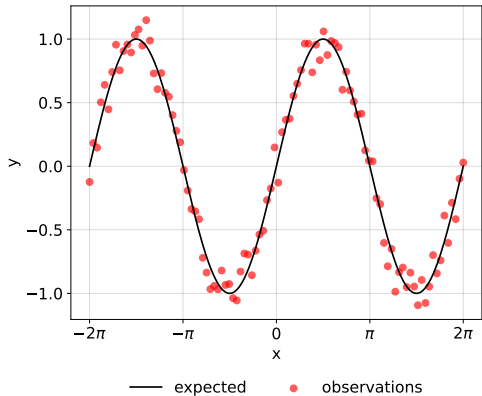
$$y = f(\vec{x}), \quad \vec{x} \in \mathbb{R}^d$$

Observed data:

$$\hat{y} = y + \varepsilon, \quad \varepsilon \sim \mathcal{P}(\vec{\theta})$$

Two Sides of the Same Coin

Data Distributions and Uncertainty Quantification



Data Analyst Approach

“Physical” process:

$$y = f(\vec{x}), \quad \vec{x} \in \mathbb{R}^d$$

Observed data:

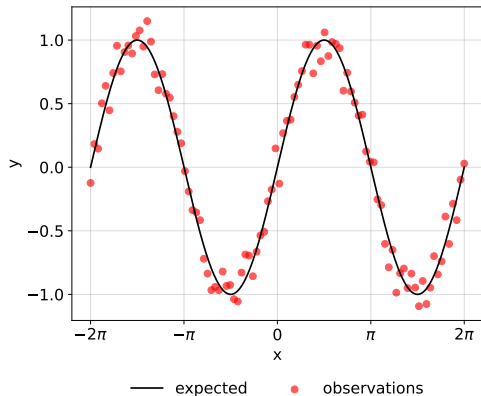
$$\hat{y} = y + \varepsilon, \quad \varepsilon \sim P(\vec{\theta})$$

Goals

- estimate y from $\hat{y}$ (i.e. find $y \approx \tilde{f}_{\Omega}(\vec{x})$);

Two Sides of the Same Coin

Data Distributions and Uncertainty Quantification



Data Analyst Approach

“Physical” process:

$$y = f(\vec{x}), \quad \vec{x} \in \mathbb{R}^d$$

Observed data:

$$\hat{y} = y + \varepsilon, \quad \varepsilon \sim P(\vec{\theta})$$

Goals

- estimate y from $\hat{y}$ (i.e. find $y \approx \tilde{f}_{\Omega}(\vec{x})$);
- quantify uncertainty on y (e.g. the c.i.)

Kriging Your Way Out of Trouble

The Gaussian Process Escape Route

Suppose:

$$y \sim \mathcal{GP}(\mu(\vec{x}), K(\vec{x}, \vec{x}'))$$

Kriging Your Way Out of Trouble

The Gaussian Process Escape Route

Suppose:

$$y \sim \mathcal{GP}(\mu(\vec{x}), K(\vec{x}, \vec{x}'))$$

and find good parameters:

- $\mu(\vec{x})$: **mean function** (consider $\mu(\vec{x}) \equiv 0$)

Kriging Your Way Out of Trouble

The Gaussian Process Escape Route

Suppose:

$$y \sim \mathcal{GP}(\mu(\vec{x}), K(\vec{x}, \vec{x}'))$$

and find good parameters:

- $\mu(\vec{x})$: mean function (consider $\mu(\vec{x}) \equiv 0$)
- $K(\vec{x}, \vec{x}')$: covariance/kernel function

Kriging Your Way Out of Trouble

The Gaussian Process Escape Route

Suppose:

$$y \sim \mathcal{GP}(\mu(\vec{x}), K(\vec{x}, \vec{x}'))$$

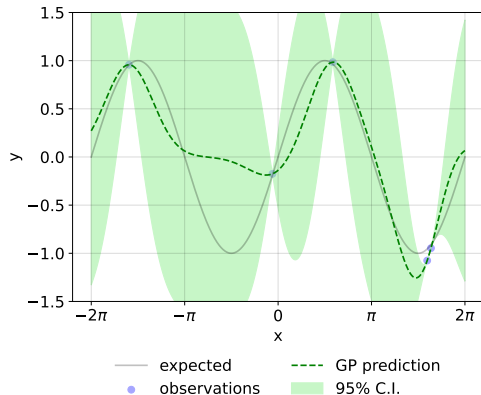
and find good parameters:

- $\mu(\vec{x})$: mean function (consider $\mu(\vec{x}) \equiv 0$)
- $K(\vec{x}, \vec{x}')$: covariance/kernel function

Or in other words, find the *best estimator*:

$$y \sim \mathcal{P}(\mu, K; \mathcal{D}_{\text{train}}) \propto \mathbf{e}^{-\frac{1}{2}(\vec{y}-\mu)^T K^{-1}(\vec{y}-\mu)}$$

where $\mathcal{D}_{\text{train}} = \{(\vec{x}^{(i)}, \hat{y}^{(i)})\}_{i=1}^N$.



Signal Detection for Spiked Models

Failure of the Gaussian Model

GPs are **unstable** in high dimension ($d \gg 1$) and for large datasets ($N \gg 1$):

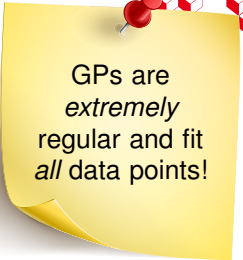
$$y = f(\vec{x}) + \varepsilon, \quad f \sim \mathcal{GP}(0, K), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Signal Detection for Spiked Models

Failure of the Gaussian Model

GPs are **unstable** in high dimension ($d \gg 1$) and for large datasets ($N \gg$

$$y = f(\vec{x}) + \varepsilon, \quad f \sim \mathcal{GP}(0, K), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$



GPs are
extremely
regular and fit
all data points!

Signal Detection for Spiked Models

Failure of the Gaussian Model

GPs are **unstable** in high dimension ($d \gg 1$) and for large datasets ($N \gg$

$$y = f(\vec{x}) + \varepsilon, \quad f \sim \mathcal{GP}(0, K), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

GPs are
extremely
regular and fit
all data points!

Consider a matrix **spiked model**:

$$Y_{ij} = \sum_{r=1}^R \beta_r u_i^{(r)} v_j^{(r)} + Z_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, P,$$

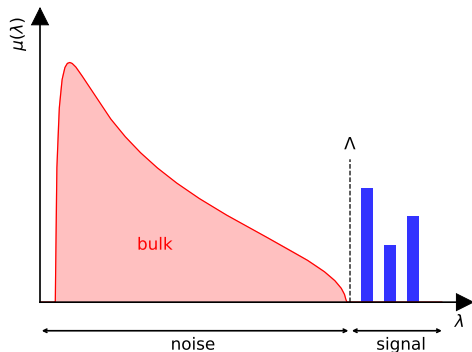
where:

- $Z_{ij} \sim \mathcal{N}(0, \sigma^2)$
- u_i uniform $\forall i = 1, 2, \dots, R$ s.t. $\|\vec{u}\| = 1$
- for simplicity, consider $\beta_1 = \dots = \beta_R = \beta$
- $\mathbb{E}[Y_{ij}] = 0$.

Principal Components and Spiked Models

Weak Signals and Almost Continuous Spectra

If we consider its **PCA**:



$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_P\}$$

s.t. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P \geq 0$.

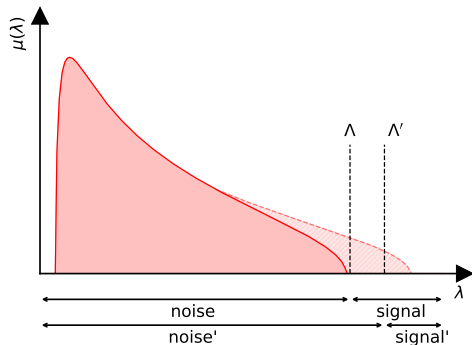
GOAL: find λ_c s.t. $\lambda_i > \lambda_c$ are **signal eigenvalues**

■ N small / β **strong** $\rightarrow$ *easily* isolate spikes

Principal Components and Spiked Models

Weak Signals and Almost Continuous Spectra

If we consider its **PCA**:



$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_P\}$$

s.t. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P \geq 0$.

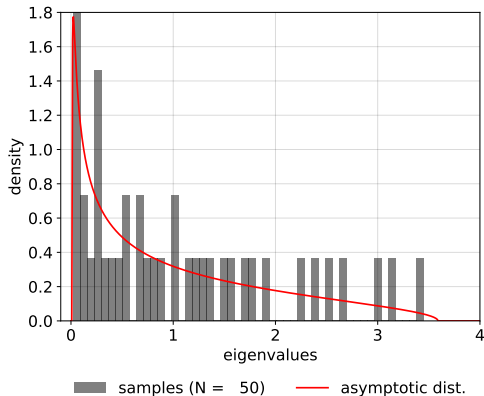
GOAL: find λ_c s.t. $\lambda_i > \lambda_c$ are **signal eigenvalues**

- N small / β **strong** $\rightarrow$ *easily* isolate spikes
- N large / β **weak** $\rightarrow$ *where to cut?*

Principal Components and Spiked Models

Weak Signals and Almost Continuous Spectra

If we consider its **PCA**:



$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_P\}$$

s.t. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P \geq 0$.

GOAL: find λ_c s.t. $\lambda_i > \lambda_c$ are **signal eigenvalues**

- N small / β **strong** $\rightarrow$ *easily* isolate spikes
- N large / β **weak** $\rightarrow$ *where to cut?*

Nearly continuous spectra

Main issue related to *large* datasets where **signal is heavily diluted** in noise.



2. Effective Field Theory Approach

Nearly Continuous Spectra

Towards Data-driven Field Theory

Is there a way to **deal with nearly continuous spectra**?

First attempt:

$$\{\vec{x}_i \in \mathbb{R}^d\}_{i=1}^N \longrightarrow \phi(\vec{x}) = \frac{1}{\sqrt{2\pi}} \int d^d p \tilde{\phi}(\vec{p}) e^{-i\vec{p} \cdot \vec{x}}$$

$$\sum_{i=1}^N y_i K_{ij}^{-1} y_j \longrightarrow S[\phi] = \frac{1}{2} \int d^d p \phi(-\vec{p}) K^{-1}(p^2) \phi(\vec{p}) \quad (p^2 = \|\vec{p}\|^2)$$

data-driven

EFT

$$C(p^2) = \mathbb{E}_{P[\phi]} [\phi(\vec{p}) \phi(-\vec{p})] = K(p^2),$$

$$P[\phi] = e^{-S[\phi]}$$

Nearly Continuous Spectra

Towards Data-driven Field Theory

Is there a way to **deal with nearly continuous spectra**?

First attempt:

$$\{\vec{x}_i \in \mathbb{R}^d\}_{i=1}^N \longrightarrow \phi(\vec{x}) = \frac{1}{\sqrt{2\pi}} \int d^d p \tilde{\phi}(\vec{p}) e^{-i\vec{p} \cdot \vec{x}}$$

$$\sum_{i=1}^N y_i K_{ij}^{-1} y_j \longrightarrow S[\phi] = \frac{1}{2} \int d^d p \phi(-\vec{p}) K^{-1}(p^2) \phi(\vec{p}) \quad (p^2 = \|\vec{p}\|^2)$$

data-driven

Still waaaaaaay too Gaussian Process-like!

$$P[\phi] = e^{-S[\phi]}$$

$$\mathbb{E}[\phi(p_1)\phi(p_2)\phi(p_3)\phi(p_4) \dots] = \sum_{\sigma(a,b,c,d,\dots)} \mathbb{E}[\phi(p_a)\phi(p_b)] \mathbb{E}[\phi(p_c)\phi(p_d)] \dots$$

Nearly Continuous Spectra

Effective Field Theory for Data Analysis

A bit further down the **EFT** rabbit hole (assume $\mathbb{Z}_2$ symmetry):

$$S[\phi] = \frac{1}{2} \int d^d p \, \phi(-\vec{p}) \tilde{K}^{-1}(p^2) \phi(\vec{p}) + \sum_{n=2}^{\infty} \frac{u_{2n}}{(2n)!} \int \prod_{i=1}^{2n} d^d p_i \, \phi(p_1) \cdots \phi(p_{2n}) \delta\left(\sum_{i=1}^n p_i\right)$$

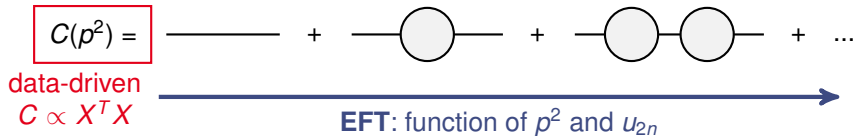
$$C(p^2) = \text{---} + \text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} \bigcirc \text{---} + \dots$$

Nearly Continuous Spectra

Effective Field Theory for Data Analysis

A bit further down the **EFT** rabbit hole (assume $\mathbb{Z}_2$ symmetry):

$$S[\phi] = \frac{1}{2} \int d^d p \, \phi(-\vec{p}) \tilde{K}^{-1}(p^2) \phi(\vec{p}) + \sum_{n=2}^{\infty} \frac{u_{2n}}{(2n)!} \int \prod_{i=1}^{2n} d^d p_i \, \phi(p_1) \cdots \phi(p_{2n}) \delta\left(\sum_{i=1}^n p_i\right)$$

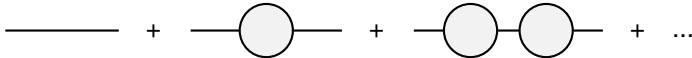


Nearly Continuous Spectra

Effective Field Theory for Data Analysis

A bit further down the **EFT** rabbit hole (assume $\mathbb{Z}_2$ symmetry):

$$S[\phi] = \frac{1}{2} \int d^d p \, \phi(-\vec{p}) \tilde{K}^{-1}(p^2) \phi(\vec{p}) + \sum_{n=2}^{\infty} \frac{u_{2n}}{(2n)!} \int \prod_{i=1}^{2n} d^d p_i \, \phi(p_1) \cdots \phi(p_{2n}) \delta\left(\sum_{i=1}^n p_i\right)$$

$C(p^2) =$ 

data-driven $C \propto X^T X$

EFT: function of p^2 and u_{2n}

Data Analysis EFT: ?

Nearly Continuous Spectra

Effective Field Theory for Data Analysis

How to include **nearly continuous spectra** in the field theory?

$$d^d p \longrightarrow dp^2 \rho(p^2)$$

which for “regular” field theory $\rho(p^2) \propto (p^2)^{\frac{d-2}{2}}$.

Nearly Continuous Spectra

Effective Field Theory for Data Analysis

How to include **nearly continuous spectra** in the field theory?

$$d^d p \longrightarrow dp^2 \rho(p^2)$$

which for “regular” field theory $\rho(p^2) \propto (p^2)^{\frac{d-2}{2}}$.

In **data analysis**, spectra are **data driven**:

$$C \propto X^T X \in \mathbb{R}^{P \times P} \rightarrow \left\{ \lambda_i = \frac{1}{p_i^2 + m^2} \right\}_{i=1}^P$$

Nearly Continuous Spectra

Effective Field Theory for Data Analysis

How to include **nearly continuous spectra** in the field theory?

$$d^d p \longrightarrow dp^2 \rho(p^2)$$

which for “regular” field theory $\rho(p^2) \propto (p^2)^{\frac{d-2}{2}}$.

In **data analysis**, spectra are **data driven**:

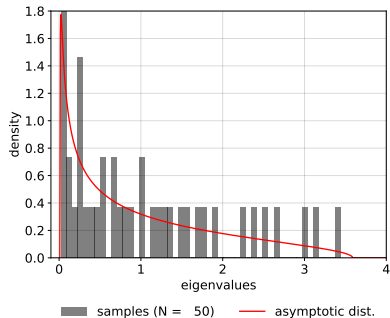
$$C \propto X^T X \in \mathbb{R}^{P \times P} \rightarrow \left\{ \lambda_i = \frac{1}{p_i^2 + m^2} \right\}_{i=1}^P$$

For **purely noisy** spectra ($X \sim \mathcal{N}(0, 1)$) $\rightarrow$

Marchenko-Pastur:

$$\{\lambda_i\}_{i=1}^P \xrightarrow{i.d.} \mu(\lambda) = \frac{1}{2\pi\sigma^2\gamma} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \mathbb{1}_{[\lambda_-, \lambda_+]},$$

where $\lambda_{\pm} = \sigma^2(1 \pm \sqrt{\gamma})^2$ and $\gamma = P/N \leq 1$.



Nearly Continuous Spectra

Effective Field Theory for Data Analysis

How to include **nearly continuous spectra** in the field theory?

$$d^d p \longrightarrow dp^2 \rho(p^2)$$

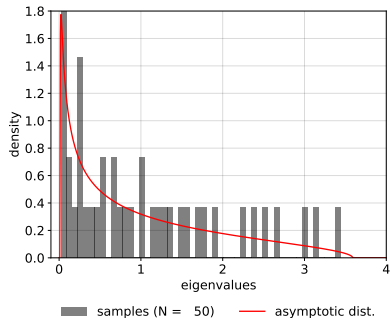
which for “regular” field theory $\rho(p^2) \propto (p^2)^{\frac{d-2}{2}}$.

In **data analysis**, spectra are **data driven**:

$$C \propto X^T X \in \mathbb{R}^{P \times P} \rightarrow \left\{ \lambda_i = \frac{1}{p_i^2 + m^2} \right\}_{i=1}^P$$

Operatively ($m^2 = \lambda_1^{-1}$):

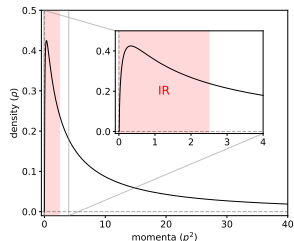
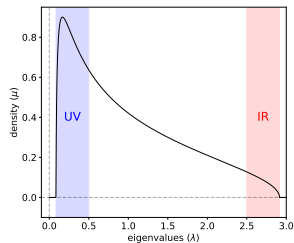
$$\{\lambda_i\}_{i=1}^P \xrightarrow{\text{hist.}} \tilde{\mu}(\lambda) \rightarrow \tilde{\rho}(p^2) = \tilde{\mu} \left(\frac{1}{p^2 + m^2} \right) \frac{1}{(p^2 + m^2)^2}$$



Nearly Continuous Spectra

Anomaly and Signal Detection

Lahoche *et al.* (2022)



Data covariance naturally defines a **spectrum**:

IR high eigenvalues $\rightarrow$ **low** momenta

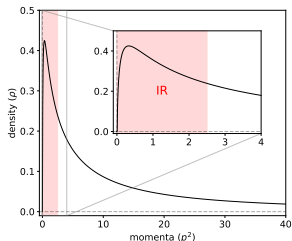
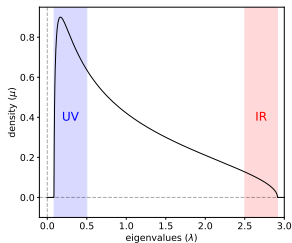
UV low eigenvalues $\rightarrow$ **high** momenta

We can thus look for **anomalies** (small perturbations) in the **IR** part of the spectrum:

Nearly Continuous Spectra

Anomaly and Signal Detection

Lahoche *et al.* (2022)



Data covariance naturally defines a **spectrum**:

IR high eigenvalues $\rightarrow$ **low** momenta

UV low eigenvalues $\rightarrow$ **high** momenta

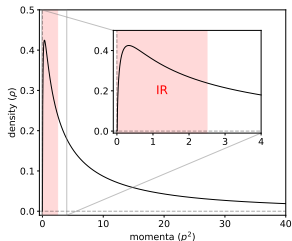
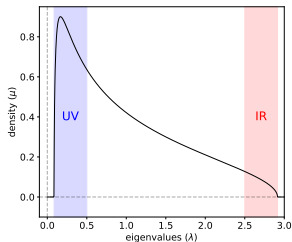
We can thus look for **anomalies** (small perturbations) in the **IR** part of the spectrum:

$$\lambda_i = \tilde{\lambda}_i + \varepsilon \Theta_{ii} + \varepsilon^2 \sum_{j \neq i} \frac{\Theta_{ij}^2}{\tilde{\lambda}_i - \tilde{\lambda}_j} + O(\varepsilon^3)$$

Nearly Continuous Spectra

Anomaly and Signal Detection

Lahoche *et al.* (2022)



Data covariance naturally defines a **spectrum**:

IR high eigenvalues $\rightarrow$ **low** momenta

UV low eigenvalues $\rightarrow$ **high** momenta

We can thus look for **anomalies** (small perturbations) in the **IR** part of the spectrum:

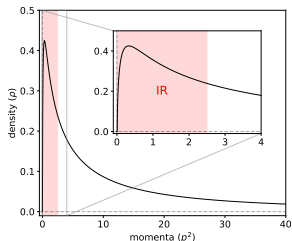
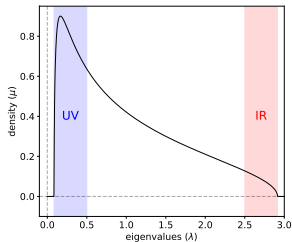
$$\Theta_{ij} \Big|_{u_4} = -\frac{\delta_{ij}}{\varepsilon} \tilde{\lambda}_i^2 \frac{u_4}{4} \overbrace{\left(\frac{1}{N} \sum_{i=1}^P \lambda_i \right)}^{\mathbb{E}[\lambda]} + O(u_4^2)$$

$$\xrightarrow{i.d.} -\frac{\delta_{ij}}{\varepsilon} \tilde{\lambda}_i^2 \frac{u_4}{4} \int d\lambda \lambda \mu(\lambda) + O(u_4^2)$$

Nearly Continuous Spectra

Anomaly and Signal Detection

Lahoche *et al.* (2022)



Data covariance naturally defines a **spectrum**:

IR high eigenvalues → **low** momenta

UV low eigenvalues → **high** momenta

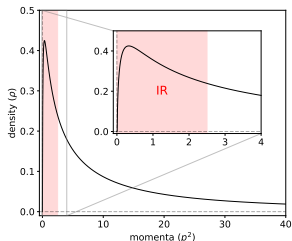
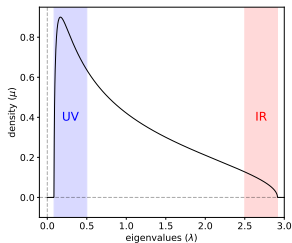
We can thus look for **anomalies** (small perturbations) in the **IR** part of the spectrum:

$$\begin{aligned} \frac{1}{p^2 + m^2} \Big|_{u_4} &= \frac{1}{\tilde{p}^2 + \tilde{m}^2} - \frac{u_4}{4} \frac{1}{\tilde{p}^2 + \tilde{m}^2} \overbrace{\int \frac{dk^2 \rho(k^2)}{k^2 + m^2}}^{\text{usual loop momenta}} \frac{1}{\tilde{p}^2 + \tilde{m}^2} + \dots \\ &= \frac{1}{\tilde{p}^2 + \tilde{m}^2 + \frac{u_4}{4} \int \frac{dk^2 \rho(k^2)}{k^2 + m^2}} \end{aligned}$$

Nearly Continuous Spectra

Anomaly and Signal Detection

Lahoche *et al.* (2022)



Data covariance naturally defines a **spectrum**:

IR high eigenvalues → **low** momenta

UV low eigenvalues → **high** momenta

We can thus look for **anomalies** (small perturbations) in the **IR** part of the spectrum:

Data	EFT (general)	EFT (“strict” LPA)
p^2	$= \int \tilde{k}^2$	$\simeq \int \tilde{k}^2$
m^2	$= \tilde{m}^2 - \Sigma(p^2)$	$\simeq \tilde{m}^2 - \Sigma(0)$

anomalies ↔ **mass (IR) corrections**

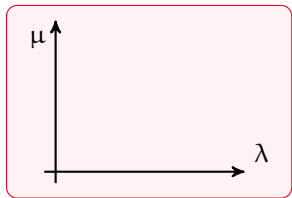
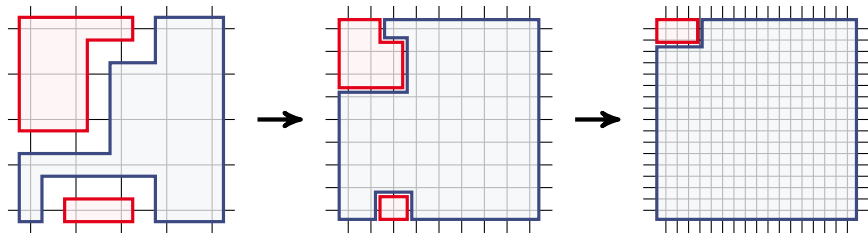
LPA = Local Potential Approximation



3. Renormalization Group Approach

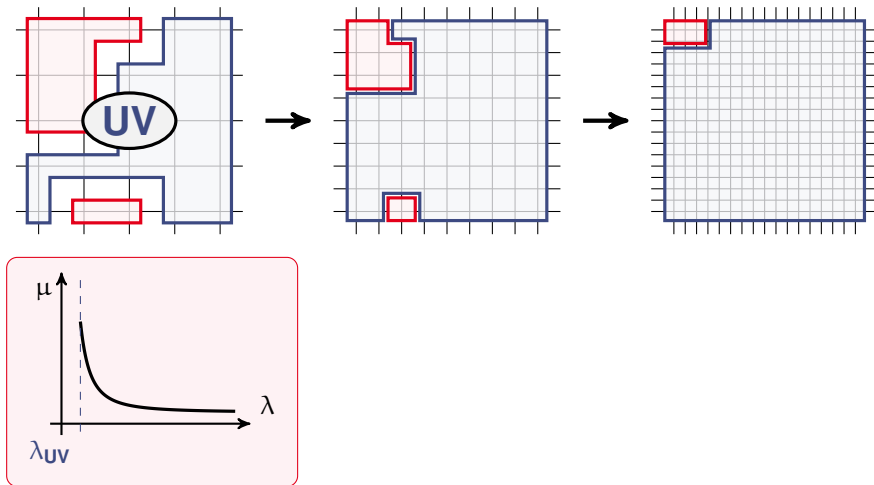
Introducing the Main Star

The Role of the Renormalization Group



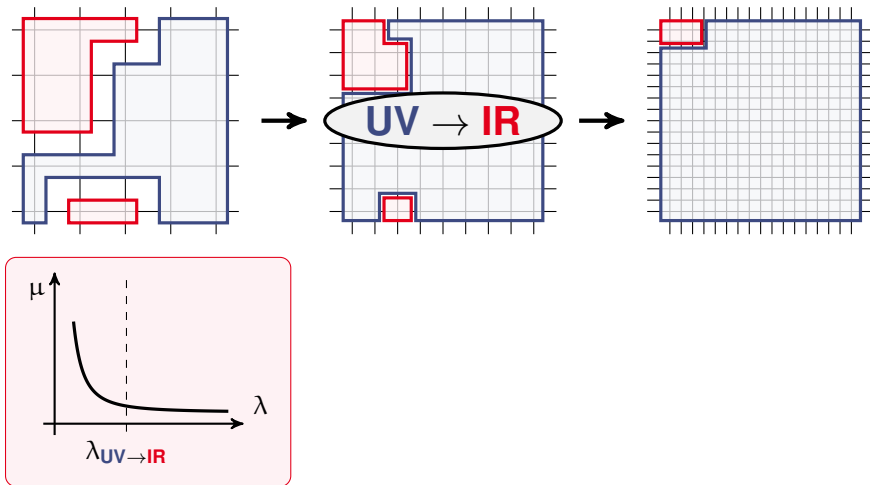
Introducing the Main Star

The Role of the Renormalization Group



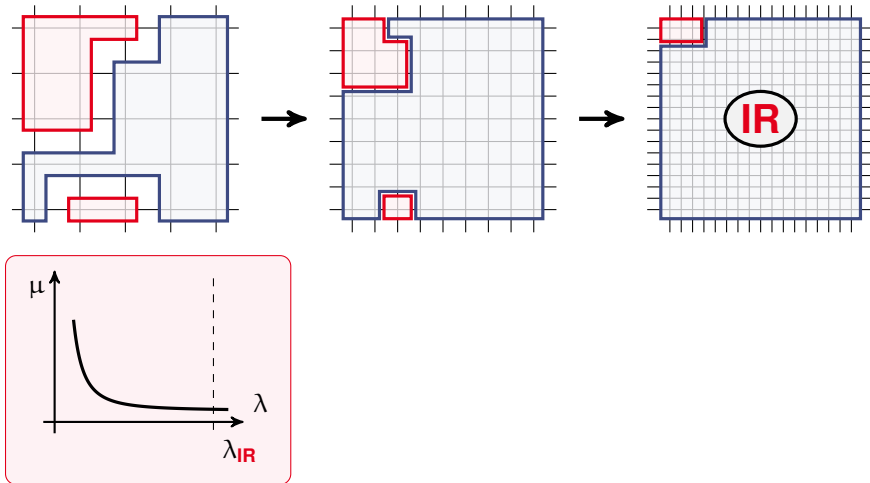
Introducing the Main Star

The Role of the Renormalization Group



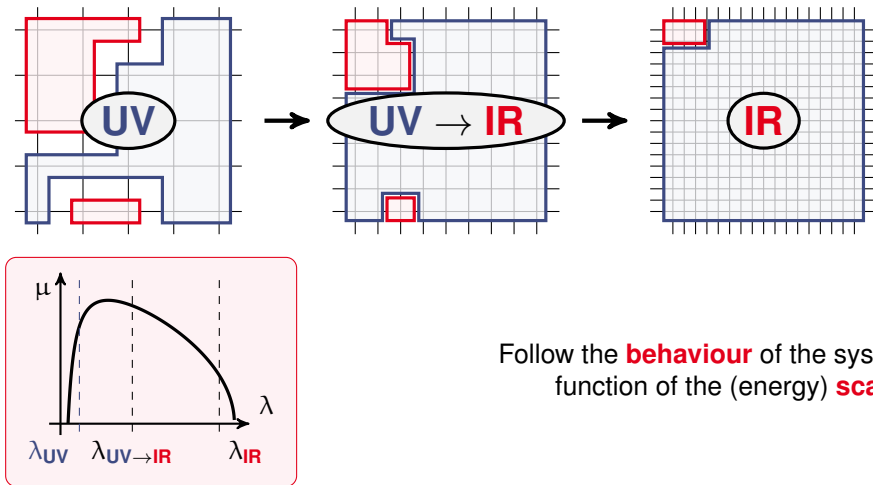
Introducing the Main Star

The Role of the Renormalization Group



Introducing the Main Star

The Role of the Renormalization Group

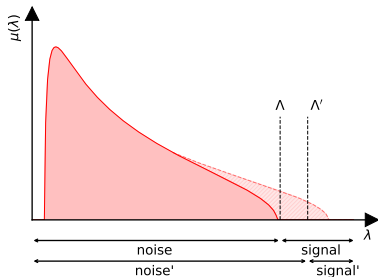


Follow the **behaviour** of the system as a function of the (energy) **scale**!

Introducing the Main Star

The Role of the Renormalization Group

Why **renormalization**?



Traditional **EFT**:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} (p^2)^{\frac{d-2}{2}}$$

Purely **noisy spectra**:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} (p^2)^{\frac{1}{2}} \quad \text{as if} \quad d = 3$$

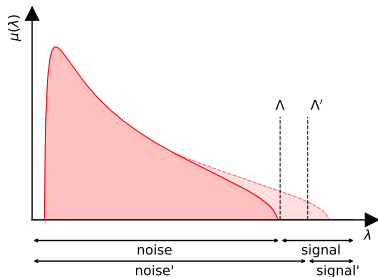
Nearly continuous spectra:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} \text{not a power law} \quad \text{as if} \quad d \neq 3$$

Introducing the Main Star

The Role of the Renormalization Group

Why **renormalization**?



Traditional **EFT**:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} (p^2)^{\frac{d-2}{2}}$$

Purely **noisy spectra**:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} (p^2)^{\frac{1}{2}} \quad \text{as if} \quad d = 3$$

Nearly continuous spectra:

$$\rho(p^2) \xrightarrow{p^2 \rightarrow 0} \text{not a power law} \quad \text{as if} \quad d \neq 3$$

We can follow the behaviour of **couplings** as functions of **momenta**!

Functional Renormalization Group

The Data Analysis Use Case

Data **covariance**

Morris *et al.* (1994)

Wetterich *et al.* (2002)

Delamotte *et al.* (2007)

$$C = \frac{1}{N} X^T X \in \mathbb{R}^{P \times P}$$

is *inherently* an “effective theory” (bottom-up) $\rightarrow$ **effective average action**:

$$S[\phi] \rightarrow S[\phi] + \Delta S_{k^2}[\phi] \quad \text{in} \quad Z_{k^2}[J] = \int \mathcal{D}\phi \, e^{-S[\phi] - \Delta S_{k^2}[\phi] + \int_p J(-p)\phi(p)}$$

Given a **good truncation** (e.g. LPA, where $M(p)$ is the **classical field**):

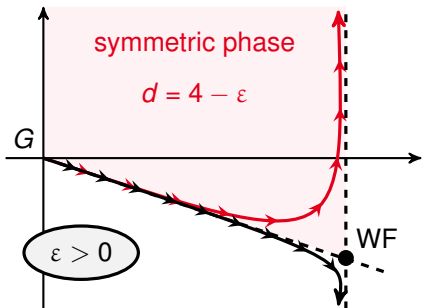
$$\Gamma_{k^2}[M] = \frac{1}{2} \oint_p M(-p) (p^2 + m^2(k^2)) M(p) + \mathcal{U}_{k^2}[M] \rightarrow \begin{cases} \Gamma[M] & \text{if } k^2 \rightarrow 0 \\ S[M] & \text{if } k^2 \rightarrow \infty \end{cases}$$

Functional Renormalization Group

Flow Equations and Data-driven Geometry

The **flow equation** for the **effective potential** involves:

$$L(k) = \int_0^k dp \, p \, \rho(p^2) = \frac{1}{2} \int_0^{k^2} dp^2 \, \rho(p^2) = \frac{1}{2} \text{CDF}[\rho](k^2)$$



FRG for Data Analysis

The **data-driven** approach modifies the underlying **geometry** of “spacetime” by the **empirical quantiles** of the spectrum!

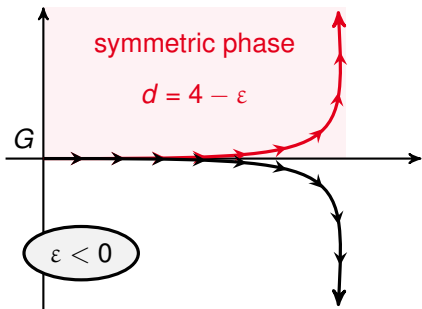
The dimensions of spacetime are **not fixed** anymore!

Functional Renormalization Group

Flow Equations and Data-driven Geometry

The **flow equation** for the **effective potential** involves:

$$L(k) = \int_0^k dp \, p \, \rho(p^2) = \frac{1}{2} \int_0^{k^2} dp^2 \, \rho(p^2) = \frac{1}{2} \text{CDF}[\rho](k^2)$$



FRG for Data Analysis

The **data-driven** approach modifies the underlying **geometry** of “spacetime” by the **empirical quantiles** of the spectrum!

The dimensions of spacetime are **not fixed** anymore!

Running Couplings

Flow Equations and Canonical Dimensions

Lahoche *et al.* (2022)

Standard **field theory** $\rightarrow$ CDF is a **power law** $\rightarrow$ rescale couplings to get **autonomous** system

Running Couplings

Flow Equations and Canonical Dimensions

Lahoche *et al.* (2022)

Standard **field theory** $\rightarrow$ CDF is a **power law** $\rightarrow$ rescale couplings to get **autonomous** system

Scale Dependence and Canonical Dimensions

We can hide the scale in the **linear term** of the flow equations (let $t = \ln k$ and $\tau = \ln L$):

$$u_2 = k^2 \bar{g}_2 \Rightarrow \dim_{u_2}(k^2) = 2t'$$
$$u_4 = \frac{k^2}{\rho(k^2)} \frac{1}{(t')^2} \bar{g}_4 \Rightarrow \dim_{u_4}(k^2) = -2 \left(\frac{t''}{t'} + t' \left(\frac{1}{2} \frac{d \ln \rho}{dt} - 1 \right) \right)$$

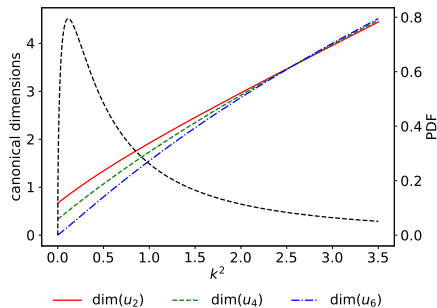
where $X' = \frac{dX}{d\tau}$. Generically: $\dim_{u_{2n}}(k^2) = (2 - n) \dim_{u_2}(k^2) + (n - 1) \dim_{u_4}(k^2)$.



4. ■ Anomaly and Signal Detection

Marchenko-Pastur Universality Class

Canonical Dimensions and Signal Detection



For the **analytical model** in the **IR** region:

- u_2 and u_4 are **relevant couplings**
- u_6 is asymptotically **marginal**
- $u_{2n \geq 8}$ are **irrelevant**

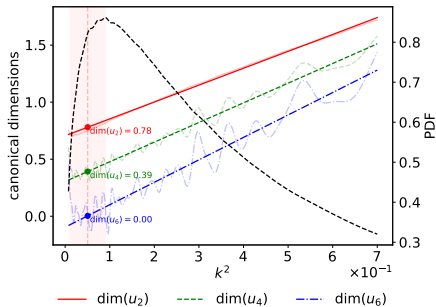
The **Gaussian theory** is **not adapted** to describe the **IR sector**!

For **nearly continuous spectra**, we need **at least** a ϕ^4 **theory**.

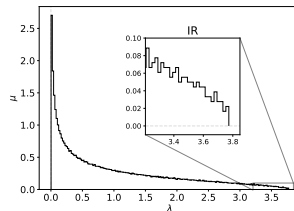
Gaussian process regression works fine for “stronger” signals, where the **IR** is not heavily polluted by noise.

Marchenko-Pastur Universality Class

Canonical Dimensions and Signal Detection



Identical results up to a “physical” **IR** scale:



The **Gaussian theory** is **not adapted** to describe the **IR sector**!

For **nearly continuous spectra**, we need **at least** a **ϕ^4 theory**.

Gaussian process regression works fine for “stronger” signals, where the **IR** is not heavily polluted by noise.

Perturbations to Marchenko-Pastur

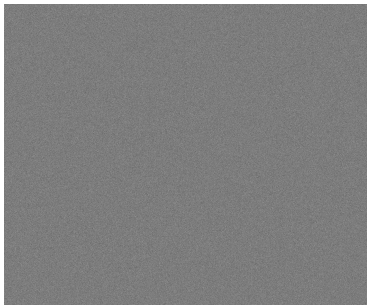
A Renormalization Group for Cats

Erbin *et al.* (2023)

Finotello *et al.* (2025)

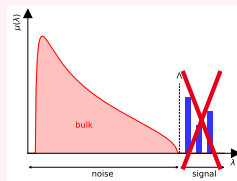
Consider a **noise model**

$$X = \beta \underbrace{S}_{\text{"signal"}} + Z, \quad Z \sim \mathcal{N}(0, 1)$$



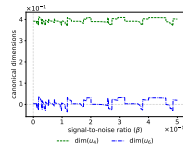
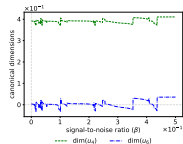
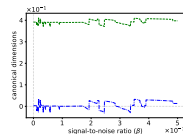
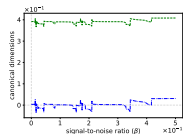
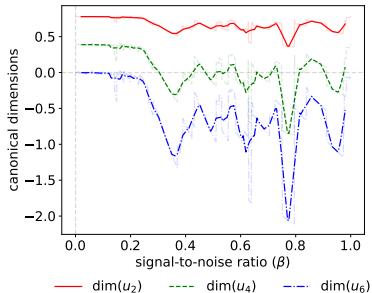
Signal Spikes

We remove the **top eigenvalues** (signal spikes) and look at the **bulk** (nearly continuous spectrum)



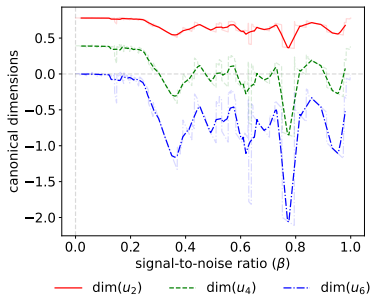
Perturbations to Marchenko-Pastur

The Effects of Signal



Perturbations to Marchenko-Pastur

The Effects of Signal



- $0 \leq \beta < \beta_t \leftarrow$ **pure noise** regime

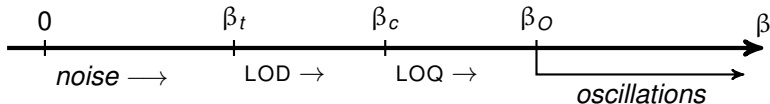
$$\dim_{u_4}(k_{\text{IR}}^2 | \beta) > 0$$

- $\beta_t \leftarrow$ **limit of detection**

$$\dim_{u_4}(k_{\text{IR}}^2 | \beta_t) = 0$$

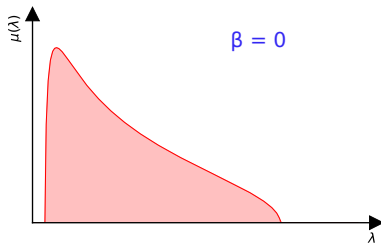
- $\beta_c \leftarrow$ **limit of quantitation**

$$\left. \frac{\partial \dim_{u_4}(k_{\text{IR}}^2 | \beta)}{\partial \beta} \right|_{\beta=\beta_c} = 0$$



Multicomponent Noise Detection

Determining a Distance from the Null Hypothesis



Under the hypothesis of **multiple sources of noise**

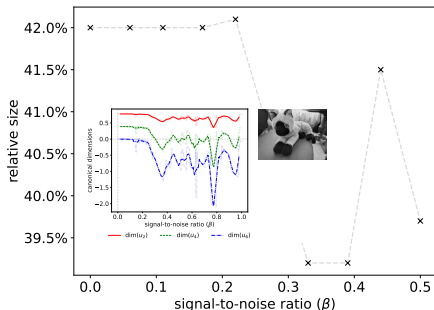
$$S = S_0 + \sum_{i=1}^M \tilde{S}_i(\omega_i) \Rightarrow X = \beta S_0 + \left(Z + \beta \sum_{i=1}^M \tilde{S}_i(\omega_i) \right) = \beta S_0 + \tilde{Z}_M(\beta),$$

Dimensional Symmetry Breaking

The Size of the Symmetric Region

In our scenario the **dimension** changes:

$$\rho(p^2) \Rightarrow d = d(p^2) \Rightarrow \text{vanishing } \mathbf{WF} \text{ fixed point}$$



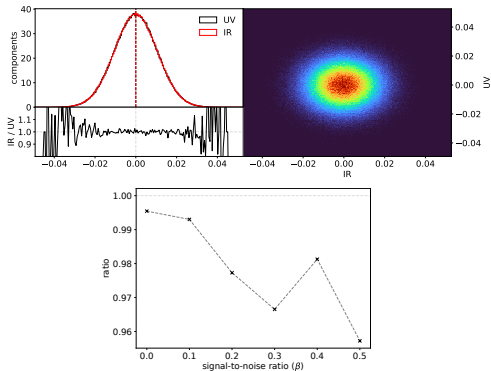
- explore different initialisations in the vicinity of the **Gaussian fixed point**
- study the **size of the symmetric region** as a function of β

Symmetric Region Size and Signal

Canonical dimensions and symmetric region size are **new observables** to detect **weak signals**!

Dimensional Symmetry Breaking

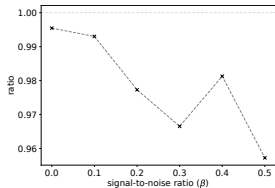
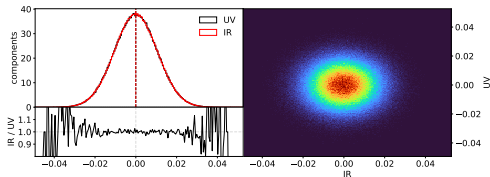
Consequences and Observables for Anomaly Detection



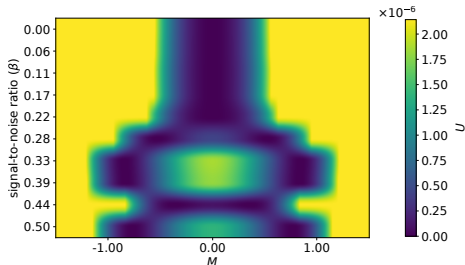
The distribution of the **components** of the **eigenvectors** changes!

Dimensional Symmetry Breaking

Consequences and Observables for Anomaly Detection



The distribution of the **components** of the **eigenvectors** changes!



Dimensional Symmetry Breaking

The flow **UV** $\rightarrow$ **IR** modifies the **effective dimension** of spacetime

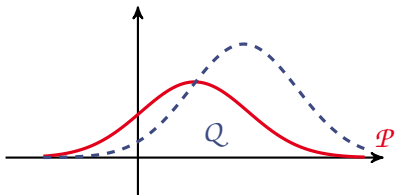
anomaly detection = dimensional symmetry breaking



5. Conclusions and a Look at the Future

Operations in Real Life

A New Definition of Distance Between Distributions?

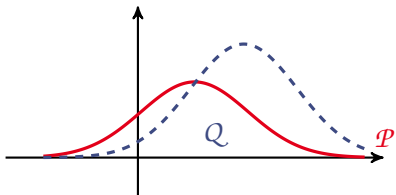


Let $\mathcal{P}$ and $\mathcal{Q}$ be two **probability distributions**:

$$D_{\text{KL}}(\mathcal{P} \parallel \mathcal{Q}) = \int dx \, \mathcal{P}(x) \ln \frac{\mathcal{P}(x)}{\mathcal{Q}(x)}$$

Operations in Real Life

A New Definition of Distance Between Distributions?



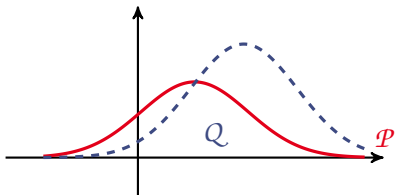
Let $\mathcal{P}$ and $\mathcal{Q}$ be two **probability distributions**:

$$D_{\text{KL}}(\mathcal{P} \parallel \mathcal{Q}) = \int dx \mathcal{P}(x) \ln \frac{\mathcal{P}(x)}{\mathcal{Q}(x)}$$

Though, it **is not a distance** (no triangle inequality)!

Operations in Real Life

A New Definition of Distance Between Distributions?



Let $\mathcal{P}$ and Q be two **probability distributions**:

$$D_{\text{KL}}(\mathcal{P} \parallel Q) = \int dx \, \mathcal{P}(x) \ln \frac{\mathcal{P}(x)}{Q(x)}$$

Though, it **is not a distance** (no triangle inequality)!

Using the **FRG**, we can define a new notion of **distance**

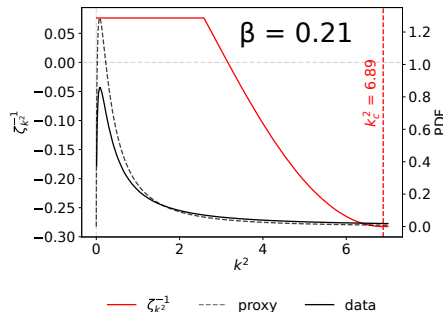
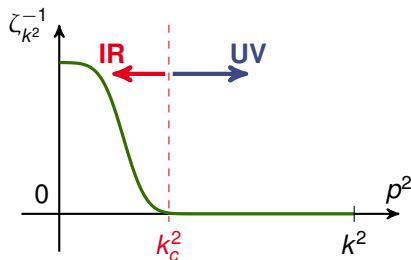
$$\text{let } \begin{cases} \rho_q(p^2) & \text{empirical dist.} \\ \tilde{\rho}_{q^*}(p^2) & \text{proxy inv.-MP dist.} \end{cases} \Rightarrow q_{k^2}^* = \underbrace{\arg \min_q \max_{(0, k^2)} \left| \dim_{u_4}(p^2) \Big|_{\rho} - \dim_{u_4}(p^2) \Big|_{\tilde{\rho}} \right|}_{\text{fit the "best" possible noise model}}$$

Operations in Real Life

Towards True Anomaly Detection and Uncertainty Quantification

We can then compute the **local adherence**:

$$\xi_{k^2}(p^2 \mid \rho) = \min_{(0, k^2)} \left(\left. \dim_{u_4}(p^2) \right|_{\bar{\rho}} - \left. \dim_{u_4}(p^2) \right|_{\rho} \right)$$



Future Research Directions

Land of Hopes and Dreams...

The **FRG approach** to data analysis enables:

- **anomaly/signal detection** via a **dimensional symmetry breaking** mechanism
- dealing with a **class of data models** via **MP universality** (sensor response, photonic noise, ...)
- detection of **very small variations** in the **bulk** of the spectrum



Future Research Directions

Land of Hopes and Dreams...

The **FRG approach** to data analysis enables:

- **anomaly/signal detection** via a **dimensional symmetry breaking** mechanism
- dealing with a **class of data models** via **MP universality** (sensor response, photonic noise, ...)
- detection of **very small variations** in the **bulk** of the spectrum

For the **future**:

- *what kind of signals can we treat?* (physical simulations, videos, HSI, ...)
- *can we formalize the notion of distance?* $\Rightarrow$ **uncertainty quantification**



Future Research Directions

Land of Hopes and Dreams...

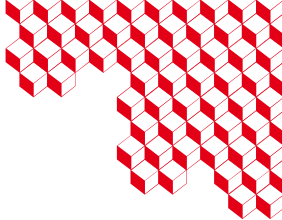
The **FRG approach** to data analysis enables:

- **anomaly/signal detection** via a **dimensional symmetry breaking** mechanism
- dealing with a **class of data models** via **MP universality** (sensor response, photonic noise, ...)
- detection of **very small variations** in the **bulk** of the spectrum

For the **future**:

- *what kind of signals can we treat?* (physical simulations, videos, HSI, ...)
- *can we formalize the notion of distance?* $\Rightarrow$ **uncertainty quantification**

work in progress with Vincent Lahoche, Dine Ousmane Samary, Parham Radpay



Renormalization Group for Signal Analysis

NPTwins2025, 7 October 2025

CEA SACLAY

91 191 Gif-sur-Yvette Cedex
France

riccardo.finotello@cea.fr