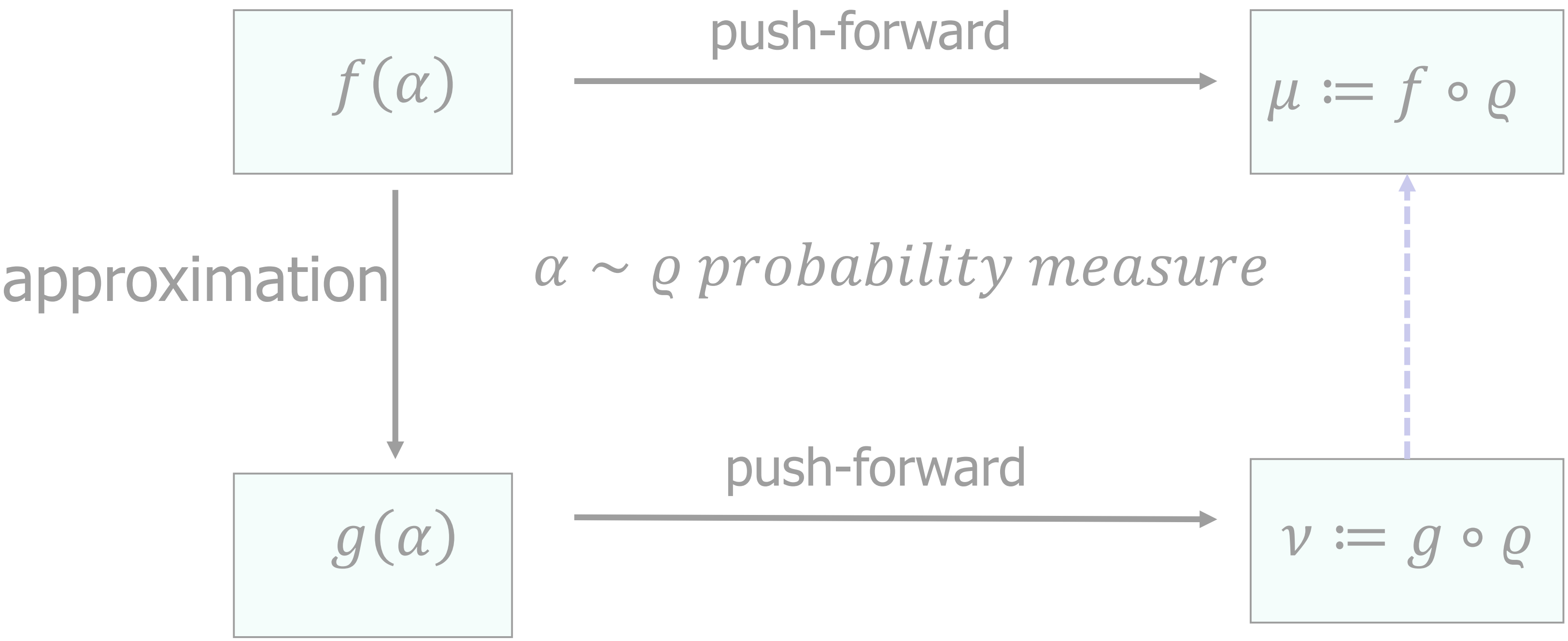
An Optimal Transport Perspective on Uncertainty Quantification



Amir Sagiv

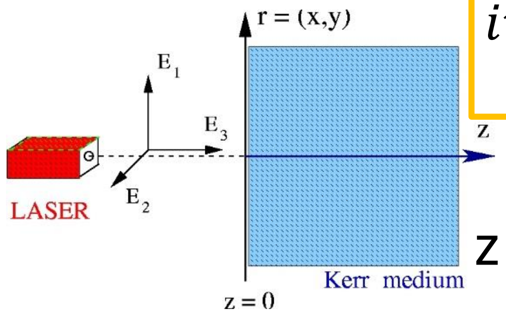
Columbia University

Two motivation slides:

nonlinear laser propagation

$$i\psi_z(z,x,y) + \Delta\psi + |\psi|^2\psi - \epsilon|\psi|^4\psi = 0$$

Nonlinear Schrodinger equation



PDE model

initial condition

$$\psi_0(x,y)$$

output

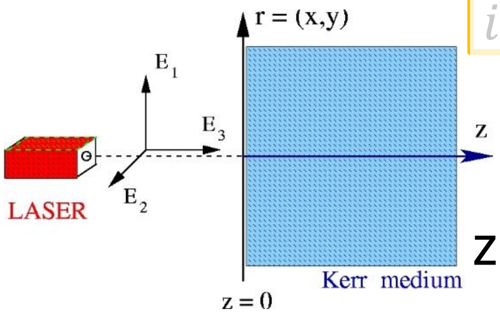
$$\psi(z,x,y)$$

Two motivation slides:

nonlinear laser propagation

$$i\psi_z(z,x,y) + \Delta\psi + |\psi|^2\psi - \epsilon|\psi|^4\psi = 0$$

Each laser shot
is different



PDE model

random initial condition

$$\psi_0(x,y;\alpha)$$

α - noise parameter

random output

$$\psi(z,x,y;\alpha)$$

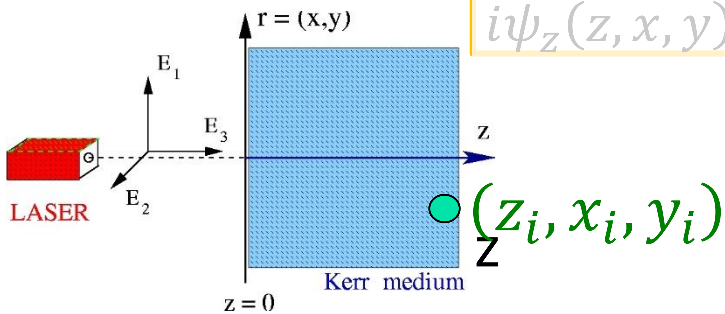
[Sagiv, Ditwkoski, Fibich, Opt. Exp. 2017]

Two motivation slides:

nonlinear laser propagation

$$i\psi_z(z,x,y) + \Delta\psi + |\psi|^2\psi - \epsilon|\psi|^4\psi = 0$$

Each laser shot
is different



PDE model

random initial condition

$$\psi_0(x,y;\alpha)$$

random output

$$\psi(z,x,y;\alpha)$$

α - noise parameter

What kind of statistics do we want to compute?

Moment estimation
e.g., $\mathrm{E}(|\psi(z_i,x_i,y_i)|^2)$,
over many realizations
(repetitions)

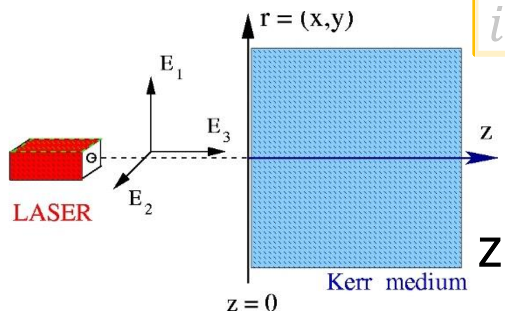
[e.g., Fibich, Eisenman, Ilan, Zigler, Opt. Lett. 2005]

Two motivation slides:

nonlinear laser propagation

$$i\psi_z(z,x,y) + \Delta\psi + |\psi|^2\psi - \epsilon|\psi|^4\psi = 0$$

Each laser shot is different



PDE model

random initial condition

$$\psi_0(x,y;\alpha)$$

random output

$$\psi(z,x,y;\alpha)$$

α - noise parameter

What kind of statistics do we want to compute?

Moment estimation

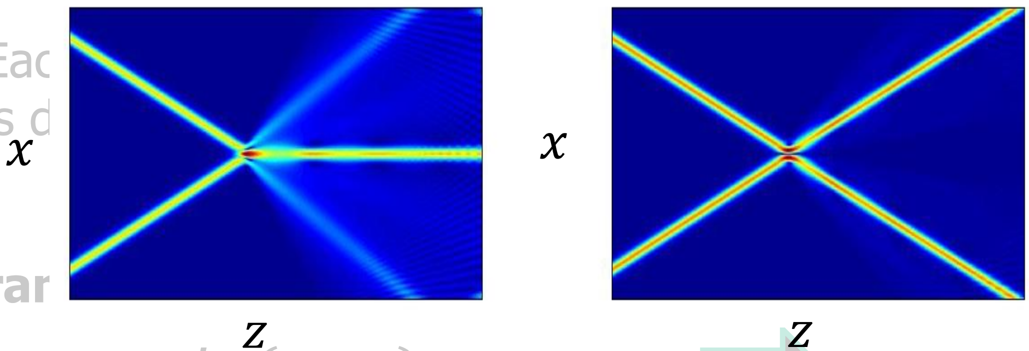
e.g., $\mathrm{E}(|\psi(z_i,x_i,y_i)|^2)$, over many realizations (repetitions)

Density estimation

Probability Density Function (PDF) of some “quantity of interest” $f(\psi)$

Why study the PDF — Examples from optics

Beam fusion | Beam repulsion

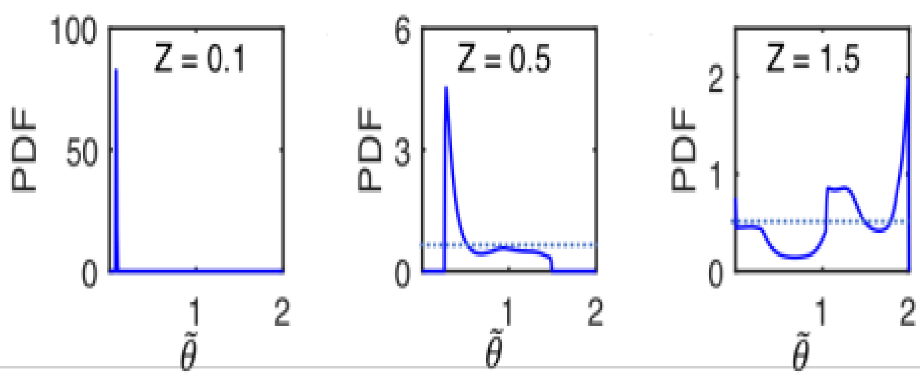


Example I: over many repetitions, what are the chances of fusion vs. repulsion?

$$\psi_0 = R_{\kappa_1}(x+d)e^{i\theta x} + (1 + 0.1\alpha)R_{\kappa_1}(x-d)e^{-i\theta x}$$

Random amplitude

[Sagiv, Ditwkoski, Fibich, Opt. Exp. 2017]



Example II: Distribution of *polarization* as a function of propagation distance

[\w Patwardhan et al, PRA 2019]

General standard nonlinear PDE settings

Initial value problem

$$\begin{cases} u_t(t,\boldsymbol{x}) = Q(\boldsymbol{x},u)u \\ u(t=0,\boldsymbol{x}) = u_0(\boldsymbol{x}) \end{cases}$$

- "quantity of interest" (model output) $f\big(u(t,\boldsymbol{x})\big)$
  - e.g., $f = u(t_i, x_i)$, $\mathrm{f} = \int dx\, |\mathrm{u}|^2$, ...
  - u & $f(u)$ are evaluated numerically

General Settings – Nonlinear PDE with randomness

Initial value problem with randomness (both i.c. u_0 and operator Q)

$$\begin{cases} u_t(t,\boldsymbol{x};\boldsymbol{\alpha}) = Q(\boldsymbol{x},u;\boldsymbol{\alpha})u \\ u(t=0,\boldsymbol{x};\boldsymbol{\alpha}) = u_0(\boldsymbol{x};\boldsymbol{\alpha}) \end{cases}$$

- $\boldsymbol{\alpha}$ distributed according to a known measure
- "quantity of interest" (model output) $f(\boldsymbol{\alpha}) := f\big(u(t,\boldsymbol{x};\boldsymbol{\alpha})\big)$
  - e.g., $f = u(t_i, x_i)$, $\mathrm{f} = \int dx\, |\mathrm{u}|^2$, ...
  - u & $f(u)$ are evaluated <u>numerically</u>

General Settings – Nonlinear PDE with randomness

Initial value problem with randomness

$$\begin{cases} u_t(t, \boldsymbol{x}; \boldsymbol{\alpha}) = Q(\boldsymbol{x}, u; \boldsymbol{\alpha})u \\ u(t = 0, \boldsymbol{x}; \boldsymbol{\alpha}) = u_0(\boldsymbol{x}; \boldsymbol{\alpha}) \end{cases}$$

- $\boldsymbol{\alpha}$ distributed according to a known measure
- "quantity of interest" (model output) $f(\boldsymbol{\alpha}) := f\big(u(t, \boldsymbol{x}; \boldsymbol{\alpha})\big)$
  - e.g., $f = u(t_i, x_i)$, $\mathrm{f} = \int dx\, |\mathrm{u}|^2$, ...

How to approximate the PDF of $f(\boldsymbol{\alpha})$,
a random variable, numerically?

Agenda

- PDF approximation
  - Is moment-estimation sufficient?
  - An algorithm & convergence results
- Transport-theory point of view

Agenda

- PDF approximation
  - Is moment-estimation sufficient?

  How does standard UQ methods perform in this task

  - An algorithm & convergence results

- Transport-theory point of view

General Settings – Nonlinear PDE with randomness

Initial value problem with randomness

$$\begin{cases} u_t(t,x;\boldsymbol{\alpha}) = Q(x,u;\boldsymbol{\alpha})u \\ u(t=0,x;\boldsymbol{\alpha}) = u_0(x;\boldsymbol{\alpha}) \end{cases}$$

- α distributed according to a known measure
- "quantity of interest" (model output) $f(\boldsymbol{\alpha}) := f\big(u(t,x;\boldsymbol{\alpha})\big)$
  - e.g., $f = u(t_i, x_i)$, $\mathrm{f} = \int dx\, |\mathrm{u}|^2, \ldots$

How to approximate the PDF of $f(\boldsymbol{\alpha})$,
a random variable, numerically?

Constraints:

- Can only compute $f(\alpha_j)$ for a <u>given</u> α_j
- Computation of $f(\boldsymbol{\alpha_j})$ is expensive (e.g., solving the (3+1)dimensional NLS)
  - Can only use a small sample $\{f(\alpha_1), \ldots, f(\alpha_N)\}$

Standard statistical methods

Step I – draw i.i.d. samples $\alpha_1, \dots, \alpha_N$
Step II – compute the samples $\{f(\alpha_1), \dots, f(\alpha_N)\}$

Moment estimation

- Monte-Carlo $\mathbf{E}_\alpha[f] \approx \frac{1}{N}\sum_{n=1}^{N} f_n$
- …

Density (PDF) estimation

- Histogram method
- Kernel density estimators (KDE)
- …

Standard statistical methods

Step I – draw i.i.d. samples $\alpha_1, \dots, \alpha_N$
Step II – compute the samples $\{f(\alpha_1), \dots, f(\alpha_N)\}$

Moment estimation

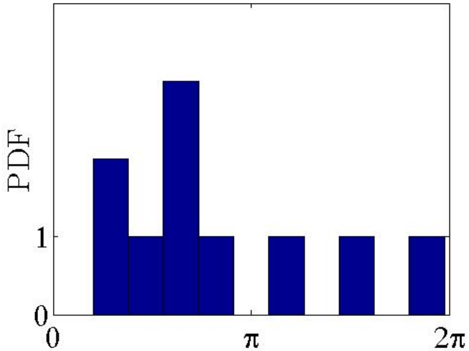
- Monte-Carlo $\mathbf{E}_{\alpha}[f] \approx \frac{1}{N}\sum_{n=1}^{N} f_n$
- …

Density (PDF) estimation

- Histogram method
- Kernel density estimators (KDE)
- …

- Poor approximations for small N $\left(\frac{1}{\sqrt{N}}\right.$ error)

e.g. Histogram method with N=10 samples



Standard statistical methods

Step I – draw i.i.d. samples $\alpha_1, \ldots, \alpha_N$
Step II – compute the samples $\{f(\alpha_1), \ldots, f(\alpha_N)\}$

Moment estimation

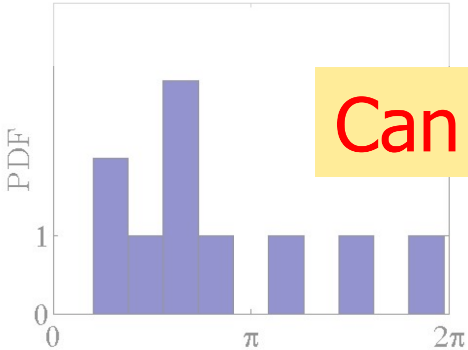
- Monte-Carlo $\mathbf{E}_{\alpha}[f] \approx \frac{1}{N}\sum_{n=1}^{N} f_n$
- …

Density (PDF) estimation

- Histogram method
- Kernel density estimators (KDE)
- …

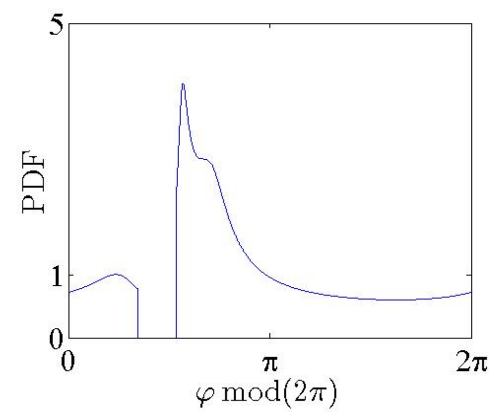
- Poor approximations for small N

e.g. Histogram method with N=10 samples



Can we improve?

Exact PDF



Standard statistical methods

Step I – draw i.i.d. samples $\alpha_1, \dots, \alpha_N$
Step II – compute the samples $\{f(\alpha_1), \dots, f(\alpha_N)\}$

Moment estimation

- Monte-Carlo $\mathbf{E}_{\alpha}[f] \approx \frac{1}{N} \sum_{n=1}^{N} f_n$
- …

Density (PDF) estimation

- Histogram method
- Kernel density estimators (KDE)
- …

Can we improve?

- Methods above only use $\{f(\alpha_1), \dots, f(\alpha_N)\}$.
- We can also use
  1. The relation $f(\boldsymbol{\alpha}) \leftrightarrow \boldsymbol{\alpha}$
  2. Smoothness of $f(\boldsymbol{\alpha})$

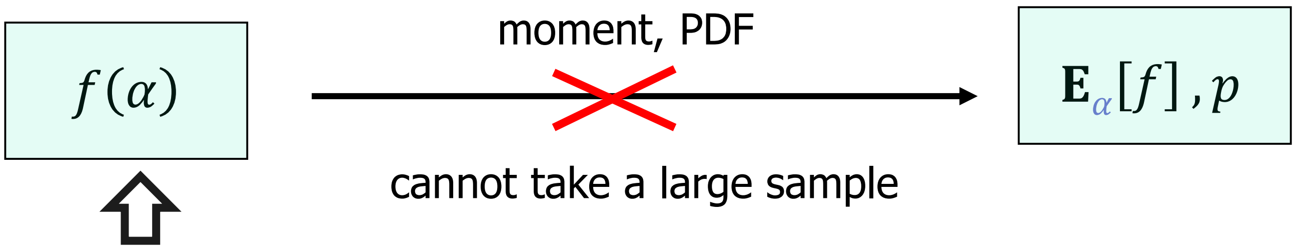
These assumptions underly many studies in uncertainty quantification (UQ), specifically in uncertainty propagation

Approximation-based estimation

p is the PDF of f

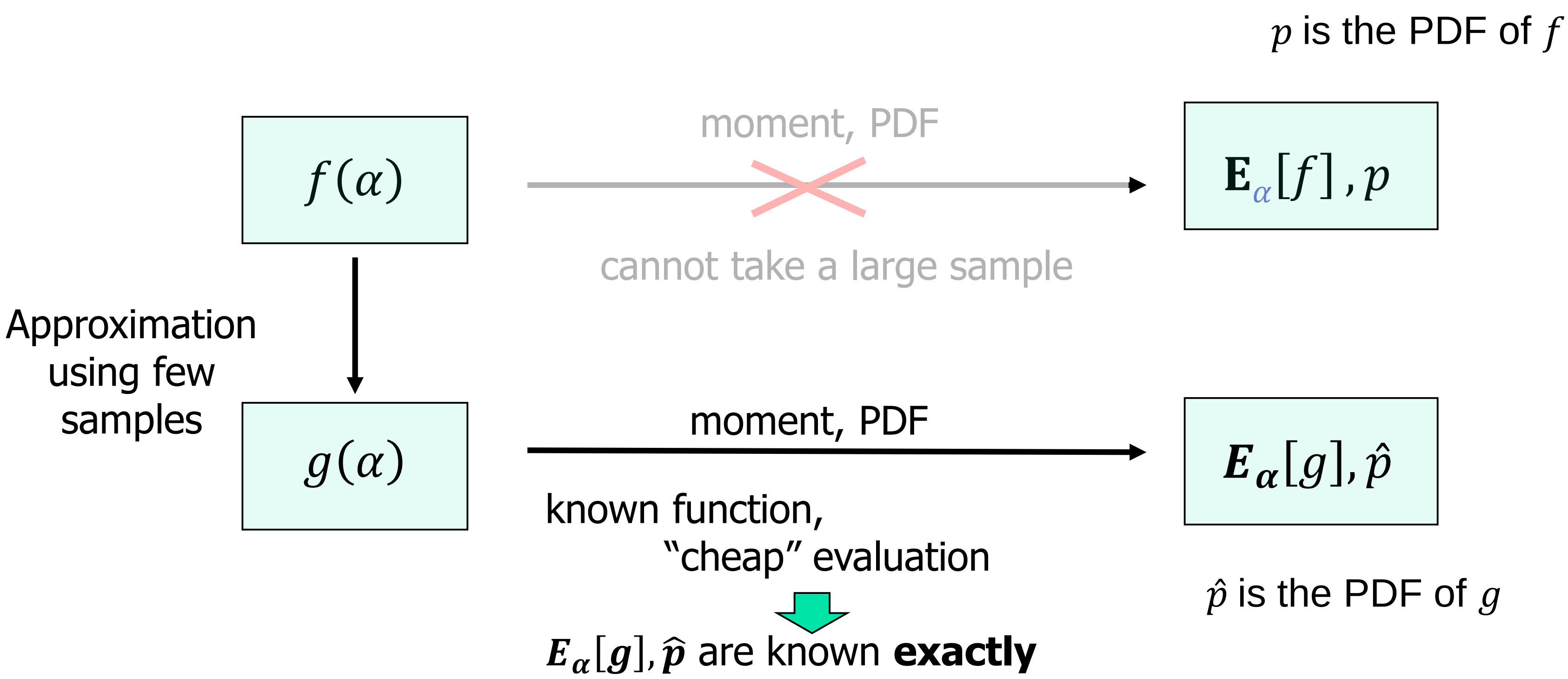
Approximation-based estimation

p is the PDF of f



- Unknown explicitly
- Each evaluation is computationally expensive

Approximation-based estimation



Approximation-based estimation

Questions

- Which approximation $g(\alpha) \approx f(\alpha)$ should be used?
- How small are $\mathrm{E}_\alpha[f] - \mathrm{E}_\alpha[g]$ and $||p - \hat{p}||$?

Attempt I - generalized polynomial chaos (gPC)

Standard in the field of uncertainty quantification (UQ)

Approximate f using orthogonal polynomials $\{q_n(\alpha)\}$

$$f_N(\alpha) = \sum_{n=0}^{N-1} \langle q_n \, , f \rangle \, q_n(\alpha)$$

- Spectral accuracy (moments and L^2)

$$\mathbf{E}_{\alpha}[f] - \mathbf{E}_{\alpha}[f_N] = O(e^{-\gamma N}), \qquad N \gg 1 \quad f \textit{ is analytic}$$

[see e.g., D. Xiu, 2010]

Attempt I - generalized polynomial chaos (gPC)

Standard in the field of uncertainty quantification (UQ)

Approximate f using orthogonal polynomials $\{q_n(\alpha)\}$

$$f_N(\alpha) = \sum_{n=0}^{N-1} \langle q_n, f \rangle \, q_n(\alpha)$$

- Spectral accuracy (moments and L^2)

$$\mathbf{E}_\alpha[f] - \mathbf{E}_\alpha[f_N] = O(e^{-\gamma N}), \qquad N \gg 1 \quad f \textit{ is analytic}$$

But,

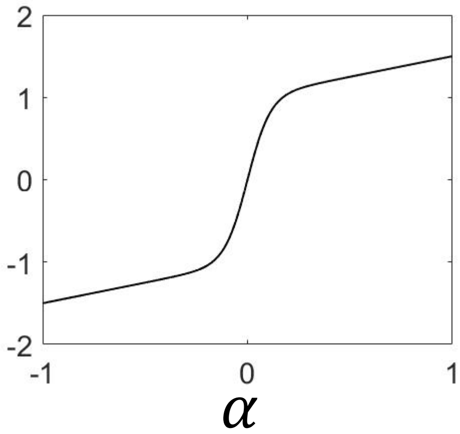
<u>PDF estimation</u>

No theory for $\|p - p_N\|$

Will it work in practice?

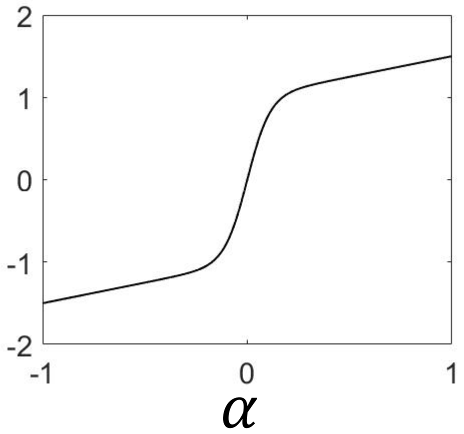
Example – gPC fails at PDF estimation

$$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim Uniform\,[-1, 1]$$

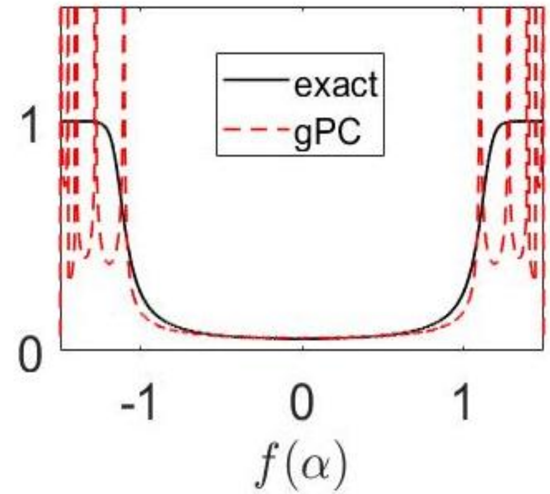


Example – gPC fails at PDF estimation

$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim Uniform\,[-1, 1]$



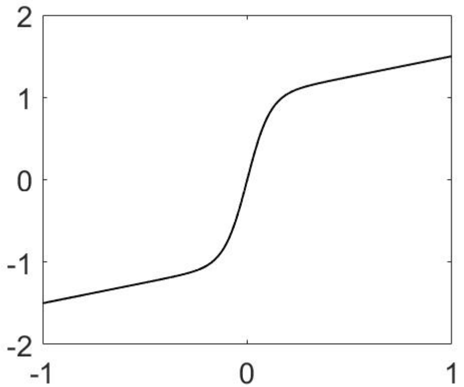
PDF approximation, $N = 12$ samples



Why does it fail?

Example – gPC fails at PDF estimation

$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim U[-1,1]$



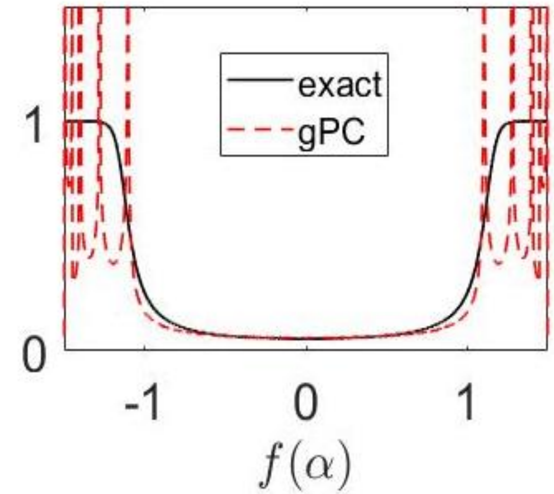
Why does it fail?

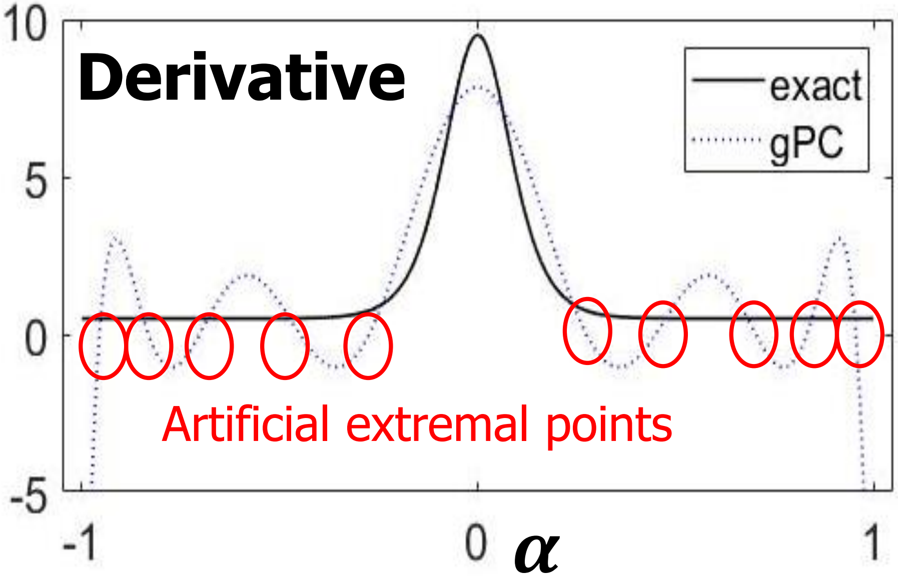
Lemma: Under general smoothness conditions

$$p(y) = \sum_{f(\alpha)=y} \frac{1}{|f'(\alpha)|}$$

Although gPC is spectrally accurate (in L^2), it produces "artificial" zero derivatives.

PDF approximation, $N = 12$ samples





Agenda

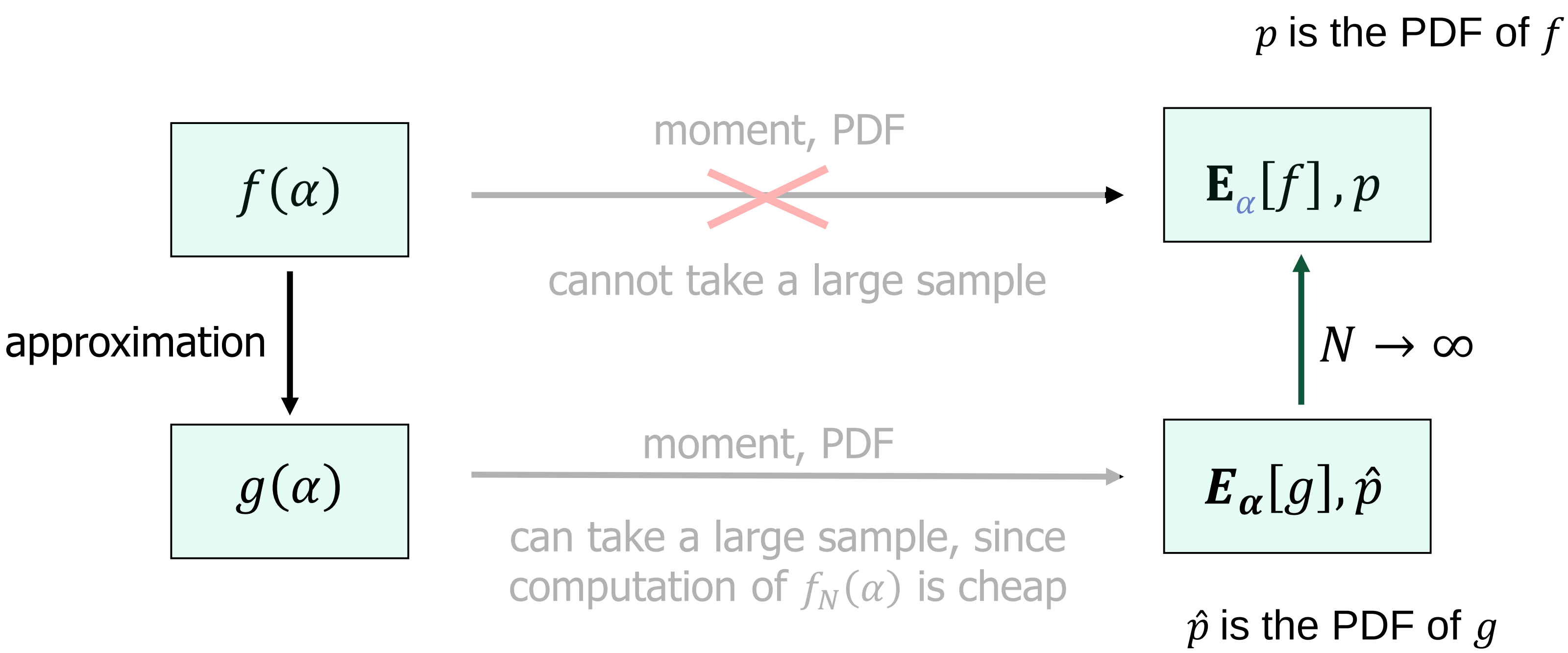
- PDF approximation
  - Is moment-estimation sufficient?

  - An algorithm & convergence results

  Approximating pushed-forward densities, provably

- Transport-theory point of view

An Alternative Approximation-based estimation



Lessons learned

- For PDF approximation, spectral moment accuracy is not sufficient.
- It is necessary that $g' \neq 0 \leftrightarrow f' \neq 0$
  $|g - f|, |g' - f'| \ll 1,$
- "Monotonicity preserving approximation"

Solution: use spline interpolation

(piece-wise polynomials)

The solution - Spline-based approach (1d)

Theorem 1 (Ditkowski, Fibich, AS ’18):

Let p and p_N be the probability density functions (**PDF**) of $f(\alpha)$ and its **m-degree spline** interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m},$$

The solution - Spline-based approach (1d)

Theorem 1 (Ditkowski, Fibich, AS ’18):

Let $f \in C^{m+1}_{piecewise}([\alpha_{\min}, \alpha_{\max}])$ with $|\boldsymbol{f'}| > \boldsymbol{a} > \boldsymbol{0}$, let α be distributed by $c(\alpha)d\alpha$ where $c \in C^1([\alpha_{\min}, \alpha_{\max}])$.

and

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m},$$

The solution - Spline-based approach (1d)

Theorem 1 (Ditkowski, Fibich, AS '18):

Let $f \in C^{m+1}_{piecewise}([\alpha_{\min}, \alpha_{\max}])$ with $|f'| > a > 0$, let α be distributed by $c(\alpha)d\alpha$ where $c \in C^1([\alpha_{\min}, \alpha_{\max}])$.

and

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m},$$

For all

$$N > \sqrt[m]{\frac{2C_m||f^{(m+1)}||_\infty}{a}}(\alpha_{\max} - \alpha_{min})$$

Proof "ingredients"

Theorem 1 (Ditkowski, Fibich, AS '18):

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m}$$

Lemma: Under general smoothness conditions

$$p(y) = \sum_{f(\alpha)=y} \frac{c(\alpha)}{|f'(\alpha)|}$$

Proof "ingredients"

Theorem 1 (Ditkowski, Fibich, AS '18):

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m}$$

Lemma: Under general smoothness conditions

$$p(y) = \sum_{f(\alpha)=y} \frac{c(\alpha)}{|f'(\alpha)|}$$

Theorem (Meyer, Hall, '76): for $f \in C^{m+1}$, Then

$$||(f-s)^{(j)}||_\infty \leq C_m(f)h^{m+1-j} \quad j = 0, \dots m-1$$

where h>0 is the maximal spacing between interpolation points

Proof "ingredients"

Theorem 1 (Ditkowski, Fibich, AS '18):

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-m}$$

Lemma: Under general smoothness conditions

$$p(y) = \sum_{f(\alpha)=y} \frac{c(\alpha)}{|f'(\alpha)|}$$

Theorem (Meyer, Hall, '76): for $f \in C^{m+1}$, Then

$$||(f - s)^{(j)}||_\infty \leq C_m(f)h^{m+1-j} \quad j = 0, \dots m-1$$

where h>0 is the maximal spacing between interpolation points

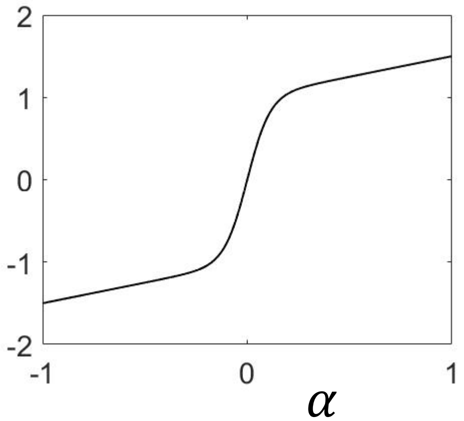
Hence, if f is monotone, N is high enough, and y in f's image, $\alpha \sim U(-1,1)$

$$||p - p_N||_1 = \int dy\, |p(y) - p_N(y)| = \int dy \left| \frac{1}{f'\left(f^{-1}(y)\right)} - \frac{1}{s'\left(s^{-1}(y)\right)} \right|$$

$$= \cdots$$

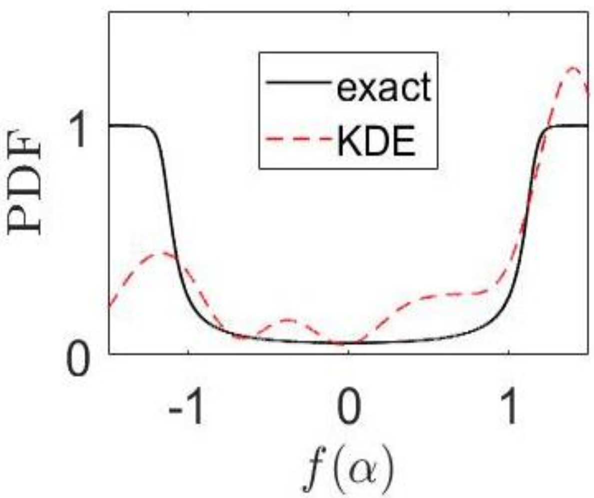
PDF estimation

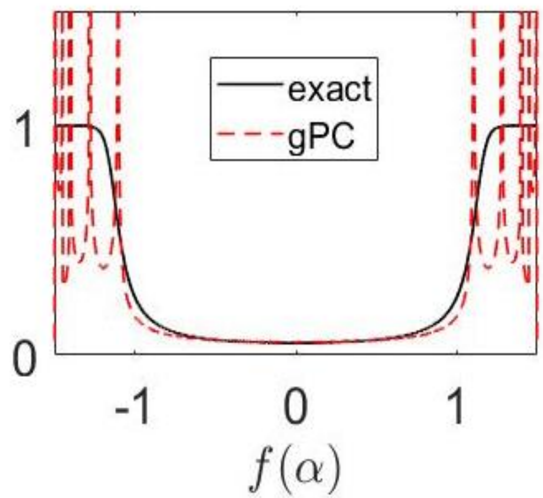
$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim Uniform\,[-1, 1]$



PDF approximation, $N = 12$

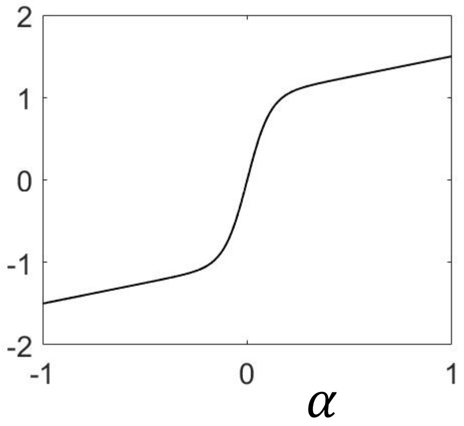
Statistically optimal



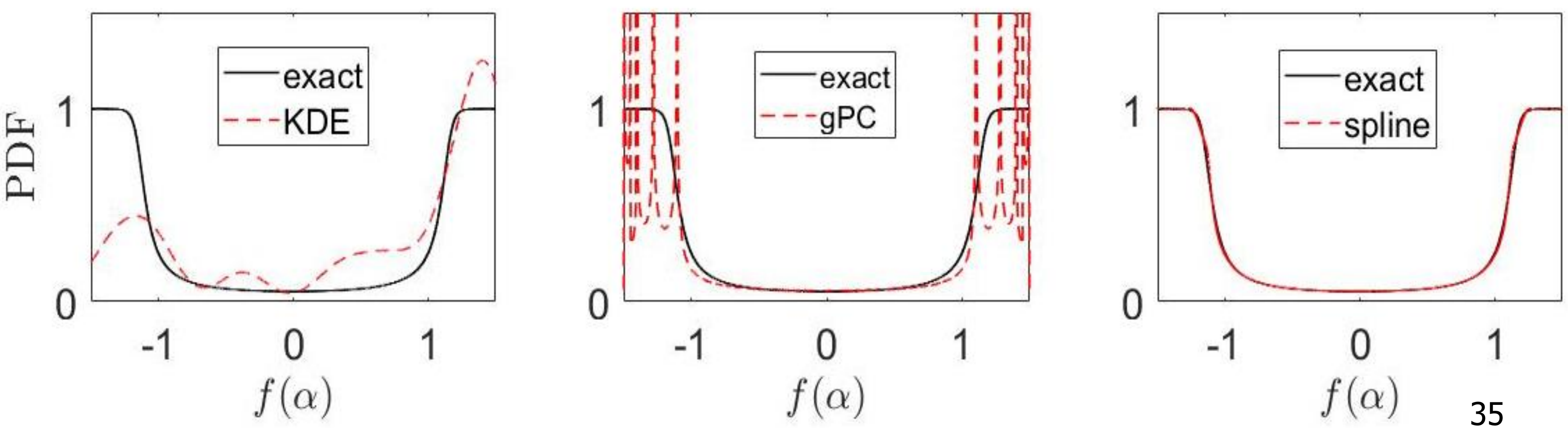


PDF estimation

$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim Uniform\,[-1, 1]$



PDF approximation, $N = 12$



Coupled NLS example

[with Patwardhan et al, PRA, 2019]

$$i\frac{\partial}{\partial t}A_\pm(t,x) + \frac{\partial^2}{\partial x^2}A_\pm + \frac{2}{3}\left(|A_\pm|^2 + 2|A_\mp|^2\right)A_\pm = 0$$

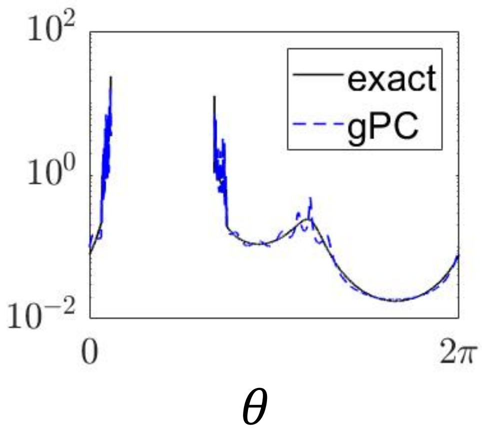
phase: $\varphi_\pm(t) = \arg\left(A_\pm(t, x=0)\right)\ mod\ (2\pi)$

polarization $\theta(t) = \varphi_+(t) - \varphi_-(t)$

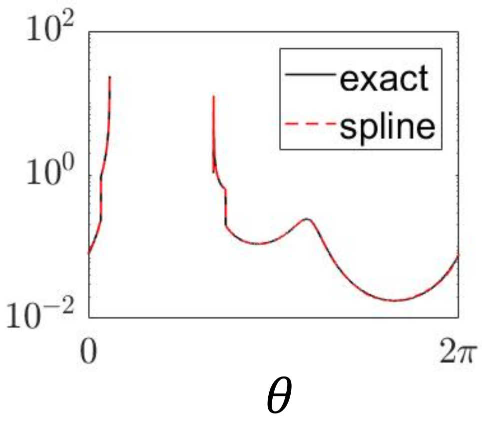
Random elliptical beam –

$$A_\pm(t=0) = (1+\alpha)C_\pm e^{-x^2}, \qquad \alpha \sim U(-0.1, 0.1)$$

PDF, N=64



PDF, N=64



Coupled NLS example

[with Patwardhan et al, PRA, 2019]

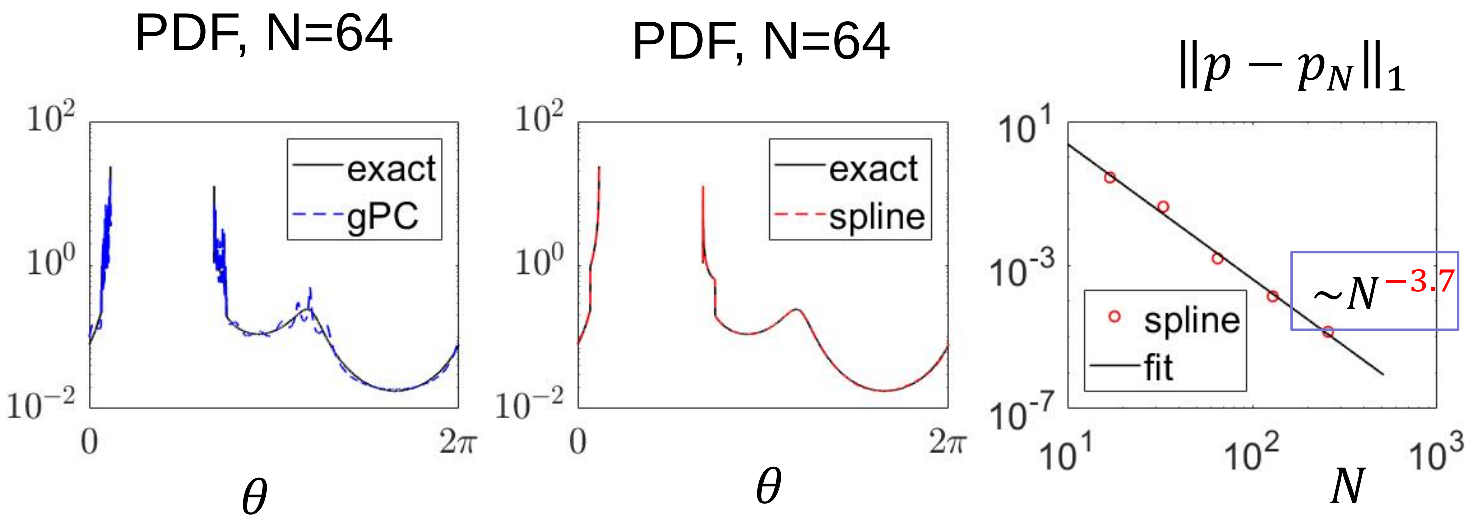
$$i\frac{\partial}{\partial t}A_\pm(t,x) + \frac{\partial^2}{\partial x^2}A_\pm + \frac{2}{3}\left(|A_\pm|^2 + 2|A_\mp|^2\right)A_\pm = 0$$

phase: $\varphi_\pm(t) = \arg\left(A_\pm(t, x=0)\right) \ mod\ (2\pi)$

polarization $\theta(t) = \varphi_+(t) - \varphi_-(t)$

Random elliptical beam –

$$A_\pm(t=0) = (1+\alpha)C_\pm e^{-x^2}, \qquad \alpha \sim U(-0.1, 0.1)$$



Burgers equation – shock location

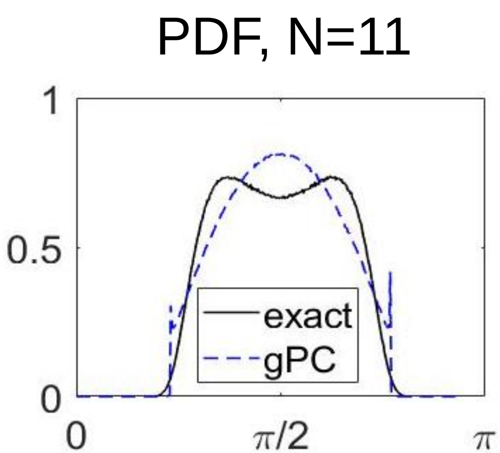
$$u_t(t,x) + \frac{1}{2}\left(u^2\right)_x = \frac{1}{2}\left(sin\,(x)\right)_x$$

Initial condition: $u_0(x) = \alpha\sin(x)$

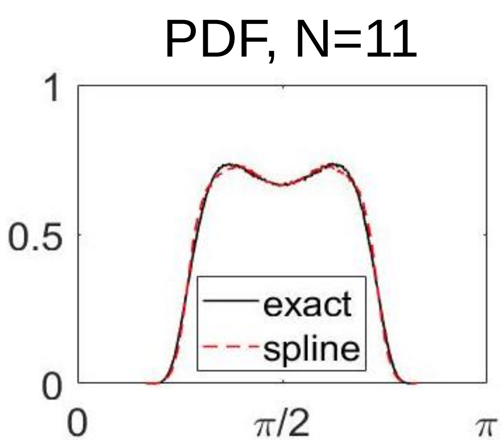
Shock location at $t \to \infty$ $\alpha = -cos(X_s)$

Distribution of random initial amplitude –

$$\alpha(\nu) = \begin{cases} \frac{-1+\sqrt{1+4\nu^2}}{2\nu} & \nu \neq 0 \\ 0 & \nu = 0 \end{cases} \qquad \nu \sim N(0,\sigma)$$



X_s



X_s

[compare Chen, Gottlieb, Hesthaven, JCP 2005]

Burgers equation – shock location

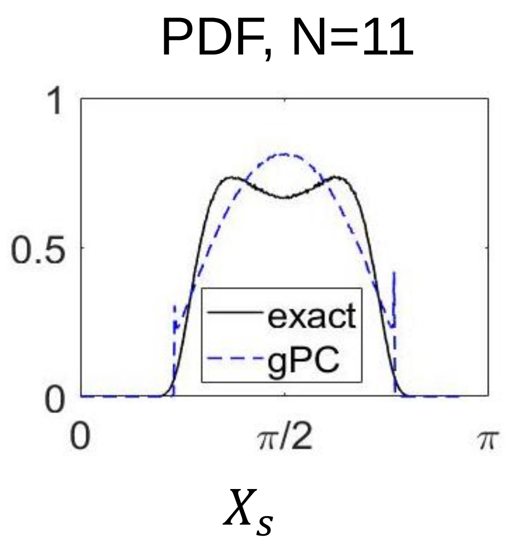
$$u_t(t,x) + \frac{1}{2}\left(u^2\right)_x = \frac{1}{2}\left(sin\,(x)\right)_x$$

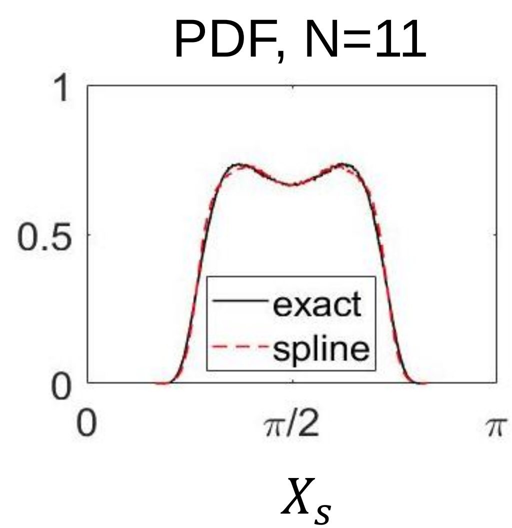
Initial condition: $u_0(x) = \alpha \sin(x)$

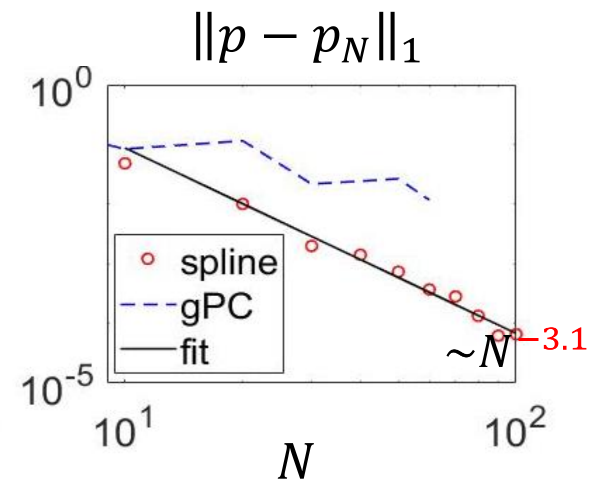
Shock location at $t \to \infty$ $\alpha = -cos(X_s)$

Distribution of random initial amplitude –

$$\alpha(\nu) = \begin{cases} \frac{-1+\sqrt{1+4\nu^2}}{2\nu} & \nu \neq 0 \\ 0 & \nu = 0 \end{cases} \qquad \nu \sim N(0,\sigma)$$







[compare Chen, Gottlieb, Hesthaven, JCP 2005]

Spline-based density estimation for multidimensional noise

Can this problem be solved if the input noise is multi-dimensional?

(physically – multiple uncertain or noisy terms in the system)

Spline-based density estimation for multidimensional noise

Theorem 2 (Ditkowski, Fibich, AS '18):

Let $\Omega = [0,1]^d$, let $f \in C^{m+1}(\Omega)$ with $|\nabla f| > a > 0$, let α be uniformly distributed in Ω,

and

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree tensor-product spline interpolant on N^d equi-distributed points. Then

$$\|p - p_N\|_1 \leq K N^{-\frac{m}{d}},$$

Curse of dimensionality

Theorem 2 (Ditkowski, Fibich, AS '18):

Let $\Omega = [0,1]^d$, let $f \in C^{m+1}(\Omega)$ with $|\nabla f| > a > 0$, let α be uniformly distributed in Ω,

and

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree tensor-product spline interpolant on $\boldsymbol{N^d}$ equi-distributed points. Then

$$\|p - p_N\|_1 \leq K \boldsymbol{N^{-\frac{m}{d}}},$$

For kernel density estimators [e.g., Devroye `84]

$$||p - p_{kde,N}||_1 \sim N^{-0.4}$$

Our method is preferable when $d \leq \frac{5m}{2}$

From d=1 to d>1:
"Same" result, more complicated proof

Theorem 1 (Ditkowski, Fibich, AS '18):

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$||p - p_N||_1 \leq KN^{-\frac{m}{d}}$$

Theorem (Schultz, '69): for $f \in C^{m+1}$,
Then

$$||(f - s)^{(j)}||_\infty \leq C_m(f)h^{m+1-j} \quad j = 0, \dots m - 1$$

where h>0 is the maximal spacing between interpolation points

From d=1 to d>1:
"Same" result, more complicated proof

Theorem 1 (Ditkowski, Fibich, AS '18):

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq K N^{-\frac{m}{d}}$$

Lemma: Under general smoothness conditions

$$p(y) \sim \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma$$

Theorem (Schultz, '69): for $f \in C^{m+1}$, Then

$$||(f - s)^{(j)}||_\infty \leq C_m(f) h^{m+1-j} \quad j = 0, \dots m-1$$

where h>0 is the maximal spacing between interpolation points

From d=1 to d>1:
"Same" result, more complicated proof

Theorem 1 (Ditkowski, Fibich, AS '18):
Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq K N^{-\frac{m}{d}}$$

Lemma: Under general smoothness conditions

$$p(y) \sim \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma$$

Theorem (Schultz, '69): for $f \in C^{m+1}$,
Then

$$||(f-s)^{(j)}||_\infty \leq C_m(f) h^{m+1-j} \quad j = 0, \dots m-1$$

where h>0 is the maximal spacing between interpolation points

Under some conditions

$$||p - p_N||_1 = \int dy \, |p(y) - p_N(y)| = \int dy \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma - \int_{g^{-1}(y)} \frac{1}{|\nabla g|} d\sigma = \cdots$$

From d=1 to d>1:
"Same" result, more complicated proof

Theorem 1 (Ditkowski, Fibich, AS '18):
Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree spline interpolant on N equi-distributed points. Then

$$\|p - p_N\|_1 \leq K N^{-\frac{m}{d}}$$

Lemma: Under general smoothness conditions

$$p(y) \sim \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma$$

Theorem (Schultz, '69): for $f \in C^{m+1}$,
Then

$$||(f-s)^{(j)}||_\infty \leq C_m(f) h^{m+1-j} \quad j = 0, \dots m-1$$

where h>0 is the maximal spacing between interpolation points

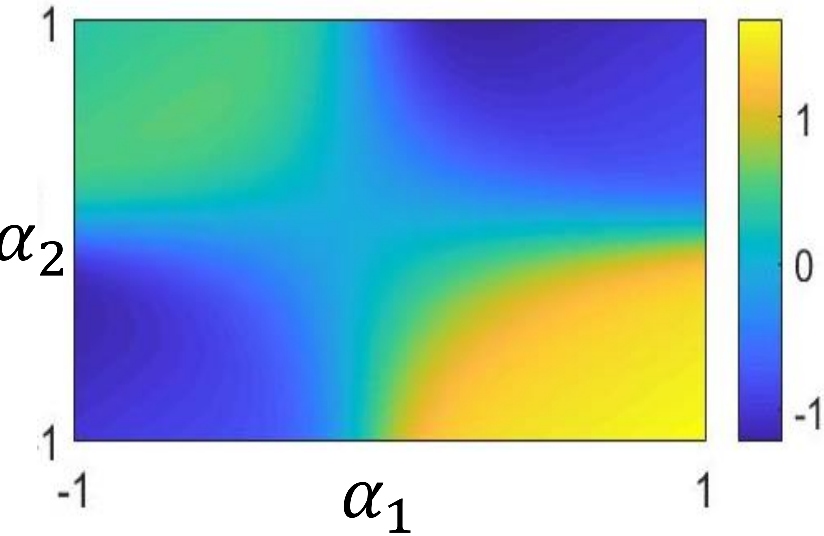
Under some conditions

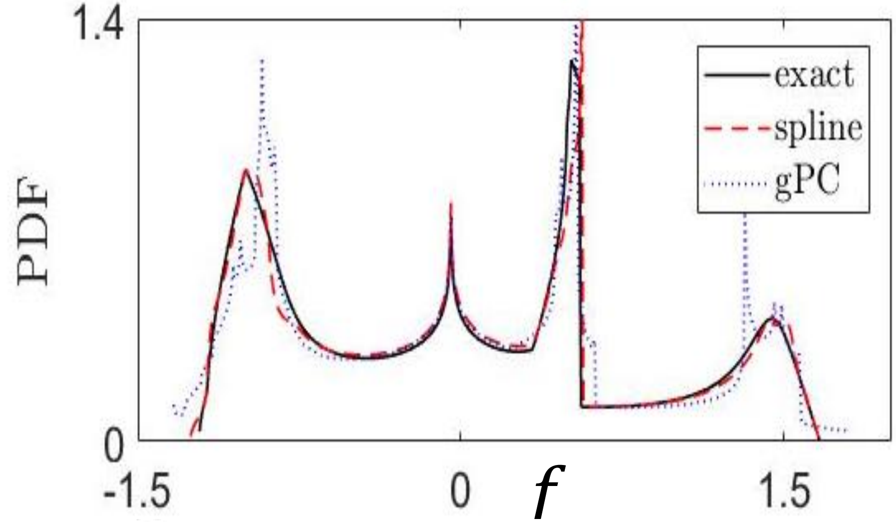
$$||p - p_N||_1 = \int dy\, |p(y) - p_N(y)| = \int dy \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma - \int_{g^{-1}(y)} \frac{1}{|\nabla g|} d\sigma = \cdots$$

Different manifolds – more complications

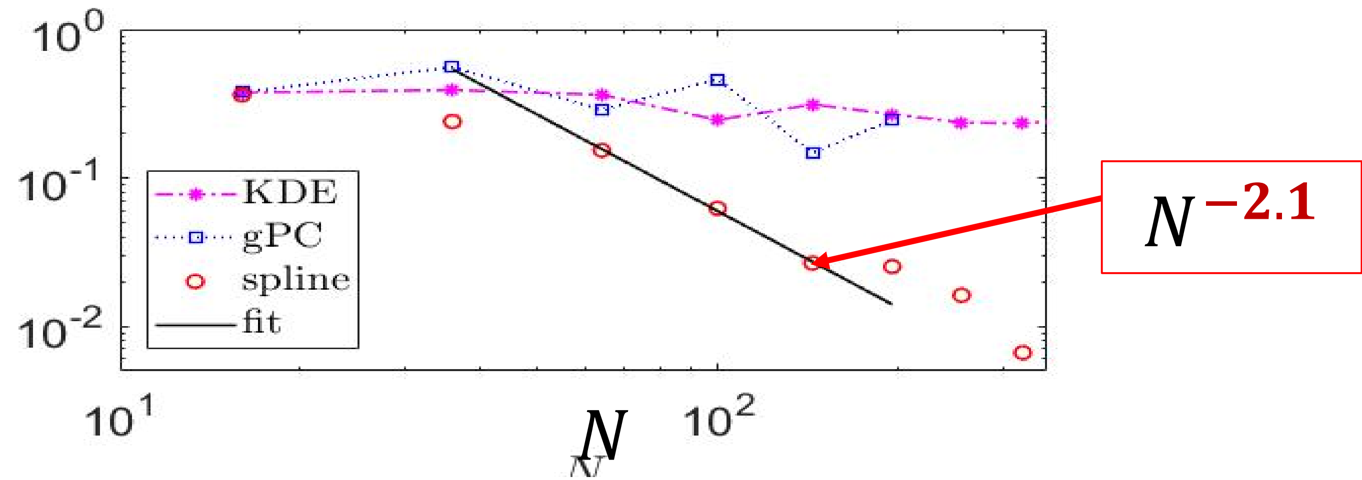
2-dimensional example

$$f(\alpha_1,\alpha_2) = \tanh\left(6\alpha_1\alpha_2 + \frac{\alpha_1}{2}\right) + \frac{\alpha_1+\alpha_2}{2}, \qquad \alpha_1,\alpha_2 \sim Uni(-1,1), \qquad i.i.d.$$





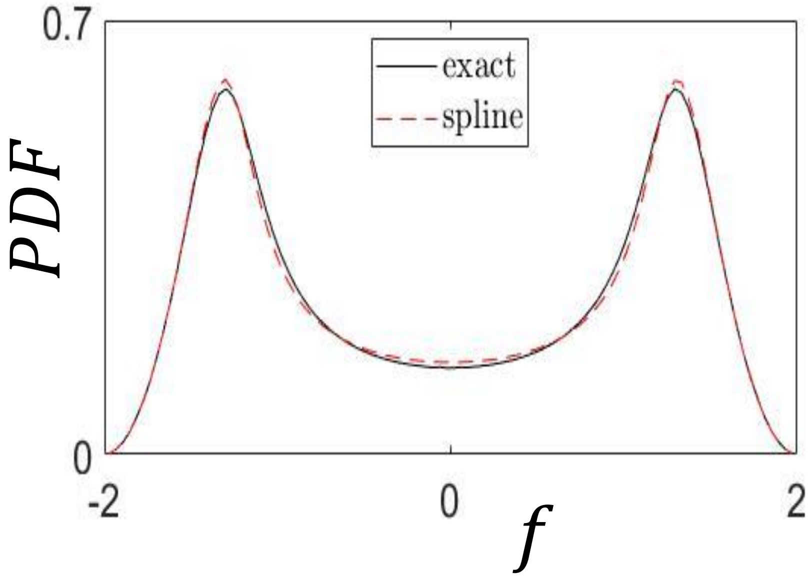
L^1 $error$



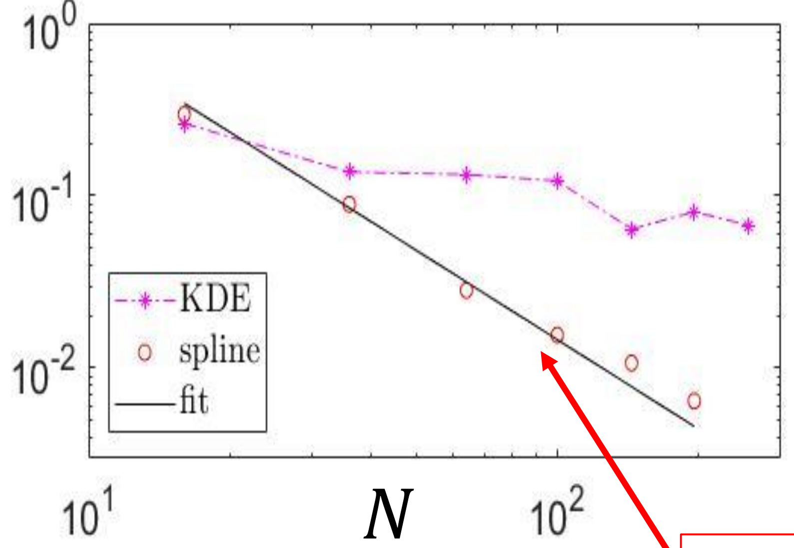
3 dimensional example

$$f(\alpha_1, \alpha_2, \alpha_3) = \tanh(2\alpha_1 + 3\alpha_2 + 3\alpha_3) + \frac{\alpha_1 + \alpha_2 + \alpha_3}{3},$$

$$\alpha_1, \alpha_2, \alpha_3 \sim Uni(-1,1), \qquad i.i.d.$$



L^1 $error$



$N^{-1.09}$

Conclusions (non-transport outlook)

- Convergence of moments and in L^2 does not guarantee convergence in PDFs

- Spline perform well for PDF approximation
  - Any other “local” method might do – RBFs, other splines, GMM,…
  - With theoretical guarantees in all dimensions.
  - With explicit “maximal dimensions” of effectiveness

A. Sagiv, A. Ditkowski, G. Fibich
Density estimation in uncertainty propagation problems using a surrogate model
arXiv 1803.10991 (under review)

Conclusions (non-transport outlook)

- Convergence of moments and in L^2 does not guarantee convergence in PDFs

- Spline perform well for PDF approximation
  - Any other “local” method might do – RBFs, other splines, GMM,…
  - With theoretical guarantees in all dimensions.
  - With explicit “maximal dimensions” of effectiveness

Open question:
Can the theory of push-forwarded densities be simplified?

Agenda

- PDF approximation
  - Is moment-estimation sufficient?
  - An algorithm & convergence results
- Transport-theory point of view

Simplifying the theory of measure approximation

Spline PDF Theory revisited

Theorem 2 (Ditkowski, Fibich, AS '18):

Let $\Omega = [0,1]^d$, let $f \in C^{m+1}(\Omega)$ with $|\nabla f| > a > 0$, let α be uniformly distributed in Ω,

and

Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree tensor-product spline interpolant on N^d equi-distributed points. Then

$$\|p - p_N\|_1 \leq KN^{-\frac{m}{d}},$$

Problem I – "arbitrary" derivative condition from application standpoint
Problem II – spline approximate derivatives in L^∞, other methods do not

Spline PDF Theory revisited

Theorem 2 (Ditkowski, Fibich, AS '18):

Let $\Omega = [0,1]^d$, let $f \in C^{m+1}(\Omega)$ with $|\boldsymbol{\nabla f}| > \boldsymbol{a} > \mathbf{0}$, **let $\boldsymbol{\alpha}$ be uniformly distributed in $\boldsymbol{\Omega}$,**

and

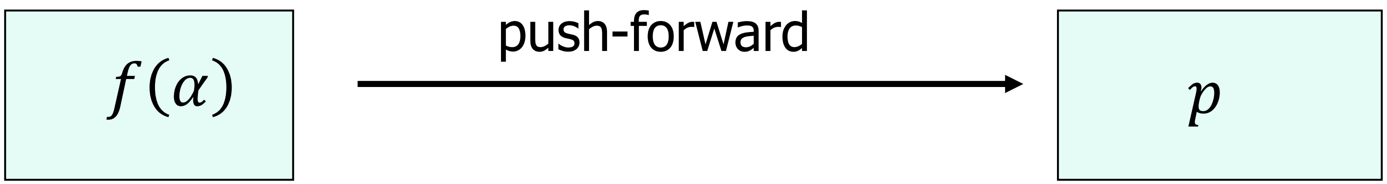
Let p and p_N be the probability density functions (PDF) of $f(\alpha)$ and its m-degree tensor-product spline interpolant on N^d equi-distributed points. Then

$$\|p - p_N\|_1 \leq K N^{-\frac{m}{d}},$$

Problem III – uniform measure (or absolutely continuous)
Problem IV – Omega is a box (compact)

Approximation-based estimation

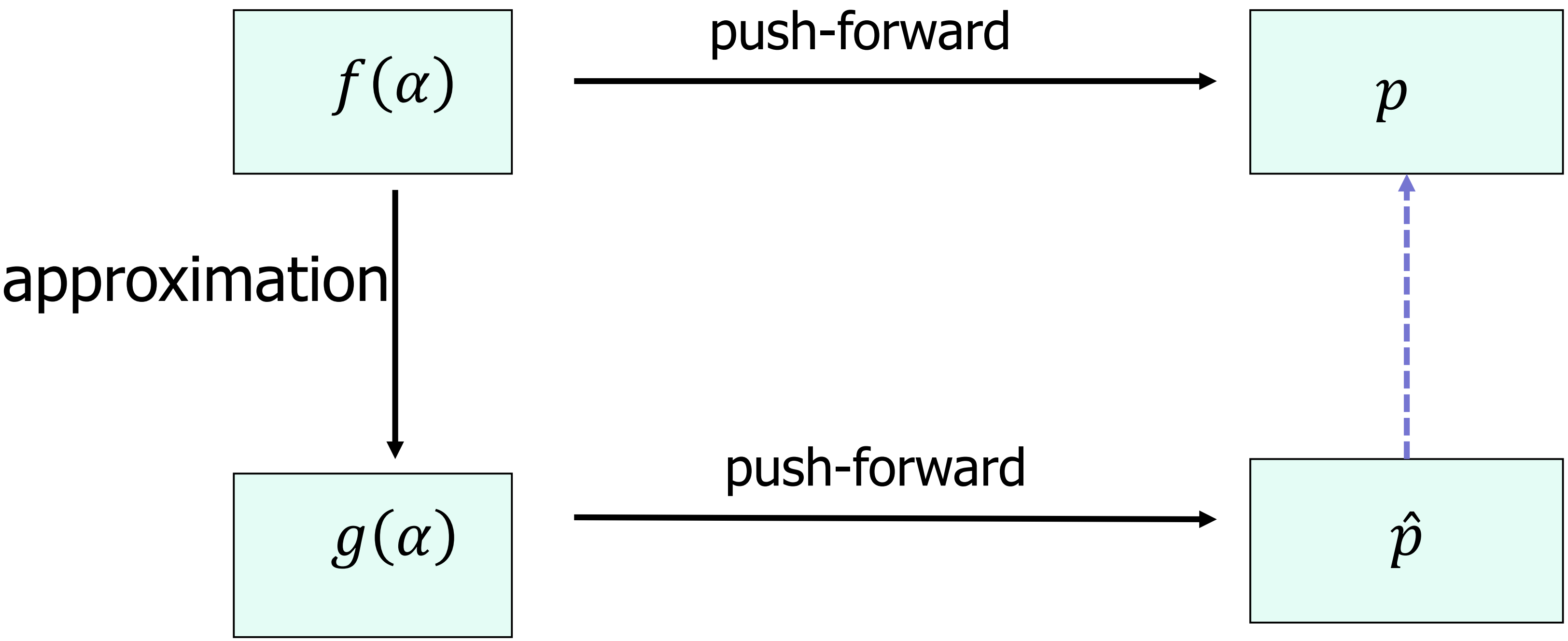
$\alpha \sim \varrho \; probability \; measure$



density of interest

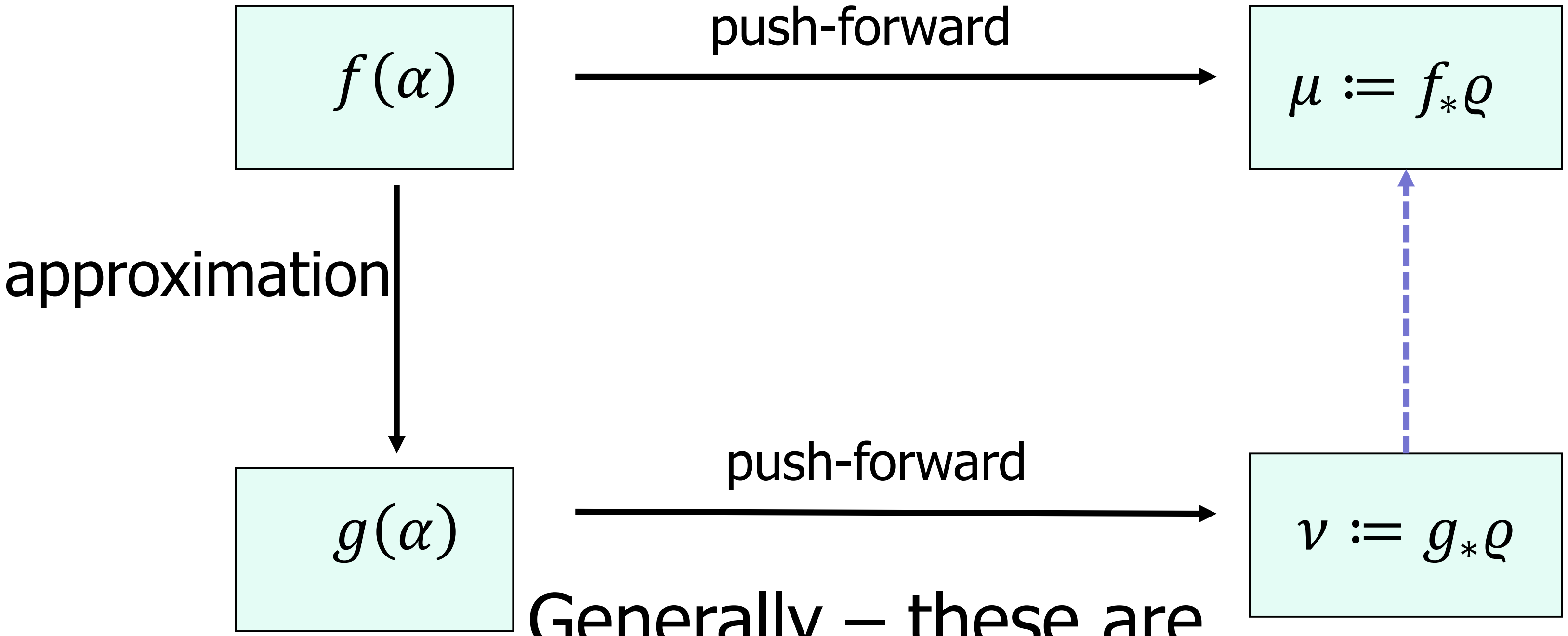
Approximation-based estimation

$\alpha \sim \varrho\ probability\ measure$



Approximation-based estimation

$\alpha \sim \varrho$ *probability measure*

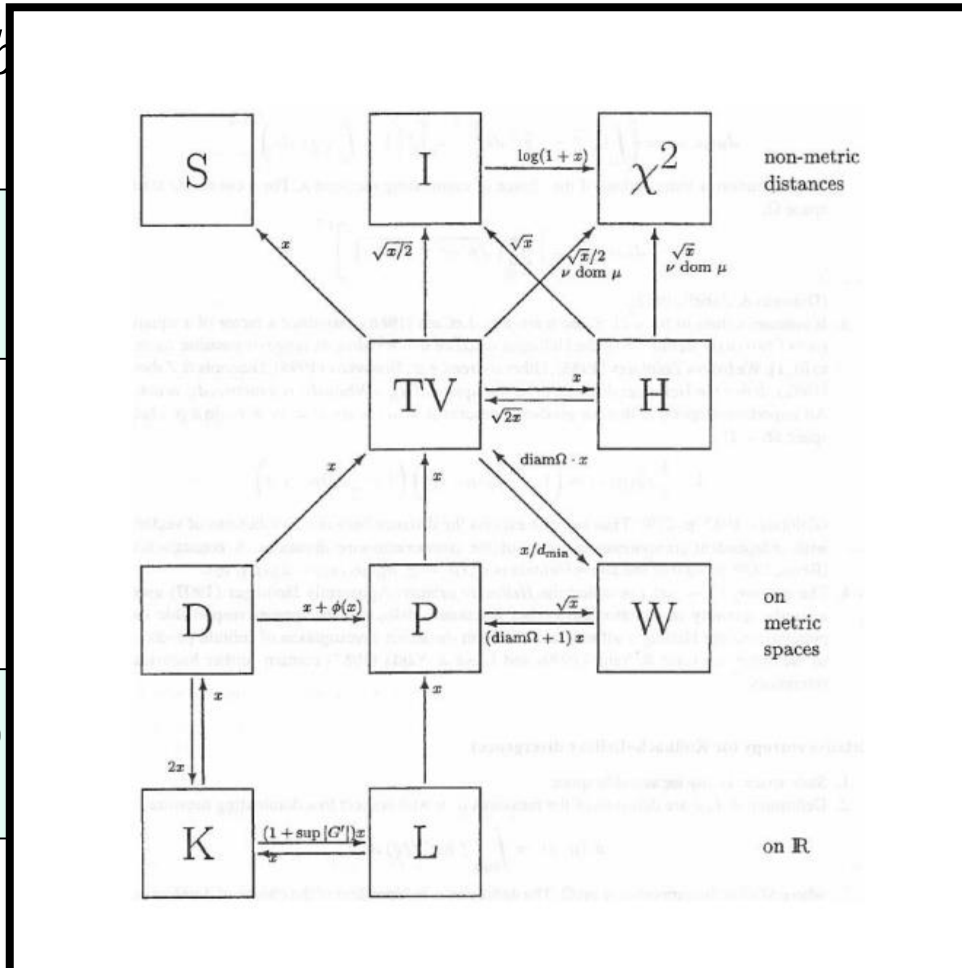


Generally – these are measures, not densities

Approximation-based estimation

$\alpha \sim \varrho$ *prob*

[Gibbs and Su, Int. Stats. Rev. 2002]



$f(\alpha)$

approximation

$g(\alpha)$

$f_*\varrho$

$g_*\varrho$

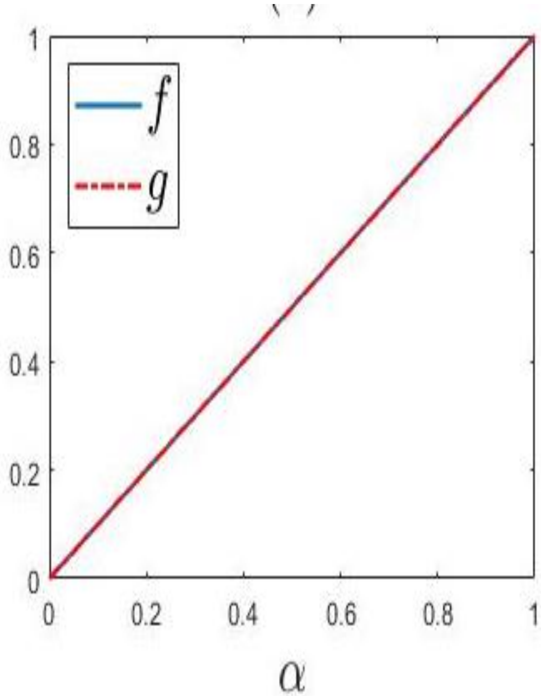
How should the difference between μ and ν be measured?

Is PDF the right way to measure?

A numerical example:

$$f(\alpha) = \alpha; \qquad g(\alpha) = \alpha + 10^{-3}\sin(100\alpha)$$

$$\Rightarrow ||f - g||_{L^q} \sim 10^{-3}$$



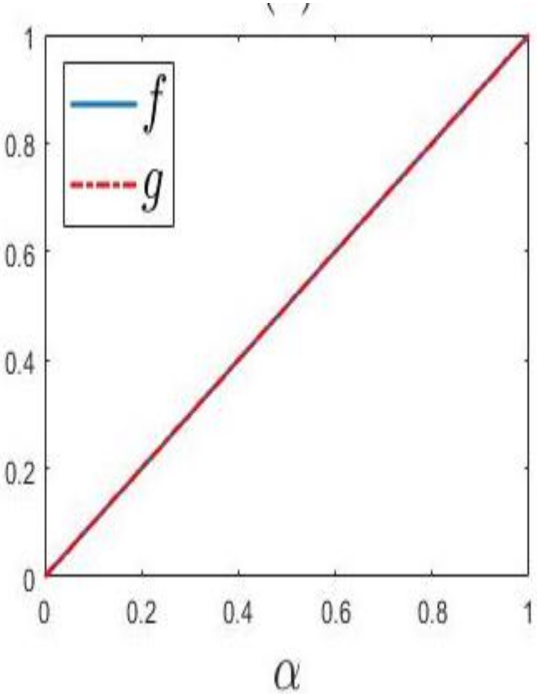
Is PDF the right way to measure?

A numerical example:

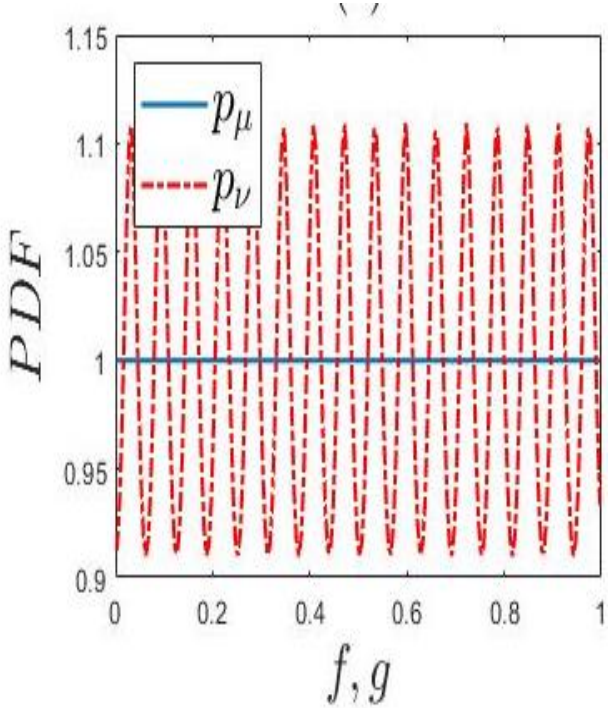
$$f(\alpha) = \alpha; \qquad g(\alpha) = \alpha + 10^{-3}\sin(100\alpha)$$

$$\varrho = Lebesgue, \qquad \mu := f_*\varrho; \quad \nu := g_*\varrho$$

$10^{-3} error$



$10^{-1} error$



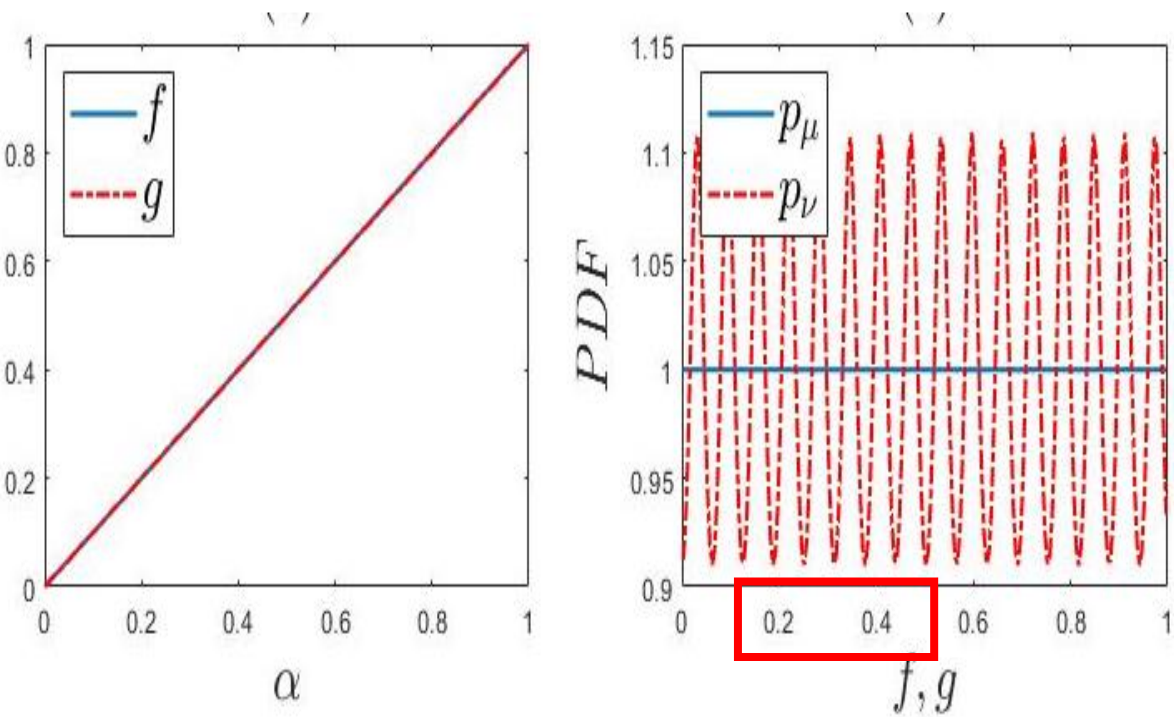
This difference can be made <u>arbitrarily</u> large

Is PDF the right way to measure?

A numerical example:

$$f(\alpha) = \alpha; \qquad g(\alpha) = \alpha + 10^{-3}\sin(100\alpha)$$

$$\varrho = Lebesgue, \qquad \mu := f_*\varrho; \quad \nu := g_*\varrho$$



PDFs are different, but

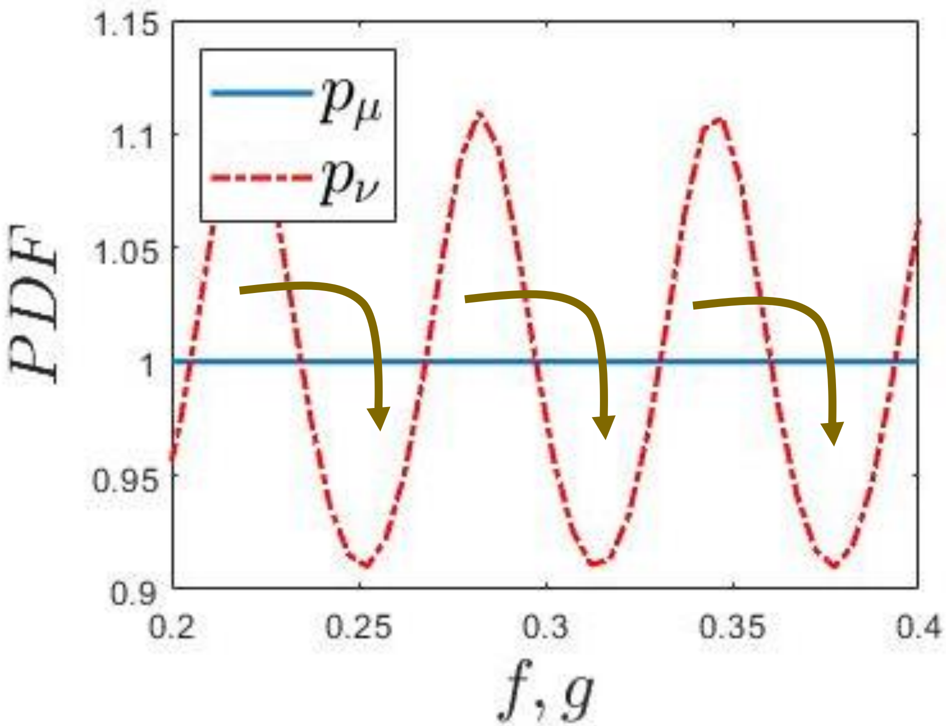
$$\mu([0.2,0.4]) \approx \nu([0.2,0.4])$$

Is PDF the right way to measure?

A numerical example:

$$f(\alpha) = \alpha; \qquad g(\alpha) = \alpha + 10^{-3}\sin(100\alpha)$$

$$\varrho = Lebesgue, \qquad \mu := f_*\varrho; \quad \nu := g_*\varrho$$



PDFs are different, but

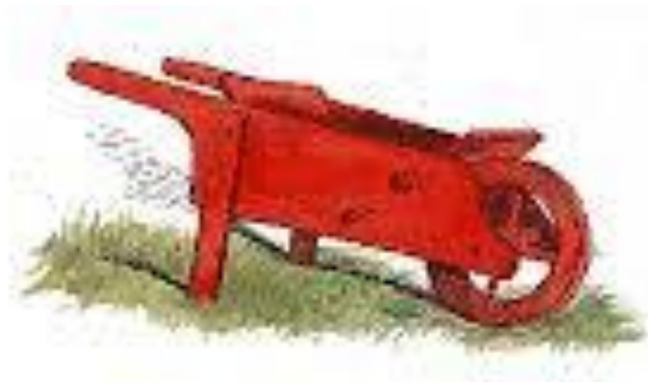
$$\mu([0.2,0.4]) \approx \nu([0.2,0.4])$$

"transfer" mass on a local scale

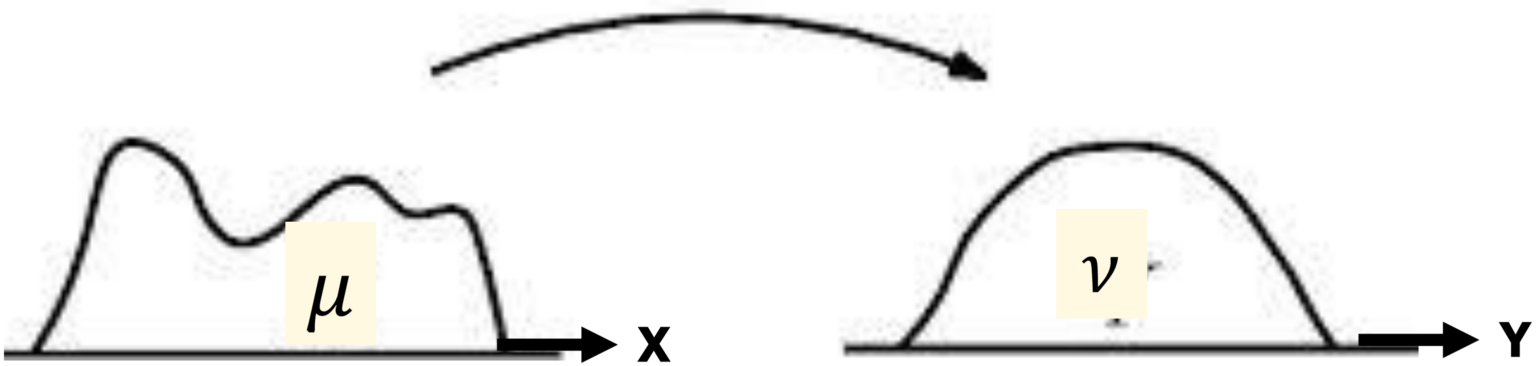
Underlying Theory – Wasserstein Metrics

$$W_p(\mu, \nu) = \left[\inf \int_{R\times R} |x-y|^p d\gamma(x,y)\right]^{\frac{1}{p}}$$

Such that μ, ν are marginals of γ



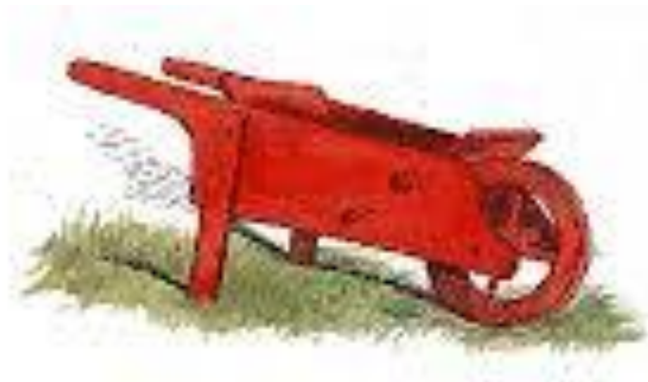
<u>Intuitively</u> (for p=1)
a transport plan: move
$\gamma(x, y)$ mass over
$|x-y|$ distance



Underlying Theory – Wasserstein Metrics

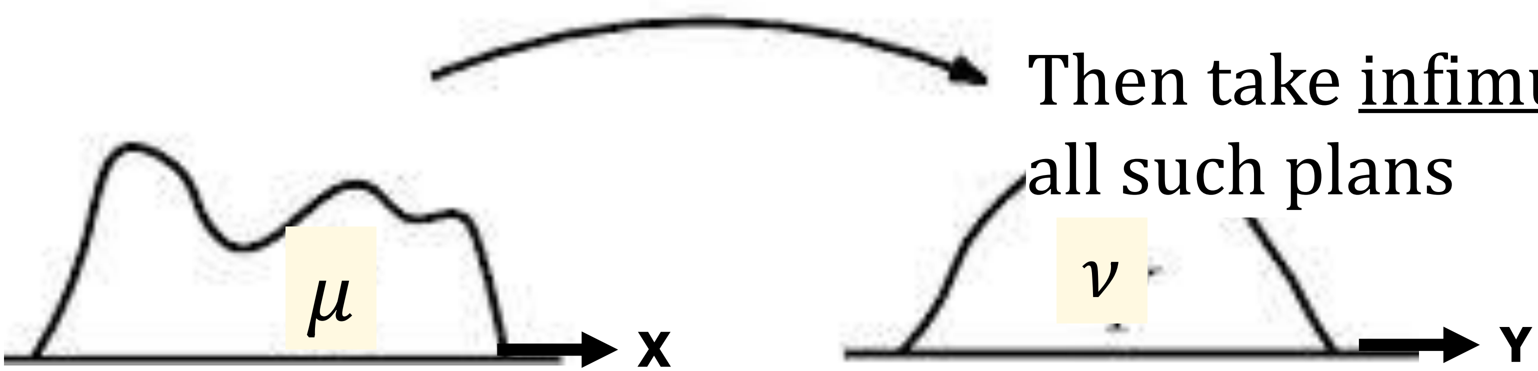
$$W_p(\mu,\nu) = \left[\inf \int_{R\times R} |x-y|^p \, d\gamma(x,y)\right]^{\frac{1}{p}}$$

Such that μ, ν are marginals of γ



<u>Intuitively</u> (for p=1)
a transport plan: move $\gamma(x,y)$ mass over $|x-y|$ distance

Then take <u>infimum</u> over all such plans



Indeed – Wasserstein theory **is** simple

Theorem 3 (AS '19):

Let $\Omega \subseteq R^d$, let $f, g \in C\,(\bar{\boldsymbol{\Omega}})$, and let ϱ be a Borel measure and $\mu = f_* \varrho, \ \nu = g_* \varrho$

1. $W_p(\mu, \nu) \leq ||f - g||_\infty$

i.e., pointwise accuracy guarantees Wasserstein accuracy

Indeed – Wasserstein theory **is** simple

Theorem 3 (AS '19):

Let $\Omega \subseteq R^d$, let $f, g \in C\,(\overline{\mathbf{\Omega}})$, and let ϱ be a Borel measure and $\mu = f_*\varrho,\ \nu = g_*\varrho$

1. $W_p(\mu, \nu) \leq ||f - g||_\infty$
2. $W_p(\mu, \nu) \leq ||f - g||_p$ *(if Ω is bounded)*

i.e., L^p accuracy guarantess Wasserstein accuracy

Indeed – Wasserstein theory **is** simple

Theorem 3 (AS ’19):

Let $\Omega \subseteq R^d$, let $f, g \in C\,(\overline{\boldsymbol{\Omega}})$, and let ϱ be a Borel measure and $\mu = f_* \varrho,\ \nu = g_* \varrho$

1. $W_p(\mu, \nu) \leq ||f - g||_\infty$
2. $W_p(\mu, \nu) \leq ||f - g||_p$ *(if Ω is bounded)*
3. $W_p(\mu, \nu) \leq C(p, q)||f - g||_q^{\frac{p}{q+p}} \cdot ||f - g||_\infty^{\frac{q}{q+p}}$ for all $q \geq 1$

Indeed – Wasserstein theory **is** simple

Theorem 3 (AS ’19):

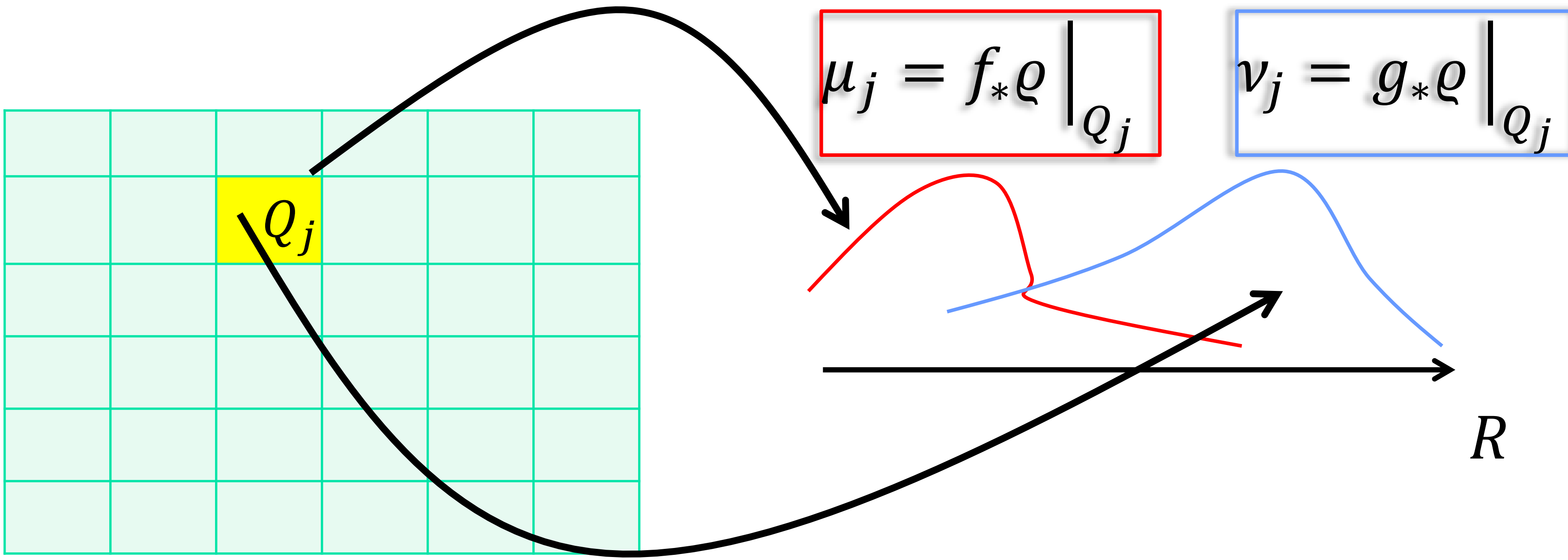
Let $\Omega \subseteq R^d$, let $f, g \in C\,(\overline{\boldsymbol{\Omega}})$, let ϱ be a Borel measure and $\mu = f_* \varrho,\ \nu = g_* \varrho$

1. $W_p(\mu, \nu) \leq ||f - g||_\infty$
2. $W_p(\mu, \nu) \leq ||f - g||_p$ *(if Ω is bounded)*
3. $W_p(\mu, \nu) \leq C(p, q)||f - g||_q^{\frac{p}{q+p}} \cdot ||f - g||_\infty^{\frac{q}{q+p}}$ for all $q \geq 1$

- No conditions on the underlying measure and domain (**=many noise models**)
- No derivative approximation conditions
- Every L^q convergence works (**=many possible approximation methods**)

Proof sketch

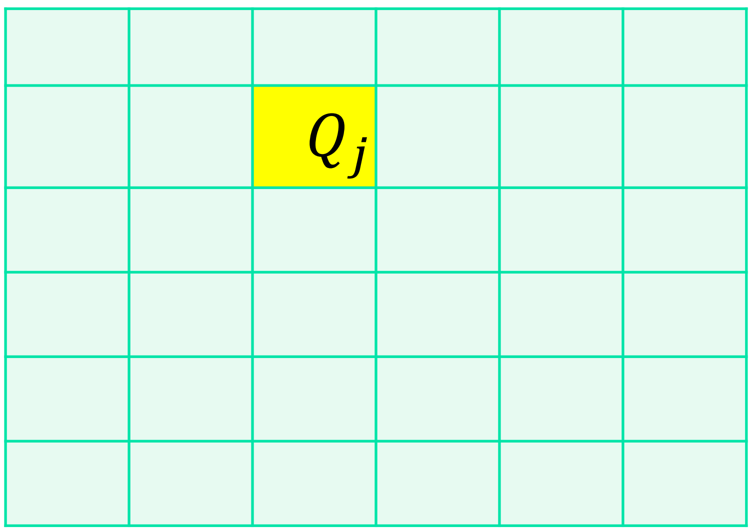
Here – Ω is a cube, ϱ is Lebesgue



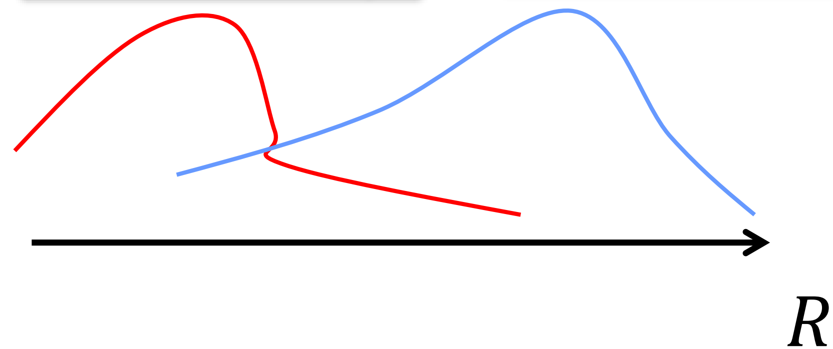
Step I – push forward a small cube Q_j to define to measures (of same mass) on R

Proof sketch

Here – Ω is a cube, ϱ is Lebesgue



$$\mu_j = f_*\varrho\Big|_{Q_j} \qquad \nu_j = g_*\varrho\Big|_{Q_j}$$



Step II – for $\varepsilon > 0$, by continuity, if $diam(Q_j) < \delta$

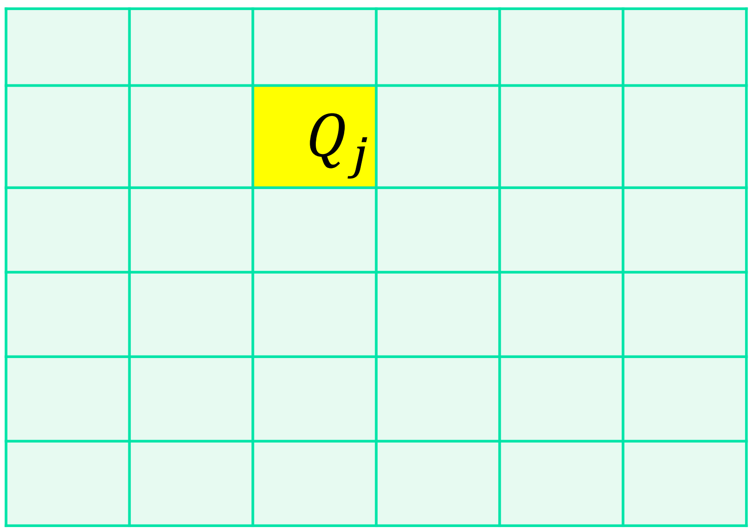
Then $|f(x) - f(y)|, \ |g(x) - g(y)| \leq \varepsilon$

And so for <u>any transport,</u>

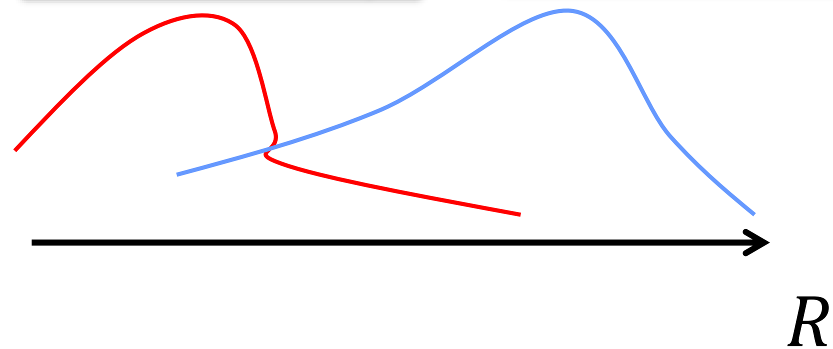
the mass ε^d travels a distance $\leq ||f - g||_{L^\infty} + o(\epsilon)$

Proof sketch

Here – Ω is a cube, ϱ is Lebesgue



$$\mu_j = f_*\varrho\Big|_{Q_j} \qquad \nu_j = g_*\varrho\Big|_{Q_j}$$



Step II – for $\varepsilon > 0$, by continuity, if $diam(Q_j) < \delta$

Then $|f(x) - f(y)|, \ |g(x) - g(y)| \leq \varepsilon$

And so for <u>any transport,</u>

the mass ε^d travels a distance $\leq ||f - g||_{L^\infty} + o(\epsilon)$

Step III – this is true for *all cubes*, for *any* $\varepsilon > 0$

Agenda

- PDF approximation
  - Is moment-estimation sufficient?
  - An algorithm & convergence results
- Transport-theory point of view

Back to the Uncertainty-quantification problem

Pause, why Wasserstein?

- The distance between PDFs is natural and intuitive to use…
  - But difficult to work with.

- Wasserstein-theory is easier to work with, better approximation results…
  - But is it useful for applications?

Wasserstein and CDFs

The CDF bounds are a result of a wider theory for **Wasserstein Metrics**, since

$$W_1(\mu, \nu) = ||F_\mu - F_\nu||_1$$

[Salvemini ‘43, Vallender ‘74]

Cumulative distribution function (CDF)

$$F_\mu(y) := \mu([y, \infty))$$

Wasserstein and CDFs

The CDF bounds are a result of a wider theory for **Wasserstein Metrics**, since

$$W_1(\mu, \nu) = ||F_\mu - F_\nu||_1$$

[Salvemini '43, Vallender '74]

Cumulative distribution function (CDF)

$$F_\mu(y) := \mu\big([y, \infty)\big)$$

Theorem 3 – for CDFs (AS '19):

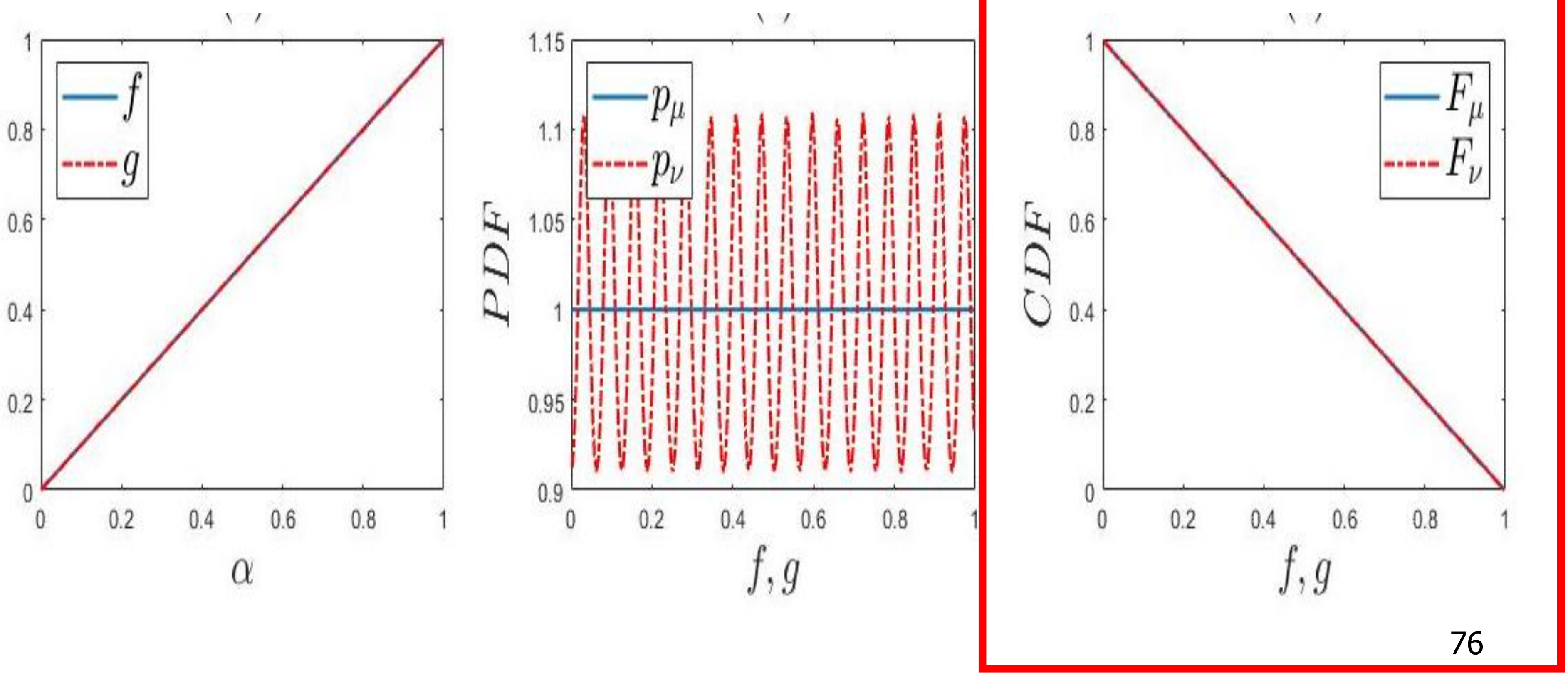
Let $\Omega \subseteq R^d$, let $f, g \in C\ (\overline{\boldsymbol{\Omega}})$, and let ϱ be a Borel measure

1. $||F_\mu - F_\nu||_1 \leq ||f - g||_\infty$
2. $||F_\mu - F_\nu||_1 \leq ||f - g||_1$ *(if Ω is bounded)*
3. $||F_\mu - F_\nu||_1 \leq ||f - g||_q^{\frac{1}{q+1}} \cdot ||f - g||_\infty^{1-\frac{1}{q+1}}$ for all $q \geq 1$

Is CDF the right way to measure?

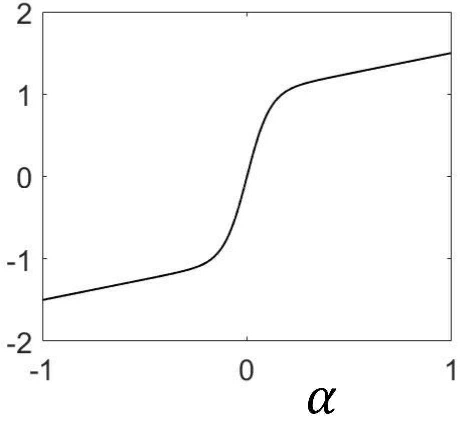
A numerical example:

$$f(\alpha) = \alpha; \qquad g(\alpha) = \alpha + 10^{-3}\sin(100\alpha)$$

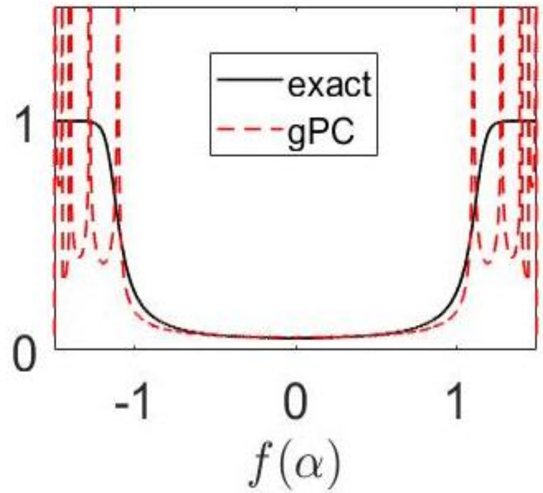


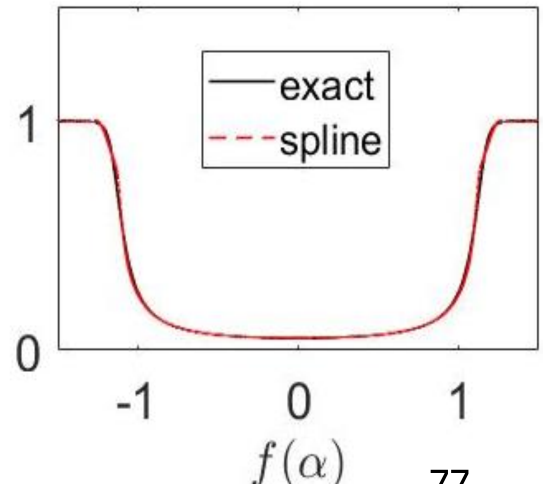
Numerical example - revisited

$f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim Uniform\,[-1,1]$



PDF approximation, $N = 12$





Numerical methods

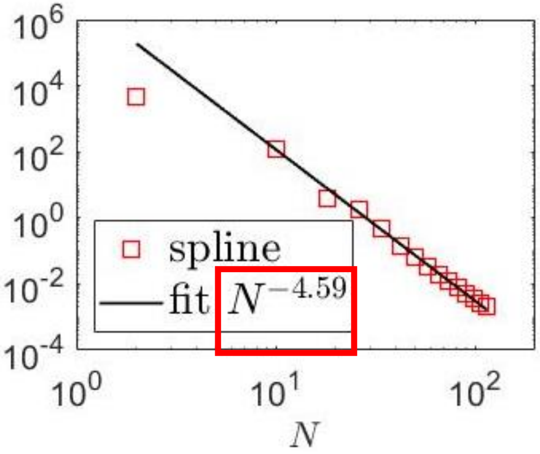
Theorems 4-5 (AS ’19):

Under general smoothness conditions

1. For *m* order spline with spacing *h>0 , then*

$$||F_\mu - F_\nu||_1 \leq K h^{m+1}$$

CDF error - spline



Numerical methods

Theorems 4-5 (AS '19):

Under general smoothness conditions

1. For *m* order spline with spacing *h>0 , then*

$$||F_\mu - F_\nu||_1 \le K h^{m+1}$$

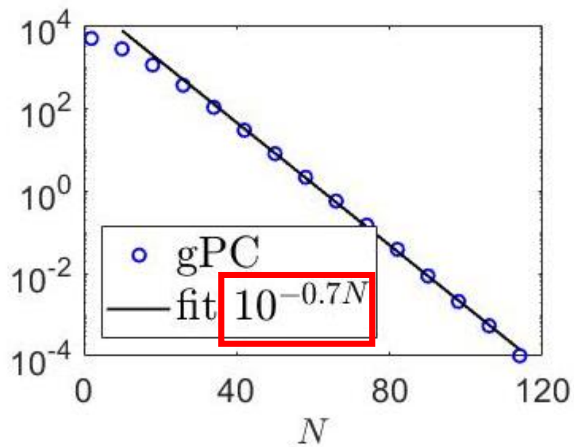
2. For analytic function f and gPC of order *N*

$$||F_\mu - F_\nu||_1 \le C exp(-\gamma N)$$

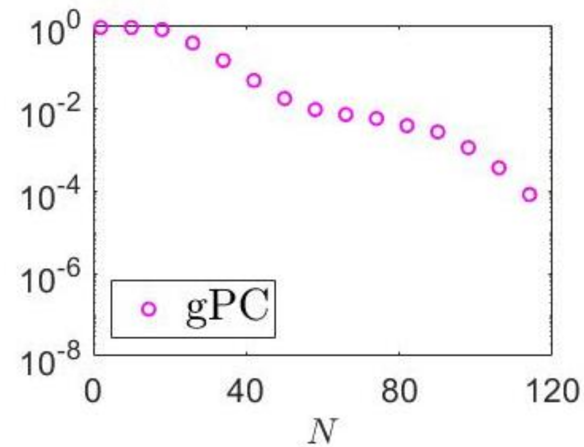
gPC result – in sharp contrast to PDF approximation

CDF error - spline

CDF error - gPC

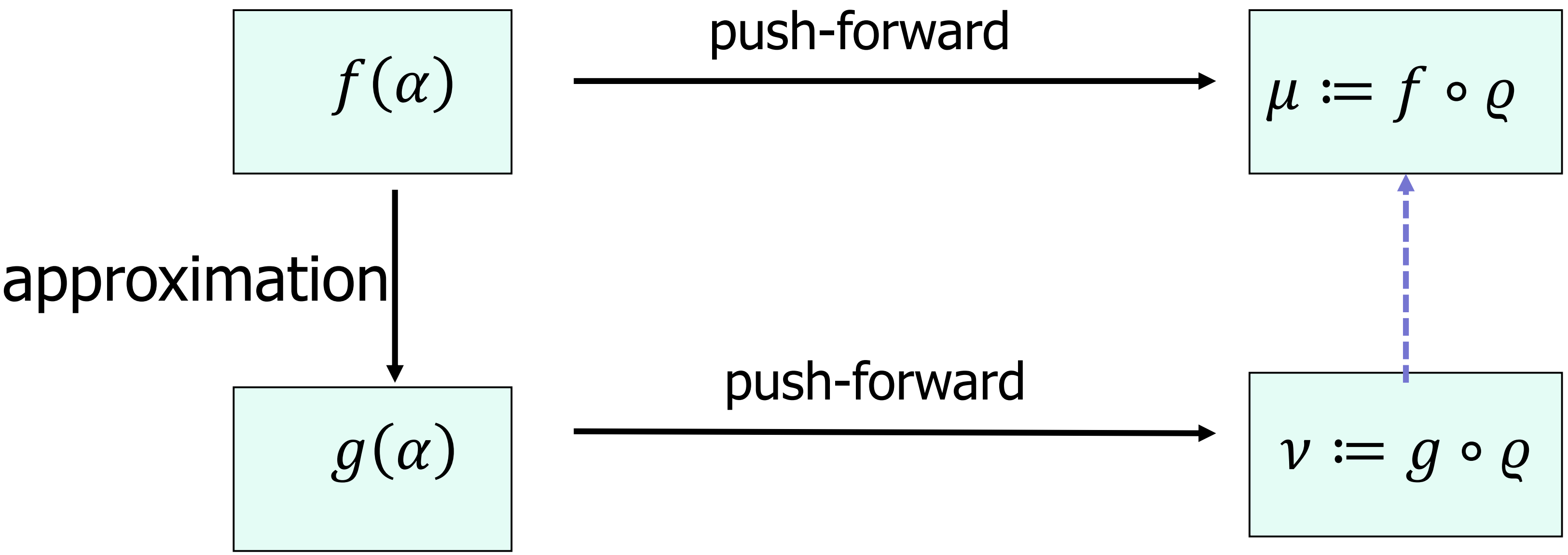


PDF error - gPC



Lower bounds

$\alpha \sim \varrho$ *probability measure*



So far

We bounded $W_p(\mu, \nu)$ by $||f - g||_{L^q}$ **from above**

What about lower bounds?

Lower bounds – key idea

Wasserstein metric is defined as an infimum, so any transport plan provides an upper bound

Can it be restated as a supremum?

Lower bounds – key idea

Wasserstein metric is defined as an infimum, so any transport plan provides an upper bound
Can it be restated as a supremum?

Monge Kantorovich–

$$W_1(\mu,\nu) = \sup\left\{\int_{\mathrm{R}} w(d\mu - d\nu) \mid Lip\,(w) \leq 1\right\}$$

Loeper (2005) & Peyre (2018)
Under certain smoothness assumptions

$W_2(\mu,\nu) \sim ||\mu - \nu||_{\dot{H}^{-1}}$ (supremum functional on $w \in \dot{H}^1$)

Lower bounds – proof sketch

Monge Kantorovich–

$$W_1(\mu,\nu) = \sup\left\{\int_{\mathrm{R}} w(d\mu - d\nu) \mid Lip\,(w) \le 1\right\}$$

Loeper (2005) & Peyre (2018)
Under certain smoothness assumptions

$W_2(\mu,\nu) \sim ||\mu - \nu||_{\dot{H}^{-1}}$ (supremum functional on $w \in \dot{H}^1$)

Proof sketch

choose $w(z) = c_k y^k$ and recover moments by change of variables, e.g.,
$\int_{\mathrm{R}} y(d\mu - d\nu) = \int_{\Omega} f(\alpha)d\varrho(\alpha) - \int_{\Omega} g(\alpha)d\varrho(\alpha)$

And similarly for $W_2(\mu,\nu)$ …

Lower bounds – proof sketch

Monge Kantorovich–

$$W_1(\mu,\nu) = \sup\left\{\int_{\mathrm{R}} w(d\mu - d\nu) \mid Lip\,(w) \leq 1\right\}$$

Loeper (2005) & Peyre (2018)
Under certain smoothness assumptions

$W_2(\mu,\nu) \sim ||\mu-\nu||_{\dot{H}^{-1}}$ (supremum functional on $w \in \dot{H}^1$)

Theorems 5&6 (AS '19):

Let $\Omega \subseteq R^d$ be bounded, let $f, g \in C\,(\overline{\boldsymbol{\Omega}})$, let ϱ be a Borel measure

$$W_1(\mu,\nu) \geq |E_\varrho f - E_\varrho g\,| \,,$$

On an interval with Lebesgue measure-

$$W_2(\mu,\nu) \geq C(f,k)\left|E_\varrho f^k - E_\varrho g^k\right| \qquad k \geq 1$$

Conclusions

- Convergence of moments and in L^2 does not guarantee convergence in PDFs
- Spline perform well for PDF approximation
  - With theoretical guarantees in all dimensions.
- Convergence in CDF is “better-behaved” than in PDFs
  - Most popular methods converge in CDF, but not always in PDF
  - Underlying theory – Wasserstein metric

Thank you!

<u>References</u>

- A. Sagiv The Wasserstein Distances Between Pushed-Forward Measures with Applications to Uncertainty Quantification arXiv 1902:05451 (under review at **Communications in Mathematical Sciences**)

- A. Sagiv, A. Ditkowski, G. Fibich Density estimation in uncertainty propagation problems using a surrogate model arXiv 1803.10991 (under review)