SAND2016-3697PE

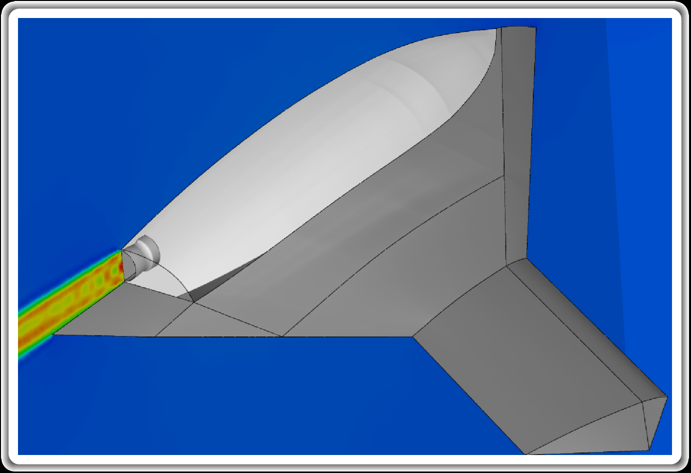
Compressed sensing and its role in designing aircraft nozzles in the presence of uncertainty

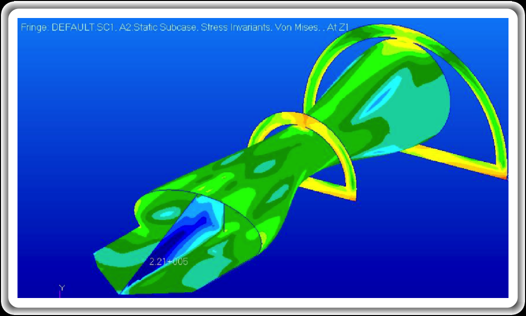
John D. Jakeman

Senior Member of Techincal Staff
Optimization and Uncertainty Quantification Department
Sandia National Laboratories
Albuquerque, NM USA

Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

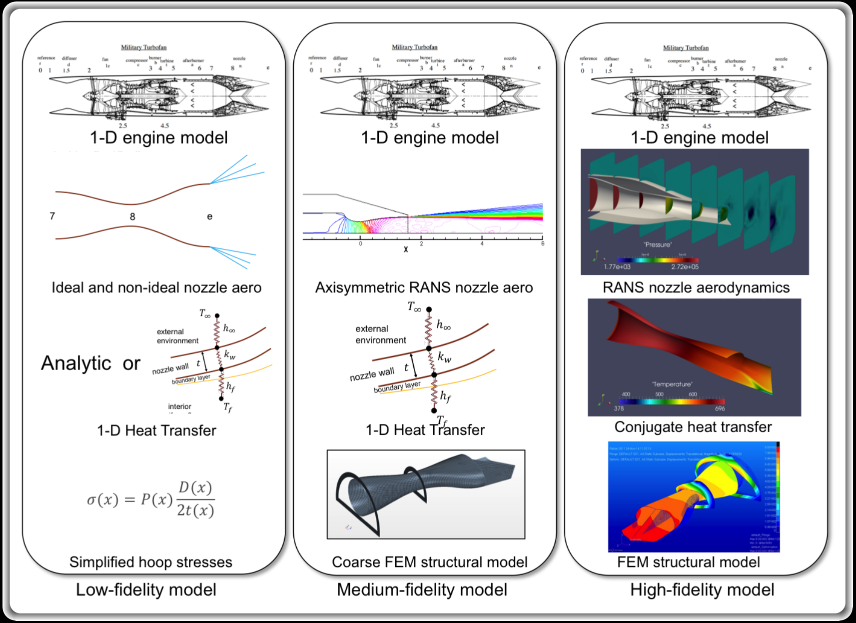
Goal: design nozzle propulsion systems of advanced aircraft that optimize aerodynamic performance, thermal and pressure loads, and fatigue





Problem Characteristics

- Nonlinear governing equations
- Computationally expensive simulations
- Large number of design variable
- Many sources of uncertainty
- Large number of uncertain variables



Military Turbofan
1-D engine model
Military Turbofan
1-D engine model
Military Turbofan
1-D engine model
7
8
e
Ideal and non-ideal nozzle aero
x
Axisymmetric RANS nozzle aero
"Pressure"
1.77e+03
2.72e+05
RANS nozzle aerodynamics
Analytic or
T_∞
h_∞
external environment
k_w
nozzle wall
t
boundary layer
h_f
interior
T_f
1-D Heat Transfer
T_∞
h_∞
external environment
k_w
nozzle wall
t
boundary layer
h_f
T_f
1-D Heat Transfer
"Temperature"
400
500
600
378
696
Conjugate heat transfer
$\sigma(x) = P(x)\frac{D(x)}{2t(x)}$
Simplified hoop stresses
Coarse FEM structural model
FEM structural model
Low-fidelity model
Medium-fidelity model
High-fidelity model

Design is expensive

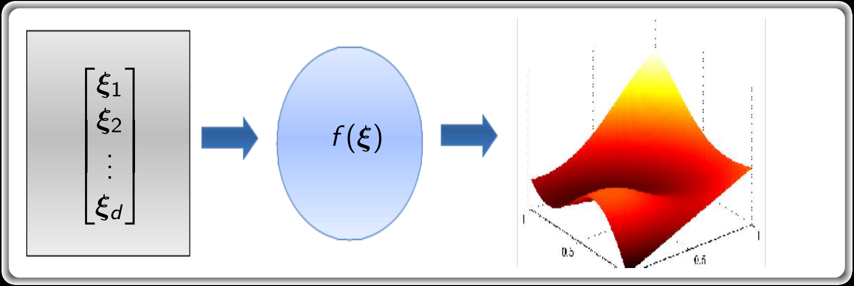
Optimization involves many evaluations of model and gradient

At each step of the design we must quantify certain statistics of uncertainty

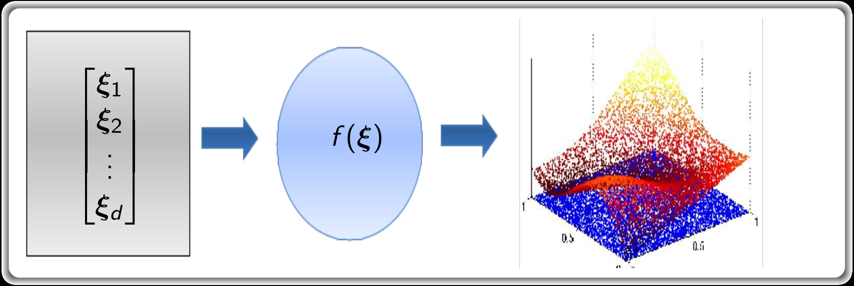
A fundamental challenge facing UQ

Can we capture the salient features of a high-fidelity, high-dimensional model from limited simulation data

Given distributions on the input data we can calculate statistical moments, distributions, etc. of the QOIs

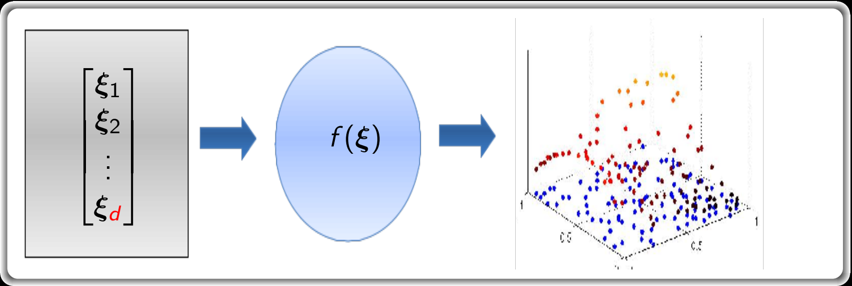


We usually must sample from the input distributions to calculate statistics



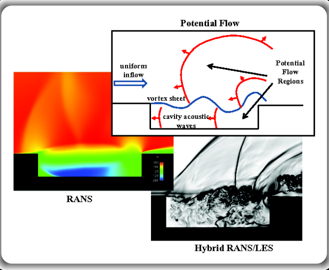
BUT...

Simulation models are computationally and financially expensive

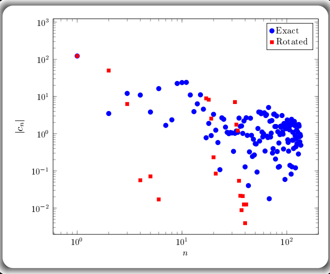


What are we doing to address these challenges?

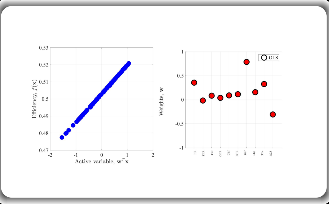
Multifidelity approaches



Compressed sensing



Dimension reduction



Today I will focus on compressed sensing

Polynomial Chaos Expansions (PCE)

Multidimensional approximation of $f(\xi)$ with finite variance

$$f(\boldsymbol{\xi}) \approx f_\Lambda(\boldsymbol{\xi}) = \sum_{\boldsymbol{\lambda}\in\Lambda} \alpha_{\boldsymbol{\lambda}} \phi_{\boldsymbol{\lambda}}(\boldsymbol{\xi}), \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$$

Orthonormal basis

$$(\phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi})) = \int_{I_{\boldsymbol{\xi}}} \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi}) w(\boldsymbol{\xi}) = \delta_{ij}$$

Assume ordering $n = 1, \ldots, N$ assigned to elements of Λ

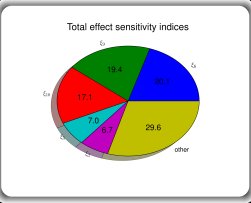
Askey scheme

Normal	Hermite $He_n(x)$	$e^{\frac{-x^2}{2}}$	$[-\infty, \infty]$
Uniform	Legendre $P_n(x)$	$\frac{1}{2}$	$[-1, 1]$

Why are PCE useful?

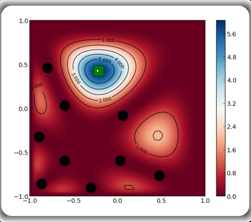
calculate moments and Sobol indices

$$\mu = \alpha_0, \quad \sigma^2 = \sum_{\boldsymbol{\lambda}\in\Lambda} \alpha_{\boldsymbol{\lambda}}^2$$



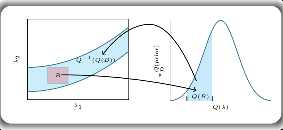
Tailor sampling to density

Weighted Leja sequences.



Surrogate for sampling methods

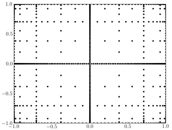
Computing PDFs, CDFs, probability of rare events, posterior-densities, etc.



How can we calculate PCE coefficients

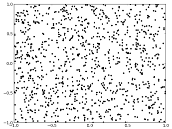
Pseudo spectral projection

$$\alpha_{\boldsymbol{\lambda}} = \int_{I_{\boldsymbol{\xi}}} f(\boldsymbol{\xi})\,\phi_{\boldsymbol{\lambda}}(\boldsymbol{\xi})\,d\rho(\boldsymbol{\xi})$$



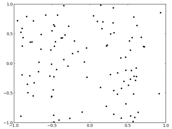
Least squares

$$\arg\min_{\boldsymbol{\alpha}} \|f(\boldsymbol{\xi}_m) - f_\Lambda(\boldsymbol{\xi}_m)\|_2$$



Compressed sensing

$$\arg\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_1 \text{ s.t. } \|f(\boldsymbol{\xi}_m) - f_\Lambda(\boldsymbol{\xi}_m)\|_2 \leq \varepsilon$$



What does compressed sensing do?

Compressed sensing attempts to find a sparse solution that is a “good” approximation of the observational data

A sparse solution

$$s = \#\{\boldsymbol{\lambda} : |\alpha_{\boldsymbol{\lambda}}| > 0\}$$

Typical “Good” approximation

$$\|f(\xi_m) - f_\Lambda(\xi_m)\|_2 \leq \varepsilon$$

Why do we care about sparsity?

Why do we care about sparsity?

Occam's razor — "when faced with many possible ways to represent a signal, the simplest choice is the best one."

Why do we care about sparsity?

“KISS”

Why do we care about sparsity?

The number of samples required grows LINERALY with dimension

Compressed Sensing

Generate M model runs

$$\Xi_M = \{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_M\}, \quad \mathbf{f} = (f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_M))^T$$

We want 'good' solution to

$$\begin{bmatrix} f(\boldsymbol{\xi}_1) \\ f(\boldsymbol{\xi}_2) \\ \vdots \\ f(\boldsymbol{\xi}_M) \end{bmatrix} = \begin{bmatrix} \phi_1(\boldsymbol{\xi}_1) & \phi_2(\boldsymbol{\xi}_1) & \dots & \phi_N(\boldsymbol{\xi}_1) \\ \phi_1(\boldsymbol{\xi}_2) & \phi_2(\boldsymbol{\xi}_2) & \dots & \phi_N(\boldsymbol{\xi}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\boldsymbol{\xi}_M) & \phi_2(\boldsymbol{\xi}_M) & \dots & \phi_N(\boldsymbol{\xi}_M) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \vdots \\ \vdots \\ \boldsymbol{\alpha}_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}$$

Sparse solutions

A sparse solution

$$s = \#\{\boldsymbol{\lambda} : |\alpha_{\boldsymbol{\lambda}}| > 0\}$$

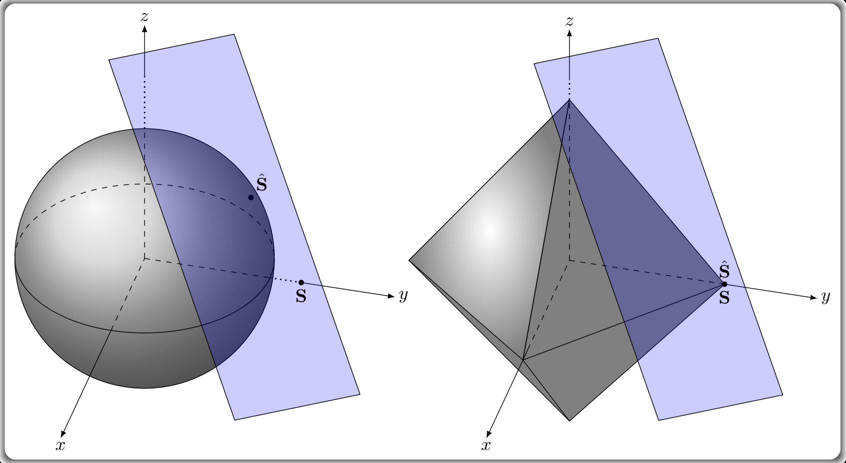
ℓ_0-minimization (NP HARD)

$$\arg\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_0 \text{ s.t. } \|f(\Xi_M) - f_\Lambda(\Xi_M)\|_2 \leq \varepsilon$$

ℓ_1-minimization

$$\arg\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_1 \text{ s.t. } \|f(\Xi_M) - f_\Lambda(\Xi_M)\|_2 \leq \varepsilon$$

Why does ℓ_1-minimization produce a sparse solution



How well does compressed sensing work?

ℓ_2-minimization

$$M \geq N \log(N) = dn^d \log(n)$$

ℓ_1-minimization

$$M \geq s \log^3(s) \log(N) L = ds \log^3(s) \log(n) L$$

Are PCE representations of models usually sparse?

Are PCE representations of models usually sparse?

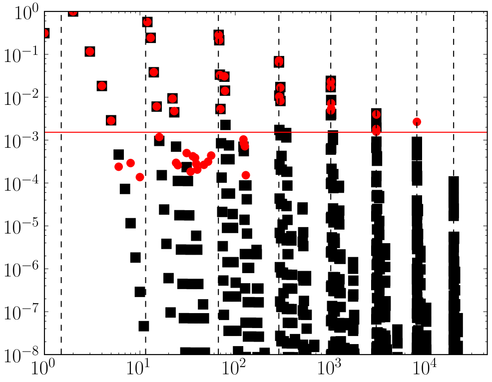
NO

But...

Are PCE representations of models usually sparse?

PCE are compressible

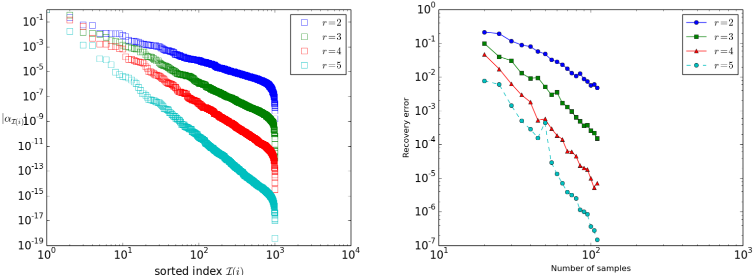
$$s = \#\{\boldsymbol{\lambda} : |\alpha_{\boldsymbol{\lambda}}| > \tau\}$$



How well does CS apply to compressible signals?

Compressible signals

$$\left|\alpha_{\mathcal{I}(i)}\right| \leq C i^{-r}$$



Recent contributions

- Reweighting
- Adaptive basis selection
- Change of measure

$$\boldsymbol{\alpha}^{\star} = \arg\min_{\boldsymbol{\alpha}} \; \|\boldsymbol{\alpha}\|_1 \quad \text{such that} \quad \|\mathbf{W}\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{W}\mathbf{f}\|_2 \leq \varepsilon$$

Requirements for finding a sparse solution

Small mutual coherence μ

$$\mu(\boldsymbol{\Phi}) = \max_{1<j<k\leq N} \frac{\left|\tilde{\phi}_j^T \tilde{\phi}_k\right|}{\left\|\tilde{\phi}_j\right\|_2 \left\|\tilde{\phi}_k\right\|_2}$$

Small RIP constant δ_s

$$(1-\delta_s)\left\|\boldsymbol{\alpha}_s\right\|_2^2 \leq \left\|\boldsymbol{\Phi}\boldsymbol{\alpha}_s\right\|_2^2 \leq (1+\delta_s)\left\|\boldsymbol{\alpha}_s\right\|_2^2$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \cdots & \phi_{\lambda_j}(\boldsymbol{\xi}_1) & \cdots & \phi_{\lambda_k}(\boldsymbol{\xi}_1) & \cdots \\ \cdots & \phi_{\lambda_j}(\boldsymbol{\xi}_2) & \cdots & \phi_{\lambda_k}(\boldsymbol{\xi}_2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \underbrace{\phi_{\lambda_j}(\boldsymbol{\xi}_M)}_{\tilde{\phi}_j} & \cdots & \underbrace{\phi_{\lambda_k}(\boldsymbol{\xi}_M)}_{\tilde{\phi}_k} & \cdots \end{pmatrix}$$

Theorem: RIP bound for Orthonormal Systems [Rahut and Ward 2010]

Consider the orthonormal system $\{\phi_j, j \in [N]\}$ with

$$\sup_{\xi \in D, j \in [N]} \|\phi_j\|_\infty \leq K$$

and the matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with entries formed by i.i.d. samples drawn from w. If

$$M \geq C\delta^{-2}K^2 s \log^3(s)\log(N),$$

then with probability at least $1 - N^{-\gamma \log^3(s)}$ the restricted isometry constant δ_s of $\frac{1}{\sqrt{M}}\mathbf{\Phi}$ satisfies $\delta_s \leq \delta$ for universal constants $C, \gamma > 0$

Change of measure

Enforce small coherence parameter

Scale basis functions by $\sqrt{W}$ so that

$$L = K^2 = O(1)$$

This can be achieved using the Chirstoffel function

Maintain orthogonality

We can no longer sample from $w(\xi)$ but must instead sample from the biased density

$$w(\xi)/W(\xi)$$

The corresponding measure of this density is the equilibrium measure μ where $w/W \approx d\mu$

The Christoffel function

$$W_\Lambda(\boldsymbol{\xi}) = \frac{N}{\sum_{n=1}^{N} \phi_n^2(\boldsymbol{\xi})}$$

Properties

- Generates bounded orthonormal system
- $w/W \to d\mu$, as $n \to \infty$

$$R_{kl} = \int_{I_\xi} \phi_k(\boldsymbol{\xi})\phi_l(\boldsymbol{\xi})W_\Lambda(\boldsymbol{\xi})d\mu, \quad 0 \le j, k \le N$$

$\mathbf{R} \to \mathbf{I}$, as $N \to \infty$

Theorem [Nevai et. al. 1994]

$$\max_{\xi\in[-1,1]} \frac{N\phi_N^2(\xi)}{\sum_{k=0}^N \phi_k^2(\xi)} \leq \frac{4N(2+\sqrt{\alpha^2+\beta^2})}{2N+\alpha+\beta+2} = K$$

[Levin and Lubinsky 1994]

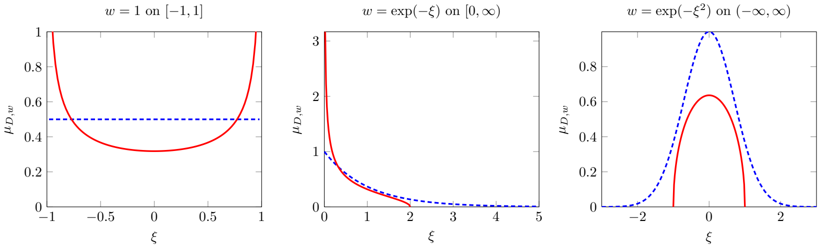
Similar more complicated bounds are known for unbounded variables with weight functions of the form

$$w(x) = \exp(-|\xi|^\alpha), \quad \alpha \geq 1$$

The equilibrium measure

Given $I_{\boldsymbol{\xi}}$ and w, we will be concerned with μ

- μ is a unique probability measure
- μ has compact support (even if $I_{\boldsymbol{\xi}}$ does not)
- With $d = 1$, μ coincides with the weighted potential-theoretic equilibrium measure (e.g., "Chebyshev-like" on 1D intervals)



Equilibrium sampling: all bounded variables

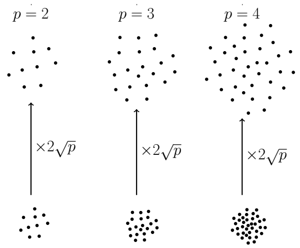
Let $z_i \sim U[0,1]$

$$\boldsymbol{\xi} = \cos(\mathbf{z}\pi)$$

Equilibrium sampling: Normal

Let $z_i \sim N(0,1)$ and $u \in [0,1]$ with PDF $w(u) = (1-u^2)^{d/2} u^{d-1}$

$$\mathbf{y} = \frac{\mathbf{z}}{\|\mathbf{z}\|_2} u, \quad \boldsymbol{\xi} = \mathbf{y} 2\sqrt{p}$$



Motivation for the equilibrium measure

ℓ_1-minimization

Allows one to bound weighted polynomials.

Regression

Ensures that the stability of the condition number can be achieved using only log-linear, i.e. $M = N \log N$.

Interpolation

It is necessary to sample from the equilibrium measure to obtain a 'good' Lebesque constant.

Theorem: [Jakeman et al.]

Suppose that M sampling points $(\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(M)})$ are drawn iid according to the equilibrium measure density v_n and the diagonal matrix $\mathbf{W}$ with entries given by $W_{ii} = W_\Lambda(\boldsymbol{\xi}^{(i)})$ Assume that the number of samples satisfies

$$M \geq L(n) \left\| \mathbf{R}^{-1/2} \right\|_1^2 s \log^3(s) \log(N),$$

where $\mathbf{R}$ measures the deviation from orthogonality. Then the coefficient vector $\mathbf{R}^{1/2}\boldsymbol{\alpha}$ is recoverable by solving the inequality-constrained ℓ^1-minimization problem

$$\mathbf{R}^{1/2}\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \ \|\mathbf{R}^{1/2}\boldsymbol{\alpha}\|_1 \quad \text{such that} \quad \|\sqrt{\mathbf{W}}\boldsymbol{\Phi}\boldsymbol{\alpha} - \sqrt{\mathbf{W}}\mathbf{f}\|_2 \leq \varepsilon$$

and the error in $\boldsymbol{\alpha}^\star$ satisfies

$$\begin{aligned} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_2 &\leq \frac{C_1 \sigma_{s,1}\left(\mathbf{R}^{1/2}\boldsymbol{\alpha}\right)}{\sqrt{s\lambda_{\min}(\mathbf{R})}} + C_2 \frac{\varepsilon}{\sqrt{\lambda_{\min}(\mathbf{R})}} \\ \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_1 &\leq D_1 \sigma_{s,1}\left(\mathbf{R}^{1/2}\boldsymbol{\alpha}\right) \left\|\mathbf{R}^{-1/2}\right\|_1 + D_2\sqrt{s} \left\|\mathbf{R}^{-1/2}\right\|_1 \varepsilon \end{aligned}$$

Theorem: [Jakeman et al.]

$L(n)$ has the following behavior:

1. There is a constant $C = C(\alpha, \beta)$ such that uniformly in $n \geq 1$,

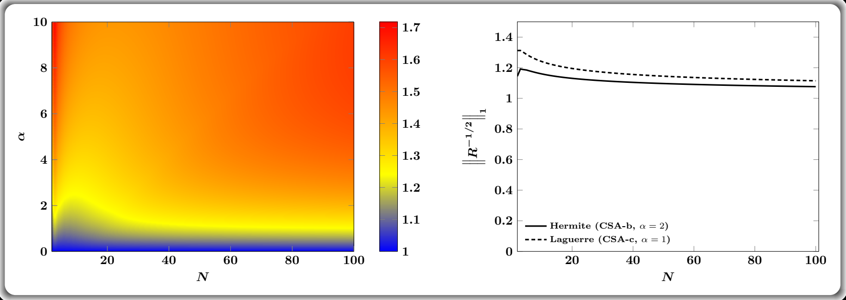
$$L(n) \leq C.$$

2. There is a constant $C = C(\alpha)$ such that uniformly in $n \geq 1$,

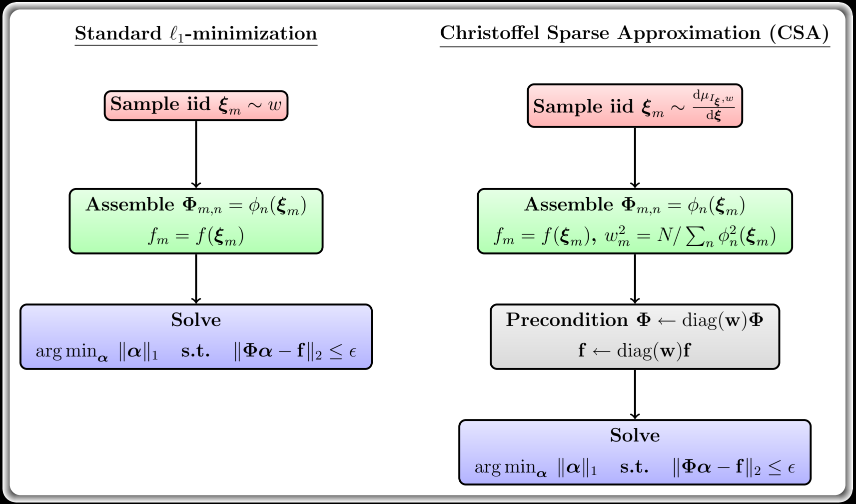
$$L(n) \leq Cn^{\max\{1/\alpha, 2/3\}} = \begin{cases} Cn^{2/3}, & \alpha \geq \frac{3}{2} \\ Cn^{1/\alpha}, & 1 < \alpha < \frac{3}{2} \end{cases}$$

3. There is a constant $C = C(\alpha)$ such that uniformly in $n \geq 1$,

$$L(n) \leq Cn^{\max\{1/2\alpha, 2/3\}} = \begin{cases} Cn^{2/3}, & \alpha \geq \frac{3}{4} \\ Cn^{1/2\alpha}, & \frac{1}{2} < \alpha < \frac{3}{4} \end{cases}$$



Left: $\left\|\mathbf{R}^{-1/2}\right\|_1$ for Jacobi polynomials $\alpha = \beta$.
Right: $\left\|\mathbf{R}^{-1/2}\right\|_1$ for Hermite polynomials with $w = \exp(-x^2)$ on $\mathbb{R}$, and Laguerre polynomials with $w = \exp(-x)$ on $[0, \infty)$.


Standard ℓ_1-minimization
Sample iid $\boldsymbol{\xi}_m \sim w$
Assemble $\boldsymbol{\Phi}_{m,n} = \phi_n(\boldsymbol{\xi}_m)$
$f_m = f(\boldsymbol{\xi}_m)$
Solve
$\arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1$ s.t. $\|\boldsymbol{\Phi\alpha} - \mathbf{f}\|_2 \leq \epsilon$
Christoffel Sparse Approximation (CSA)
Sample iid $\boldsymbol{\xi}_m \sim \frac{\mathrm{d}\mu_{I_{\boldsymbol{\xi}},w}}{\mathrm{d}\boldsymbol{\xi}}$
Assemble $\boldsymbol{\Phi}_{m,n} = \phi_n(\boldsymbol{\xi}_m)$
$f_m = f(\boldsymbol{\xi}_m)$, $w_m^2 = N/\sum_n \phi_n^2(\boldsymbol{\xi}_m)$
Precondition $\boldsymbol{\Phi} \leftarrow \mathrm{diag}(\mathbf{w})\boldsymbol{\Phi}$
$\mathbf{f} \leftarrow \mathrm{diag}(\mathbf{w})\mathbf{f}$
Solve
$\arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1$ s.t. $\|\boldsymbol{\Phi\alpha} - \mathbf{f}\|_2 \leq \epsilon$

Manufactured solutions

- Generate s-sparse vectors $\boldsymbol{\alpha}$
  - Index of each non-zero entries chosen $i = 1, \ldots, s \ \sim U(1, N)$ without replacement
  - Value of each non-zero entry $\alpha_i \ \sim N(0, 1)$
- Use Basis Pursuit to recover coefficients $\boldsymbol{\alpha}^\star$ from noiseless data $f(\boldsymbol{\xi}_m) = \sum_{n=1}^{N} \alpha_n \phi_n(\boldsymbol{\xi})$
  - Generate samples from w and $\rho_{I_{\boldsymbol{\xi}}, w}$
- Recovery successful if $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_2 / \|\boldsymbol{\alpha}\|_2 \leq 0.01$
- Measure probability of recovery using 100 trials

Alternative pre-conditioning schemes

Beta

$$\text{Let } z_i \sim U(0,1)$$

$$\boldsymbol{\xi} = \cos(\pi \mathbf{z}), \quad w_{m,m} = \prod_{i=1}^{d} (1 - \xi_i^2)^{1/4} \sqrt{w(\xi)}$$

Gaussian

$$\text{Let } z_i \sim N(0,1) \text{ and } u \sim U[0,1]$$

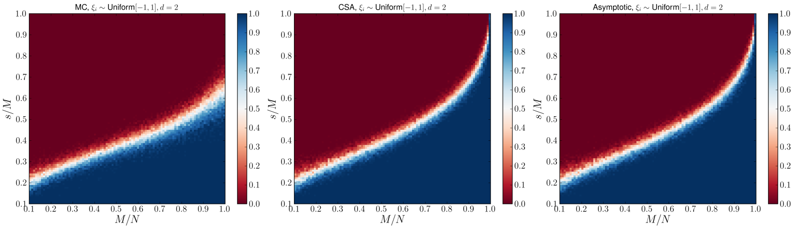
$$\mathbf{y} = \frac{\mathbf{z}}{\|\mathbf{z}\|_2} u^{1/d}, \quad \boldsymbol{\xi} = \mathbf{y}\sqrt{2}\sqrt{2p+1}$$

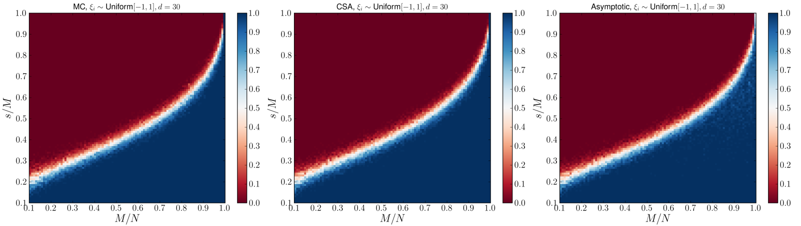
$$w_{m,m} = \exp(-\|\boldsymbol{\xi}\|_2^2/4)$$

- Asymptotic sampling: **y** are uniformly sampled in the unit ball.
- Equilibrium sampling: **y** are concentrated towards the center of the unit ball.

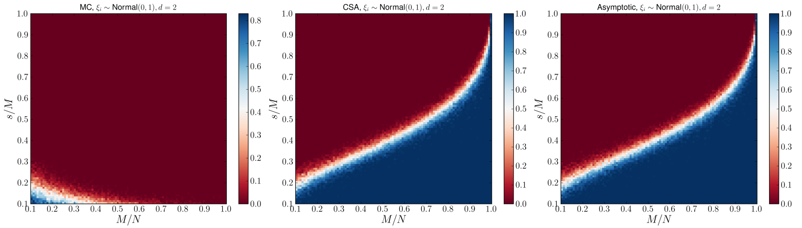
Special mention: coherence optimal sampling based upon MCMC. (Hampton and Doostan 2015)

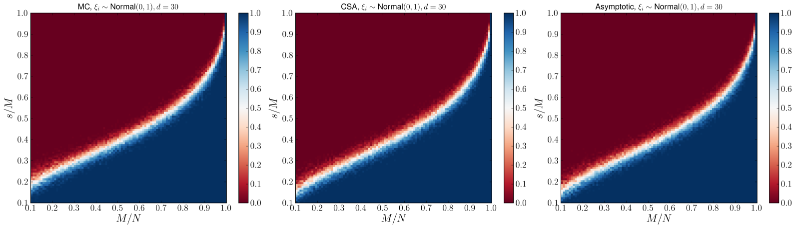
Uniform Variables (d=2,30)



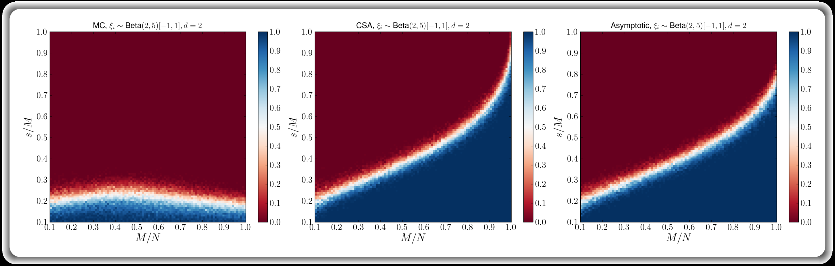


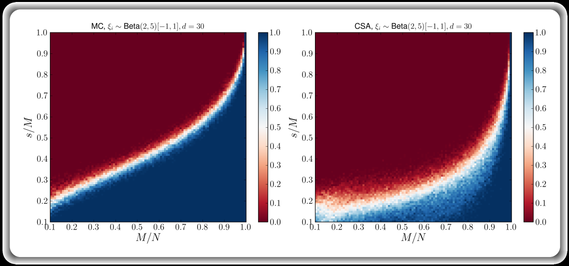
Normal Variables (d=2,30)





Beta(2,5) Variables (d=2,30)





Approximating an Eliptic PDE

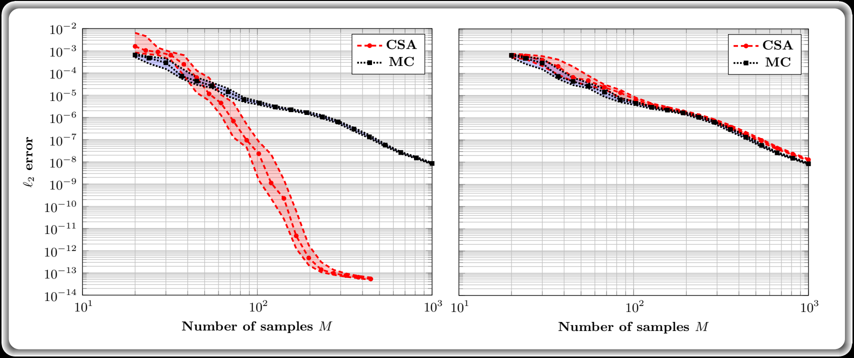
We want to approximate $q(\boldsymbol{\xi}) = u(1/2, \boldsymbol{\xi})$ where

$$-\frac{d}{dx}\left[a(x,\boldsymbol{\xi})\frac{du}{dx}(x,\boldsymbol{\xi})\right] = 1 \quad (x,\boldsymbol{\xi}) \in (0,1)\times I_{\boldsymbol{\xi}}$$
$$u(0,\boldsymbol{\xi}) = u(1,\boldsymbol{\xi}) = 0$$

with diffusivity $\log(a(x,\boldsymbol{\xi})) = \bar{a} + \sigma_a \sum_{k=1}^{d} \sqrt{\lambda_k}\varphi_k(x)\xi_k$, where $\{\lambda_k\}_{k=1}^{d}$ and $\{\varphi_k(x)\}_{k=1}^{d}$ are determined by $C_a(x_1, x_2) = \exp\left[-\frac{(x_1-x_2)^2}{l_c^2}\right]$

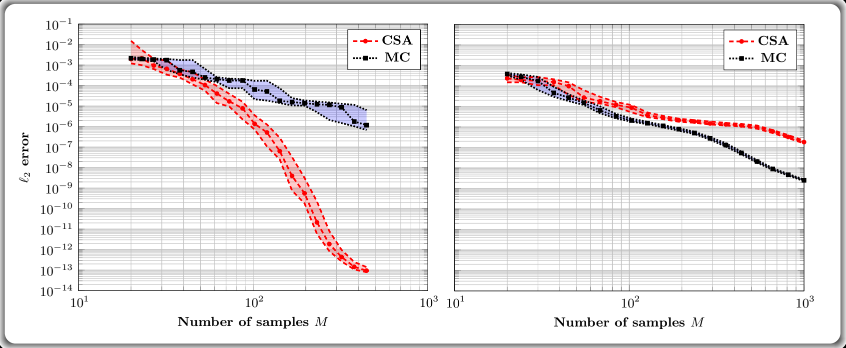
- Compute PCE using Basis Pursuit (Least Angle Regression)
- Measure accuracy in PCE approximation q_Λ by computing $M_{\text{test}}^{-1/2}\left\|q - q_\Lambda\right\|_{\ell_2(w)}$ using $M_{\text{test}} = 10000$ samples from $w(\boldsymbol{\xi})$.
- Measure mean error using 20 trials

Effect of dimension on performance of CSA



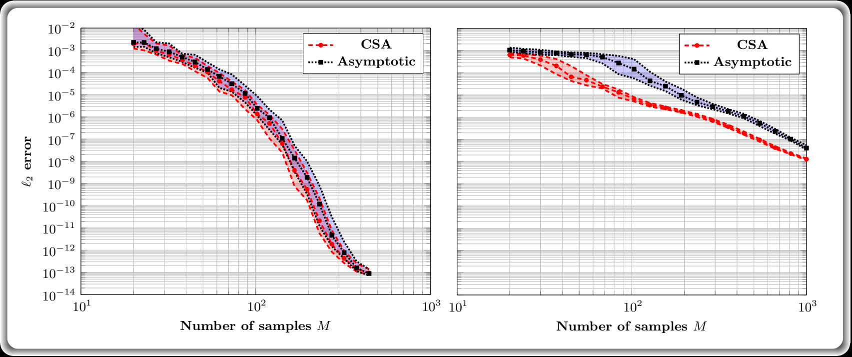
(Left) 30th degree Legendre polynomial in 2 dimensions. (Right) 4th degree Legendre polynomial in 20 dimensions

Effect of dimension on pefrformance of CSA



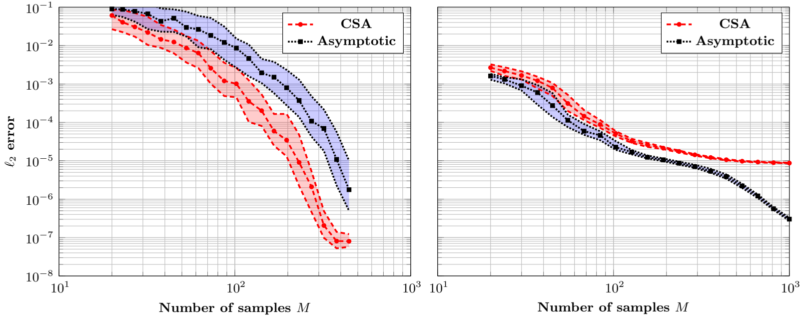
(Left) 30th degree Jacobi polynomial in 2 dimensions. (Right) 4th degree Jacobi polynomial in 20 dimensions

CSA vs. asymptotic method



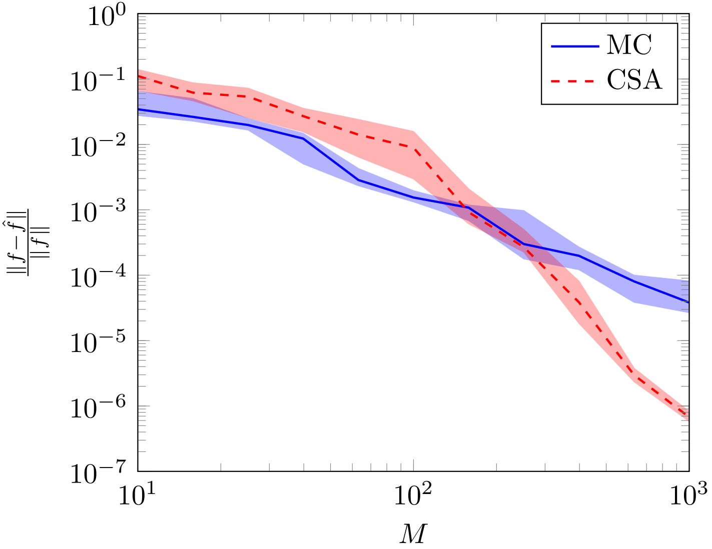
(left) $d = 2$ Jacobi approximation and (right) $d = 20$ Legendre approximation.

CSA vs. asymptotic method



(left) $d = 2$ and (right) $d = 20$ Hermite approximation.

Applying CSA to the nozzle problem



Applying CSA to the nozzle problem