An Optimal Transport Perspective on Uncertainty Quantification

\[
f(\alpha) \quad \xrightarrow{\text{push-forward}} \quad \mu := f \circ \varrho
\]

\[
g(\alpha) \quad \xrightarrow{\text{push-forward}} \quad \nu := g \circ \varrho
\]

\(\alpha \sim \varrho\) probability measure

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 Schrödinger equation

**initial condition**

\[ \psi_0(x, y) \]

**PDE model**

**output**

\[ \psi(z, x, y) \]
Each laser shot is different

**random initial condition**

\[ \psi_0(x, y; \alpha) \]

\( \alpha \) - noise parameter

**PDE model**

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

**random output**

\[ \psi(z, x, y; \alpha) \]

Two motivation slides: nonlinear laser propagation

[Sagiv, Ditwkoski, Fibich, Opt. Exp. 2017]
Each laser shot is different

random initial condition \( \psi_0(x, y; \alpha) \)

\( \alpha \) - noise parameter

random output \( \psi(z, x, y; \alpha) \)

What kind of statistics do we want to compute?

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

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

Each laser shot is different

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

\( \alpha \) - noise parameter

PDE model

random output

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

Two motivation slides:

nonlinear laser propagation

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

What kind of statistics do we want to compute?

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

Density estimation
Probability Density Function (PDF) of some “quantity of interest” \( f(\psi) \)
Why study the PDF — Examples from optics

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

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

- \( \alpha \) - noise parameter

Random amplitude

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

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

[\( \backslash \text{w Patwardhan et al, PRA 2019} \)]
General standard nonlinear PDE settings

Initial value problem

\[
\begin{align*}
    u_t(t, x) &= Q(x, u)u \\
    u(t = 0, x) &= u_0(x)
\end{align*}
\]

- “quantity of interest” (model output) \( f(u(t, x)) \)
- e.g., \( f = u(t_i, x_i) \), \( f = \int dx \ |u|^2 \), ...
- \( u \) & \( f(u) \) are evaluated **numerically**
Initial value problem with randomness (both i.c. $u_0$ and operator $Q$)

\[
\begin{align*}
&\begin{aligned}
u_t(t, x; \alpha) &= Q(x, u; \alpha)u \\
u(t = 0, x; \alpha) &= u_0(x; \alpha)
\end{aligned}
\end{align*}
\]

- $\alpha$ distributed according to a known measure
- “quantity of interest” (model output) $f(\alpha) := f(u(t, x; \alpha))$
  - e.g., $f = u(t_i, x_i)$, $f = \int dx |u|^2$, ...
- $u$ & $f(u)$ are evaluated **numerically**
General Settings – Nonlinear PDE with randomness

Initial value problem with randomness

\[
\begin{align*}
  u_t(t, x; \alpha) &= Q(x, u; \alpha)u \\
  u(t = 0, x; \alpha) &= u_0(x; \alpha)
\end{align*}
\]

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

How to approximate the PDF of $f(\alpha)$, 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

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

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

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

Constraints:
- Can only compute $f(\alpha_j)$ for a given $\alpha_j$
- Computation of $f(\alpha_j)$ is expensive (e.g., solving the (3+1)dimensional NLS)
  - Can only use a small sample $\{f(\alpha_1), ..., f(\alpha_N)\}$
Standard statistical methods

**Step I** – draw i.i.d. samples $\alpha_1, ..., \alpha_N$

**Step II** – compute the samples $\{f(\alpha_1), ..., f(\alpha_N)\}$

**Moment estimation**
- Monte-Carlo $E_\alpha[f] \approx \frac{1}{N} \sum_{n=1}^{N} f_n$
- ...

**Density (PDF) estimation**
- Histogram method
- Kernel density estimators (KDE)
- ...

\[ \approx \frac{1}{N} \sum_{n=1}^{N} f_n \]
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 \( E_\alpha[f] \approx \frac{1}{N} \sum_{n=1}^{N} f_n \)
  - …

- **Density (PDF) estimation**
  - Histogram method
  - Kernel density estimators (KDE)
  - …

- Poor approximations for small \( N \) \( \frac{1}{\sqrt{N}} \) 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 $E_\alpha[f] \approx \frac{1}{N} \sum_{n=1}^{N} f_n$
- ...

Density (PDF) estimation
- Histogram method
- Kernel density estimators (KDE)
- ...

Can we improve?

- Poor approximations for small N
  e.g. Histogram method with $N=10$ samples

Exact PDF
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 $E_{\alpha}[f] \approx \frac{1}{N} \sum_{n=1}^{N} f_n$
- ...

**Density (PDF) estimation**

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

**Can we improve?**

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

These assumptions underly many studies in uncertainty quantification (UQ), specifically in uncertainty propagation
Approximation-based estimation

\[ f(\alpha) \quad \text{moment, PDF} \quad E_\alpha[f], p \]

\[ p \text{ is the PDF of } f \]

Quantity of interest
Approximation-based estimation

- $f(\alpha)$
- $\alpha$ moment, PDF
- $E_\alpha[f], p$
- $p$ is the PDF of $f$

- Unknown explicitly
- Each evaluation is computationally expensive

- cannot take a large sample
Approximation-based estimation

- Approximation using few samples:
  - $f(\alpha)$
  - $g(\alpha)$

- moment, PDF
  - cannot take a large sample

- known function, “cheap” evaluation
  - $E_\alpha[f], p$

- $E_\alpha[g], \hat{p}$ are known exactly

- $p$ is the PDF of $f$

- $\hat{p}$ is the PDF of $g$
Approximation-based estimation

Questions

- Which approximation \( g(\alpha) \approx f(\alpha) \) should be used?
- How small are \( E_\alpha[f] - E_\alpha[g] \) and \( ||p - \hat{p}|| \)?
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 \( L^2 \))**

\[
E_\alpha[f] - E_\alpha[f_N] = O(e^{-\gamma N}), \quad N \gg 1 \quad f \text{ is analytic}
\]

[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 $L^2$)
  $$E_\alpha[f] - E_\alpha[f_N] = O(e^{-\gamma N}), \quad N \gg 1 \quad f \text{ is analytic}$$

But,

**PDF estimation**
No theory for $\|p - p_N\|$

Will it work in practice?
Example – gPC fails at PDF estimation

\[ f = \tanh(9\alpha) + \frac{\alpha}{2}, \quad \alpha \sim \text{Uniform } [-1, 1] \]
Example – gPC fails at PDF estimation

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

PDF approximation, \( N = 12 \) samples

Why does it fail?
Example – gPC fails at PDF estimation

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

**PDF approximation, \( N = 12 \) samples**

**Derivative**

**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.

Artificial extremal points
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)

\( p \) is the PDF of \( f \)

\( N \rightarrow \infty \)

\( \hat{p} \) is the PDF of \( g \)
Theorem 1 (Ditkowsk, 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

$$\|p - p_N\|_1 \leq KN^{-m},$$
Theorem 1 (Ditkowski, Fibich, AS ’18):
Let $f \in C_{\text{piecewise}}^{m+1}([\alpha_{\min}, \alpha_{\max}])$ with $|f'| > \alpha > 0$, let $\alpha$ be distributed by $c(\alpha)d\alpha$ where $c \in C^1([\alpha_{\min}, \alpha_{\max}])$.

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

$$\|p - p_N\|_1 \leq KN^{-m},$$
Theorem 1 (Ditkowski, Fibich, AS ’18):

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

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

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

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

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

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, $\alpha \sim U(-1,1)$

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

$$= \ldots$$
PDF estimation

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

\[ \sigma(f) - \sigma(f_N) \]

PDF approximation, \( N = 12 \)

Statistically optimal
PDF estimation

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

PDF approximation, \( N = 12 \)
Coupled NLS example

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

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

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

Random elliptical beam –
\[
A_\pm(t = 0) = (1 + \alpha)C_\pm e^{-x^2}, \quad \alpha \sim U(-0.1, 0.1)
\]

PDF, N=64

[with Patwardhan et al, PRA, 2019]
**Coupled NLS example**

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

**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** -
\[
A_{\pm}(t = 0) = (1 + \alpha) C_{\pm} e^{-x^2}, \quad \alpha \sim U(-0.1, 0.1)
\]

[with Patwardhan et al, PRA, 2019]
Burgers equation – shock location

\[ u_t(t, x) + \frac{1}{2} (u^2)_x = \frac{1}{2} (\sin(x))_x \]

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

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

Distribution of random initial amplitude –

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

[compare Chen, Gottlieb, Hesthaven, JCP 2005]
Burgers equation – shock location

\[
\begin{align*}
    u_t(t, x) + \frac{1}{2} (u^2)_x &= \frac{1}{2} (\sin(x))_x \\
    \text{Initial condition: } u_0(x) &= \alpha \sin(x) \\
    \text{Shock location at } t \to \infty: \quad \alpha &= -\cos(X_s) \\
    \text{Distribution of random initial amplitude: } \quad \alpha(\nu) &= \begin{cases} 
        -1 + \sqrt{1 + 4\nu^2} & \nu \neq 0 \\
        \frac{2\nu}{\nu} & \nu = 0 
    \end{cases} \\
    \nu &\sim N(0, \sigma)
\end{align*}
\]

PDF, N=11

PDF, N=11

\[\|p - p_N\|_1 \sim N^{3.1}\]

[compare Chen, Gottlieb, Hesthaven, JCP 2005]
Can this problem be solved if the input noise is multi-dimensional?

(physically – multiple uncertain or noisy terms in the system)
Theorem 2 (Ditkowski, Fibich, AS ’18):
Let $\Omega = [0,1]^d$, let $f \in C^{m+1} (\Omega)$ with $|\nabla f| > a > 0$, let $\alpha$ be uniformly distributed in $\Omega$, 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

$$\|p - p_N\|_1 \leq KN^{-\frac{m}{d}},$$
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 $\alpha$ be uniformly distributed in $\Omega$, 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

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

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 (Ditkowskí, 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

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

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

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

**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

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

where $h>0$ is the maximal spacing between interpolation points
From $d=1$ to $d>1$:
“Same” result, more complicated proof

**Theorem 1 (Ditkowskii, 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

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

**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

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

where $h>0$ is the maximal spacing between interpolation points.

Under some conditions

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

**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
$$\|(f - s)^{(j)}\|_\infty \leq C_m(f) h^{m+1-j} \quad j = 0, \ldots, m - 1$$
where $h > 0$ is the maximal spacing between interpolation points

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

Different manifolds – more complications
2-dimensional example

\[ f(\alpha_1, \alpha_2) = \tanh \left( 6\alpha_1 \alpha_2 + \frac{\alpha_1}{2} \right) + \frac{\alpha_1 + \alpha_2}{2}, \quad \alpha_1, \alpha_2 \sim \text{Uni}(-1, 1), \quad i. i. d. \]
3 dimensional example

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

\[ \alpha_1, \alpha_2, \alpha_3 \sim \text{Uni}(-1,1), \quad \text{i.i.d.} \]
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
Theorem 2 (Ditkowski, Fibich, AS ’18):

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

Let $p$ and $p_N$ be the probability density functions (PDF) of $f(α)$ 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 $L^\infty$, other methods do not
Theorem 2 (Ditkowski, Fibich, AS ’18):

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

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

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

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

\[ \alpha \sim q \text{ probability measure} \]

\[ f(\alpha) \xrightarrow{\text{push-forward}} p \]

density of interest
Approximation-based estimation

\[ \alpha \sim q \text{ probability measure} \]

\[ f(\alpha) \rightarrow \text{push-forward} \rightarrow p \]

\[ g(\alpha) \rightarrow \text{push-forward} \rightarrow \hat{p} \]
Approximation-based estimation

$$\alpha \sim q$$ probability measure

Generally – these are measures, not densities
Approximation-based estimation

\[ \alpha \sim q \text{ probability measure} \]

- \[ f(\alpha) \]
- \[ g(\alpha) \]

approximation

- \[ f \ast q \]
- \[ g \ast q \]

How should the difference between \( \mu \) and \( \nu \) be measured?

Is PDF the right way to measure?

A numerical example:

\[ f(\alpha) = \alpha; \quad g(\alpha) = \alpha + 10^{-3} \sin(100\alpha) \]

\[ \Rightarrow ||f - g||_{L^q} \sim 10^{-3} \]
Is PDF the right way to measure?

A numerical example:

\[ f(\alpha) = \alpha; \quad g(\alpha) = \alpha + 10^{-3} \sin(100\alpha) \]
\[ q = \text{Lebesgue}, \quad \mu := f \ast q; \quad \nu := g \ast q \]

This difference can be made **arbitrarily large**.
Is PDF the right way to measure?

A numerical example:
\[ f(\alpha) = \alpha; \quad g(\alpha) = \alpha + 10^{-3} \sin(100\alpha) \]
\[ q = \text{Lebesgue}, \quad \mu := f \ast q; \quad \nu := g \ast q \]

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:

\[ f(\alpha) = \alpha; \quad g(\alpha) = \alpha + 10^{-3} \sin(100\alpha) \]
\[ \mathcal{Q} = \text{Lebesgue}, \quad \mu := f_*\mathcal{Q}; \quad \nu := g_*\mathcal{Q} \]

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

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

Such that \( \mu, \nu \) are marginals of \( \gamma \)

**Intuitively (for \( p=1 \))**

a transport plan: move \( \gamma(x, y) \) mass over \( |x - y| \) distance
Underlying Theory – Wasserstein Metrics

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

Such that \( \mu, \nu \) are marginals of \( \gamma \)

Intuitively (for \( p=1 \))

a transport plan: move \( \gamma(x, y) \) mass over \( |x - y| \) distance

Then take \text{infimum} over all such plans
Indeed – Wasserstein theory is simple

Theorem 3 (AS ’19):
Let $\Omega \subseteq \mathbb{R}^d$, let $f, g \in C(\Omega)$, and let $\rho$ be a Borel measure and $\mu = f_*\rho$, $\nu = g_*\rho$

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

1. $W_p(\mu, \nu) \leq \|f - g\|_\infty$
2. $W_p(\mu, \nu) \leq \|f - g\|_p \quad (if \ \Omega \ \text{is bounded})$

i.e., $L^p$ accuracy guarantees Wasserstein accuracy
Indeed – Wasserstein theory **is** simple

**Theorem 3 (AS ’19):**

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

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

2. $W_p(\mu, \nu) \leq \|f - g\|_p$ \hspace{1cm} (if $\Omega$ 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 $\Omega \subseteq \mathbb{R}^d$, let $f, g \in C(\overline{\Omega})$, let $\mathcal{Q}$ be a Borel measure and $\mu = f_*\mathcal{Q}$, $\nu = g_*\mathcal{Q}$

1. $W_p(\mu, \nu) \leq ||f - g||_\infty$
2. $W_p(\mu, \nu) \leq ||f - g||_p$ (if $\Omega$ is bounded)
3. $W_p(\mu, \nu) \leq C(p, q)||f - g||_{q+\frac{p}{q}}^{q+p} \cdot ||f - g||_{\infty}^{q+p}$ 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 – $\Omega$ is a cube, $\rho$ is Lebesgue

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

Here – $\Omega$ is a cube, $\varrho$ is Lebesgue

\[ \mu_j = f_* \varrho \bigg|_{Q_j} \quad \nu_j = g_* \varrho \bigg|_{Q_j} \]

Step II – for $\varepsilon > 0$, by continuity, if $\text{diam}(Q_j) < \delta$
Then $|f(x) - f(y)|, \ |g(x) - g(y)| \leq \varepsilon$
And so for any transport,
the mass $\varepsilon^d$ travels a distance $\leq ||f - g||_{L^\infty} + o(\varepsilon)$
Proof sketch

Here – \( \Omega \) is a cube, \( \varrho \) is Lebesgue

\[
\mu_j = f_\ast \varrho \bigg|_{Q_j} \quad \nu_j = g_\ast \varrho \bigg|_{Q_j}
\]

Step II – for \( \varepsilon > 0 \), by continuity, if \( \text{diam}(Q_j) < \delta \)
Then \( |f(x) - f(y)|, |g(x) - g(y)| \leq \varepsilon \)
And so for any transport,
the mass \( \varepsilon^d \) travels a distance \( \leq \|f - g\|_{L^\infty} + o(\varepsilon) \)
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)

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

**Theorem 3 – for CDFs (AS ’19):**

Let \( \Omega \subseteq \mathbb{R}^d \), let \( f, g \in C(\overline{\Omega}) \), and let \( \rho \) 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 \( \Omega \) is bounded)
3. \( \| F_\mu - F_\nu \|_1 \leq \| f - g \|_q^{\frac{1}{q+1}} \cdot \| f - g \|_\infty^{\frac{1}{q+1}} \) for all \( q \geq 1 \)
Is CDF the right way to measure?

A numerical example:

\[ f(\alpha) = \alpha; \quad g(\alpha) = \alpha + 10^{-3} \sin(100\alpha) \]
Numerical example - revisited

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

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
   $$\|F_\mu - F_\nu\|_1 \leq K h^{m+1}$$
Theorems 4-5 (AS ’19):
Under general smoothness conditions
1. For \( m \) order spline with spacing \( h>0 \), then
\[
\| F_\mu - F_\nu \|_1 \leq K h^{m+1}
\]
2. For analytic function \( f \) and gPC of order \( N \)
\[
\| F_\mu - F_\nu \|_1 \leq C \exp(-\gamma N)
\]
gPC result – in sharp contrast to PDF approximation
Lower bounds

\[ \alpha \sim q \text{ probability measure} \]

\[ f(\alpha) \xrightarrow{\text{push-forward}} \mu := f \circ q \]

\[ g(\alpha) \xrightarrow{\text{push-forward}} \nu := g \circ q \]

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

\textit{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—

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

Under certain smoothness assumptions

$$W_2(\mu, \nu) \sim \|\mu - \nu\|_{\dot{H}^{-1}} \quad \text{(supremum functional on } w \in \dot{H}^1)$$
Lower bounds – proof sketch

Monge Kantorovich–

\[ W_1(\mu, \nu) = \sup \left\{ \int_R w(d\mu - dv) \mid \text{Lip}(w) \leq 1 \right\} \]

Under certain smoothness assumptions

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

Proof sketch

choose \( w(z) = c_k y^k \) and recover moments by change of variables, e.g.,

\[ \int_R y(d\mu - dv) = \int_{\Omega} f(\alpha) d\sigma(\alpha) - \int_{\Omega} g(\alpha) d\sigma(\alpha) \]

And similarly for \( W_2(\mu, \nu) \) ...
Lower bounds – proof sketch

Monge Kantorovich–

\[
W_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}} w(d\mu - d\nu) \mid \text{Lip}(w) \leq 1 \right\}
\]

Under certain smoothness assumptions

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

Theorems 5&6 (AS ’19):
Let \( \Omega \subseteq \mathbb{R}^d \) be bounded, let \( f, g \in C(\overline{\Omega}) \), let \( \varrho \) be a Borel measure

\[
W_1(\mu, \nu) \geq |E_{\varrho}f - E_{\varrho}g|,
\]

On an interval with Lebesgue measure-

\[
W_2(\mu, \nu) \geq C(f, k)|E_{\varrho}f^k - E_{\varrho}g^k| \quad k \geq 1
\]
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!

References

• 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)