SAND2015-3730C

Sampling and Preconditioning Strategies for ℓ_1-minimization

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.

Polynomial Chaos Expansions (PCE)

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

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

Orthonormal basis

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

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

Askey scheme

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

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

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

Compressed Sensing

Generate M model runs

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

We want 'good' solution to

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

How does one find a “good” sparse solution?

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

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

Requirements for finding a sparse solution

Small mutual coherence μ

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

Small RIP constant δ_s

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

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

$$M \geq C\delta^{-2} K^2 s \log^3(s) \log(N), \quad (1)$$

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 $C, \gamma > 0$

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. $M = N \log N$.

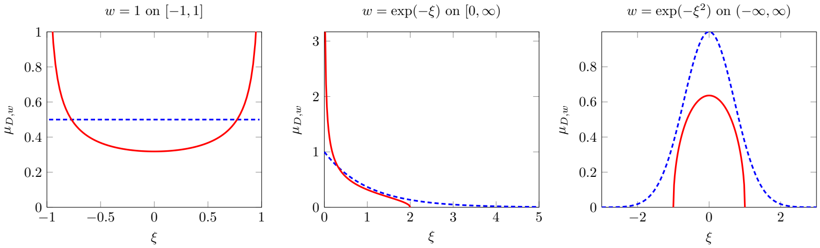
Interpolation

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

The equilibrium measure

Given D and w, we will be concerned with $\mu_{I_{\xi},w}$

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



Equilibrium sampling: Normal

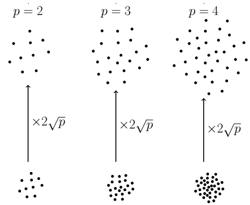
Let $z_i \sim N(0,1)$ and $u \sim U[0,1]$

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

Equilibrium sampling: Gamma

Let $z_i \sim \text{Gamma}(1/2, 1)$ and $u \sim U[0,1]$

$$\mathbf{y} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} u, \quad \xi = \mathbf{y} 4p$$



The Christoffel function

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

When applied to Gauss quadrature points the Christoffel function returns the Gauss quadrature weights

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

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

Theorem: [Jakeman et al.]

Let $M, N, s \in \mathbb{N}$ be given such that

$$M \geq Cs \log^3(s) \log(N).$$

Suppose that M sampling points $\boldsymbol{\xi}_m \sim$ i.i.d $\rho_{I_{\boldsymbol{\xi}},w}$ and consider $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ and the $\mathbf{W}$ with non-zero entries

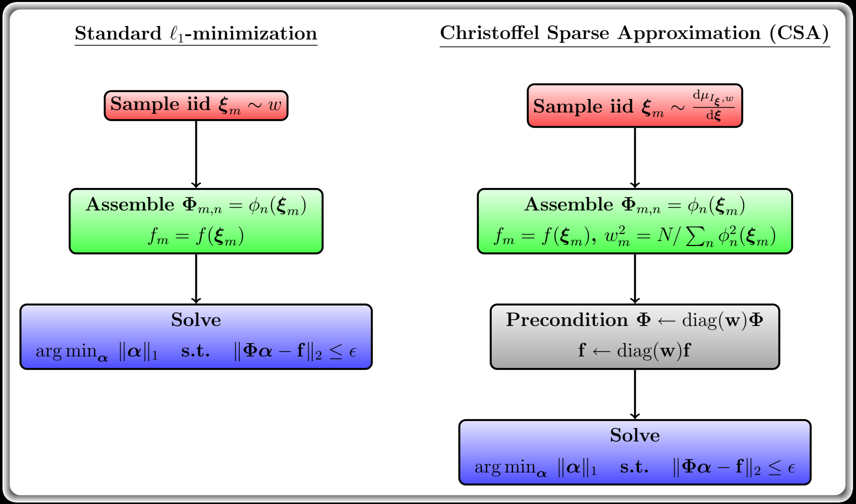
$$W_{ii} = N^{-1} \sum_{j=1}^{N} \phi_j^2(\boldsymbol{\xi}_i)$$

Then with probability exceeding $1 - N^{-\gamma \log^3(s)}$ the following holds for all polynomials $p_N(x)$. Suppose that noisy sample values $\mathbf{f} = \boldsymbol{\Phi}\boldsymbol{\alpha} + \boldsymbol{\eta}$ are observed, and $\|\mathbf{W}\eta\|_\infty \leq \varepsilon$. Then $\boldsymbol{\alpha}$ is recoverable to within a factor of its best s-term approximation ℓ_p-error $\sigma_s(z)_p$ and to a factor of the noise level ϵ by solving the inequality-constrained ℓ_1-minimization problem

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

Precisely,

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_2 \leq \frac{C_1\sigma_s(c)_1}{\sqrt{s}} + C_2\varepsilon, \quad \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_1 \leq D_1\sigma_s(c)_1 + D_2\sqrt{s}\varepsilon$$



Standard ℓ_1-minimization
Christoffel Sparse Approximation (CSA)
Sample iid $\boldsymbol{\xi}_m \sim w$
Assemble $\boldsymbol{\Phi}_{m,n} = \phi_n(\boldsymbol{\xi}_m)$
$f_m = f(\boldsymbol{\xi}_m)$
Solve
$\arg\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{f}\|_2 \le \epsilon$
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
$\arg\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{f}\|_2 \le \epsilon$

Manufactured solutions

- Generate s-sparse vectors $\boldsymbol{\alpha}$
  - Index of each non-zero entries chosen $i \ \sim U(1, N)$ 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

Uniform

Let $z_i \sim U(0,1)$

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

Gaussian

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

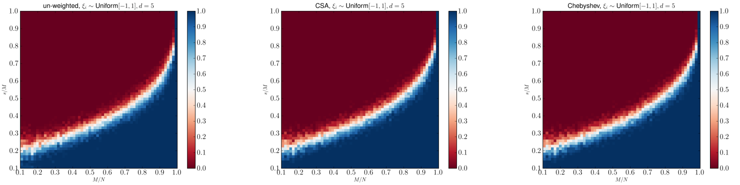
$$\mathbf{y} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} u^{1/d}, \quad \boldsymbol{\xi} = \mathbf{y}\sqrt{2}\sqrt{2p+1}$$

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

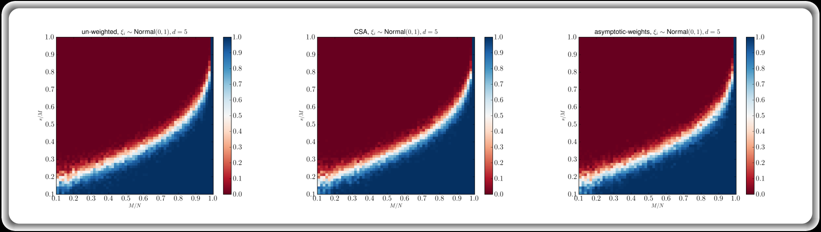
- Asymptotic sampling: $\mathbf{y}$ are uniformly sampled in the unit ball.
- Equilibrium sampling: $\mathbf{y}$ are concentrated towards the center of the unit ball.

Special mention: coherence optimal sampling based upon MCMC for uniform and Gaussian variables.

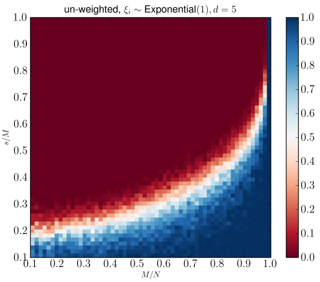
Uniform Variables



Gaussian Variables



Exponential Variables



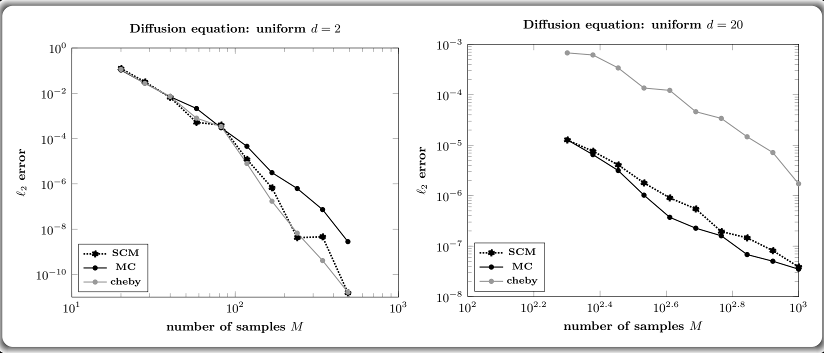
Approximating an Eliptic PDE

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

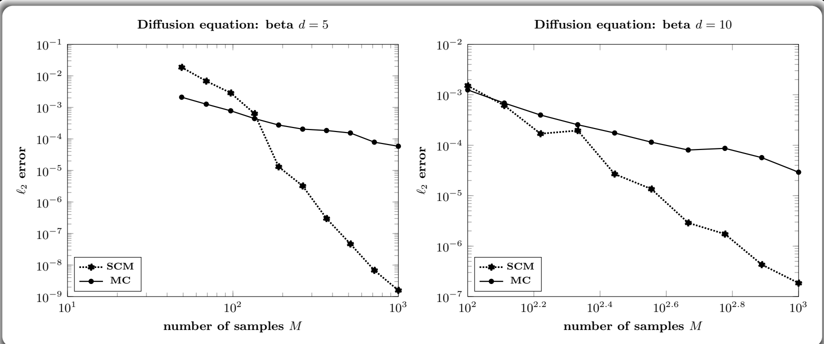
$$-\frac{d}{dx}\left[a(x,\boldsymbol{\xi})\frac{du}{dx}(x,\boldsymbol{\xi})\right] = 1 \quad (x,\boldsymbol{\xi}) \in (0,1) \times I_{\boldsymbol{\xi}}$$
$$u(0,\boldsymbol{\xi}) = u(1,\boldsymbol{\xi}) = 0$$

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 $C_a(x_1, x_2) = \exp\left[-\frac{(x_1-x_2)^2}{l_c^2}\right]$

Measure accuracy in PCE approximation $\hat{q}$ by computing $M_{\text{test}}^{-1/2}\|q - \hat{q}\|_{\ell_2(w)}$ using $M_{\text{test}} = 10000$ samples from $w(\boldsymbol{\xi})$.



Diffusion equation: uniform $d = 2$
ℓ_2 error
number of samples M
SCM
MC
cheby
Diffusion equation: uniform $d = 20$
ℓ_2 error
number of samples M
SCM
MC
cheby



Diffusion equation: beta $d = 5$
ℓ_2 error
number of samples M
SCM
MC
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{1}
10^{2}
10^{3}
Diffusion equation: beta $d = 10$
ℓ_2 error
number of samples M
SCM
MC
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{2}
$10^{2.2}$
$10^{2.4}$
$10^{2.6}$
$10^{2.8}$
10^{3}