# Optimal sampling in least-squares approximations

Albert Cohen

Laboratoire Jacques-Louis Lions
Sorbonne Université
Paris

Collaborators : Benjamin Arras, Markus Bachmayr,
Ronald DeVore, Giovanni Migliorati, Christoph Schwab

Nantes, 23-03-2018

## Problems that motivated this work

### 1. Reconstruction of acoustic fields :

An acoustic pressure field $p(x,t)$ generated by a source is measured by $n$ microphones at positions $x_1,\dots,x_n \in X \subset \mathbb{R}^2$ or $\mathbb{R}^3$, for $t \in [0,T]$.

Fourier analysis in time $p(x_i,t) \mapsto \hat{p}(x_i,\omega)$ and focus at a frequency $\omega$ of interest.

One wants to reconstruct the function $u(x) := \hat{p}(x,\omega)$ on $X$, from the observed data $u(x_i)$, $i = 1,\dots,n$.

## Problems that motivated this work

### 1. Reconstruction of acoustic fields :

An acoustic pressure field $p(x,t)$ generated by a source is measured by $n$ microphones at positions $x_1, \ldots, x_n \in X \subset \mathbb{R}^2$ or $\mathbb{R}^3$, for $t \in [0, T]$.

Fourier analysis in time $p(x_i, t) \mapsto \hat{p}(x_i, \omega)$ and focus at a frequency $\omega$ of interest.

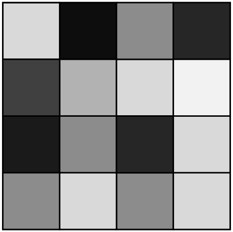
One wants to reconstruct the function $u(x) := \hat{p}(x, \omega)$ on $X$, from the observed data $u(x_i)$, $i = 1, \ldots, n$.

## 2. Fast solutions to high dimensional parametric PDE's :

Partial differential equation $\mathcal{P}(u,x) = 0$ depending on a parameter vector $x \in X \subset \mathbb{R}^d$ with $d >> 1$.

For each $x \in X$, the PDE is well posed in some Hilbert space $V$ : solution map $x \mapsto u(x) \in V$.

Example : $-\mathrm{div}(a\nabla u) = f$ on a domain $D$ (with boundary conditions), where diffusion $a$ is piecewise constant on subdomains $D_1, \dots, D_d$, with values $a_1, \dots, a_d$, which define the parameter vector $x = (a_1, \dots, a_d) \in X = [a_{\min}, a_{\max}]^d$.



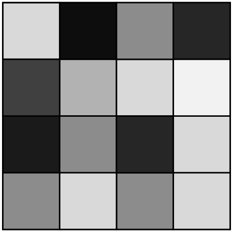
We want to reconstruct the solution map, from "snapshots" : particular instances of solutions $u(x_i)$, $i = 1, \dots, n$ computed by some numerical solver.

2. Fast solutions to high dimensional parametric PDE's :

Partial differential equation $\mathcal{P}(u,x)=0$ depending on a parameter vector $x \in X \subset \mathbb{R}^d$ with $d >> 1$.

For each $x \in X$, the PDE is well posed in some Hilbert space $V$ : solution map $x \mapsto u(x) \in V$.

Example : $-\mathrm{div}(a\nabla u) = f$ on a domain $D$ (with boundary conditions), where diffusion $a$ is piecewise constant on subdomains $D_1, \dots, D_d$, with values $a_1, \dots, a_d$, which define the parameter vector $x = (a_1, \dots, a_d) \in X = [a_{\min}, a_{\max}]^d$.



We want to reconstruct the solution map, from "snapshots" : particular instances of solutions $u(x_i)$, $i = 1, \dots, n$ computed by some numerical solver.

## Common features

Reconstruction of unknown function $u$ from scattered data.

Measurements $y_i = u(x_i)$ are costly : one cannot afford to have $n >> 1$.

Measurements could be noisy : $y_i = u(x_i) + \eta_i$.

The $x_i$ can be chosen by us (this talk) or imposed, deterministic or random.

Questions : how should we sample ? how should we reconstruct ?

Extra information on unknown function $u$ from the model (acoustic or PDE).

Approximability prior : analysis from these models shows that in both there exists sequences of $m$ dimensional linear spaces $(V_m)_{m>0}$ such that the unknown function $u$ is well approximated by such spaces

$$e_m(u) := \min_{v \in V_m} \|u - v\| \leq \varepsilon(m),$$

where $\varepsilon(m)$ is a known bound (such as $Cm^{-s}$) and where

$$\|v\| := \|v\|_{L^2(X,\rho)},$$

with $\rho$ some probability measure on $X$.

## Common features

Reconstruction of unknown function $u$ from scattered data.

Measurements $y_i = u(x_i)$ are costly : one cannot afford to have $n >> 1$.

Measurements could be noisy : $y_i = u(x_i) + \eta_i$.

The $x_i$ can be chosen by us (this talk) or imposed, deterministic or random.

Questions : how should we sample ? how should we reconstruct ?

Extra information on unknown function $u$ from the model (acoustic or PDE).

Approximability prior : analysis from these models shows that in both there exists sequences of $m$ dimensional linear spaces $(V_m)_{m>0}$ such that the unknown function $u$ is well approximated by such spaces

$$e_m(u) := \min_{v \in V_m} \|u - v\| \leq \varepsilon(m),$$

where $\varepsilon(m)$ is a known bound (such as $Cm^{-s}$) and where

$$\|v\| := \|v\|_{L^2(X,\rho)},$$

with $\rho$ some probability measure on $X$.

## Weighted least-squares approximation

The exact $L^2(X,\rho)$ projection $P_m u = \text{argmin}_{v \in V_m} \|u - v\|$ is out of reach.

For a certain value of $m \leq n$ solve :

$$u_W = \text{Argmin}_{v \in V_m} \frac{1}{n} \sum_{i=1}^{n} w(x_i)|y_i - v(x_i)|^2.$$

Widely used since its introduction by Gauss.

Standard (unweighted) least-squares : $w = 1$.

When $y_i = u(x_i)$ (noiseless case), then $u_W$ can be viewed as the orthogonal projection of $u$ onto $V_m$ in the sense of the Hilbertian (semi)-norm $\|\cdot\|_n$ defined by

$$\|v\|_n^2 := \frac{1}{n} \sum_{i=1}^{n} w(x_i)|v(x_i)|^2.$$

## Weighted least-squares approximation

The exact $L^2(X,\rho)$ projection $P_m u = \mathrm{argmin}_{v\in V_m}\|u-v\|$ is out of reach.

For a certain value of $m \le n$ solve :

$$u_W = \mathrm{Argmin}_{v\in V_m} \frac{1}{n}\sum_{i=1}^n w(x_i)|y_i - v(x_i)|^2.$$

Widely used since its introduction by Gauss.

Standard (unweighted) least-squares : $w = 1$.

When $y_i = u(x_i)$ (noiseless case), then $u_W$ can be viewed as the orthogonal projection of $u$ onto $V_m$ in the sense of the Hilbertian (semi)-norm $\|\cdot\|_n$ defined by

$$\|v\|_n^2 := \frac{1}{n}\sum_{i=1}^n w(x_i)|v(x_i)|^2.$$

## Implementation

The minimization problem is solved by using a given basis $L_1, \dots, L_m$ of $V_m$ and searching

$$u_W = \sum_{j=1}^{m} c_j L_j.$$

The vector $\mathbf{c} = (c_1, \dots, c_m)^t$ is solution to the normal equations

$$\mathbf{G}\mathbf{c} = \mathbf{a},$$

with $\mathbf{G} = (G_{k,j})_{k,j=1,\dots,m}$ and $\mathbf{a} = (a_1, \dots, a_n)^t$, where

$$G_{k,j} := \frac{1}{n} \sum_{i=1}^{n} w(x_i) L_k(x_i) L_j(x_i) \quad \text{and} \quad a_k := \frac{1}{n} \sum_{i=1}^{n} w(x_i) y_i L_k(x_i).$$

The solution always exists and is unique if $\mathbf{G}$ is invertible.

## Implementation

The minimization problem is solved by using a given basis $L_1, \dots, L_m$ of $V_m$ and searching

$$u_W = \sum_{j=1}^{m} c_j L_j.$$

The vector $\mathbf{c} = (c_1, \dots, c_m)^t$ is solution to the normal equations

$$\mathbf{G}\mathbf{c} = \mathbf{a},$$

with $\mathbf{G} = (G_{k,j})_{k,j=1,\dots,m}$ and $\mathbf{a} = (a_1, \dots, a_n)^t$, where

$$G_{k,j} := \frac{1}{n}\sum_{i=1}^{n} w(x_i)L_k(x_i)L_j(x_i) \quad \text{and} \quad a_k := \frac{1}{n}\sum_{i=1}^{n} w(x_i)y_iL_k(x_i).$$

The solution always exists and is unique if $\mathbf{G}$ is invertible.

## General questions

1. How accurate is the least square approximation ?
2. Stability with respect to data perturbations ?
3. How large should we take $n$ compared to $m$ ?

A typical trade-off :

If $m$ is small : high amount of regularization, stabilizes the method, but $V_m$ has poor approximation properties.

If $m$ is large : better approximation properties, but the method may become unstable and therefore unaccurate (also in the noiseless case).

How can we describe the optimal compromise ?

Can we have stable and accurate approximations with $n = \mathcal{O}(m)$ samples ?

How does this depend on the distribution of the samples $x_i$ ?

## General questions

1. How accurate is the least square approximation ?
2. Stability with respect to data perturbations ?
3. How large should we take $n$ compared to $m$ ?

A typical trade-off :

If $m$ is small : high amount of regularization, stabilizes the method, but $V_m$ has poor approximation properties.

If $m$ is large : better approximation properties, but the method may become unstable and therefore unaccurate (also in the noiseless case).

How can we describe the optimal compromise ?

Can we have stable and accurate approximations with $n = \mathcal{O}(m)$ samples ?

How does this depend on the distribution of the samples $x_i$ ?

## General questions

1. How accurate is the least square approximation ?
2. Stability with respect to data perturbations ?
3. How large should we take $n$ compared to $m$ ?

A typical trade-off :

If $m$ is small : high amount of regularization, stabilizes the method, but $V_m$ has poor approximation properties.

If $m$ is large : better approximation properties, but the method may become unstable and therefore unaccurate (also in the noiseless case).

How can we describe the optimal compromise ?

Can we have stable and accurate approximations with $n = \mathcal{O}(m)$ samples ?

How does this depend on the distribution of the samples $x_i$ ?

Recall that we measure approximation error in the $L^2(X,\rho)$ norm,

$$\|v\|^2 := \int_X |v(x)|^2 d\rho,$$

where $\rho$ is a probability measure.

Pick the $x_i$ independently at random according to another probability measure $\mu$ over $X$, requiring that

$$d\rho = w \, d\mu.$$

Therefore, as $n$ gets large

$$\|v\|_n^2 = \frac{1}{n}\sum_{i=1}^n w(x_i)|v(x_i)|^2 \sim \int_X w(x)|v(x)|^2 d\mu = \|v\|^2,$$

and one has $\mathbb{E}(\|v\|_n^2) = \|v\|^2$.

Trivial choice : $w = 1$ and $\rho = \mu$, unweighted least-squares.

Our analysis reveals that there is a substantial interest in not going for this choice (similar to importance sampling).

Earlier avocated in work by Narayan, Doostan-Hampton on polynomial regression (2015).

# A stochastic setting

Recall that we measure approximation error in the $L^2(X,\rho)$ norm,

$$\|v\|^2 := \int_X |v(x)|^2 d\rho,$$

where $\rho$ is a probability measure.

Pick the $x_i$ independently at random according to another probability measure $\mu$ over $X$, requiring that

$$d\rho = w\, d\mu.$$

Therefore, as $n$ gets large

$$\|v\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} w(x_i)|v(x_i)|^2 \sim \int_X w(x)|v(x)|^2 d\mu = \|v\|^2,$$

and one has $\mathbb{E}(\|v\|_n^2) = \|v\|^2$.

Trivial choice : $w = 1$ and $\rho = \mu$, unweighted least-squares.

Our analysis reveals that there is a substantial interest in not going for this choice (similar to importance sampling).

Earlier avocated in work by Narayan, Doostan-Hampton on polynomial regression (2015).

Recall that we measure approximation error in the $L^2(X,\rho)$ norm,

$$\|v\|^2 := \int_X |v(x)|^2 d\rho,$$

where $\rho$ is a probability measure.

Pick the $x_i$ independently at random according to another probability measure $\mu$ over $X$, requiring that

$$d\rho = w\, d\mu.$$

Therefore, as $n$ gets large

$$\|v\|_n^2 = \frac{1}{n}\sum_{i=1}^n w(x_i)|v(x_i)|^2 \sim \int_X w(x)|v(x)|^2 d\mu = \|v\|^2,$$

and one has $\mathbb{E}(\|v\|_n^2) = \|v\|^2$.

Trivial choice : $w = 1$ and $\rho = \mu$, unweighted least-squares.

Our analysis reveals that there is a substantial interest in not going for this choice (similar to importance sampling).

Earlier avocated in work by Narayan, Doostan-Hampton on polynomial regression (2015).

# Stability and accuracy

We want to compare the least-square approximation error $\|u - u_W\|$ with the best approximation error in the $L^2(X, \rho)$ norm

$$e_m(u) := \inf_{v \in V_m} \|u - v\|,$$

This comparison is tied to the stability of the weighted least-squares method.

If $L_1, \dots, L_m$ is an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm, the Grammian matrix

$$\mathbf{G} = (G_{k,j}) := \Big(\frac{1}{n}\sum_{i=1}^{n} w(x_i) L_k(x_i) L_j(x_i)\Big),.$$

involved in the normal equations satisfies $\mathbb{E}(\mathbf{G}) = I$.

Our analysis relies on a probabilisty control of $\|\mathbf{G} - I\|$, where $\|M\|$ is the spectral norm of a matrix, or equivalently of the condition number $\kappa(\mathbf{G})$.

Stable sampling : note that

$$\|\mathbf{G} - I\| \leq \delta \iff (1-\delta)\|v\|^2 \leq \|v\|_n^2 \leq (1+\delta)\|v\|^2, \quad v \in V_m$$

By convention, we set $u_W = 0$ in the event where $\|\mathbf{G} - I\| \geq \frac{1}{2}$ and retain it otherwise.

## Stability and accuracy

We want to compare the least-square approximation error $\|u - u_W\|$ with the best approximation error in the $L^2(X, \rho)$ norm

$$e_m(u) := \inf_{v \in V_m} \|u - v\|,$$

This comparison is tied to the stability of the weighted least-squares method.

If $L_1, \dots, L_m$ is an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm, the Grammian matrix

$$\mathbf{G} = (G_{k,j}) := \Big(\frac{1}{n}\sum_{i=1}^{n} w(x_i) L_k(x_i) L_j(x_i)\Big),.$$

involved in the normal equations satisfies $\mathbb{E}(\mathbf{G}) = I$.

Our analysis relies on a probabilisty control of $\|\mathbf{G} - I\|$, where $\|M\|$ is the spectral norm of a matrix, or equivalently of the condition number $\kappa(\mathbf{G})$.

Stable sampling : note that

$$\|\mathbf{G} - I\| \leq \delta \iff (1-\delta)\|v\|^2 \leq \|v\|_n^2 \leq (1+\delta)\|v\|^2, \quad v \in V_m$$

By convention, we set $u_W = 0$ in the event where $\|\mathbf{G} - I\| \geq \frac{1}{2}$ and retain it otherwise.

We want to compare the least-square approximation error $\|u - u_W\|$ with the best approximation error in the $L^2(X, \rho)$ norm

$$e_m(u) := \inf_{v \in V_m} \|u - v\|,$$

This comparison is tied to the stability of the weighted least-squares method.

If $L_1, \dots, L_m$ is an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm, the Grammian matrix

$$\mathbf{G} = (G_{k,j}) := \Big( \frac{1}{n} \sum_{i=1}^{n} w(x_i) L_k(x_i) L_j(x_i) \Big),.$$

involved in the normal equations satisfies $\mathbb{E}(\mathbf{G}) = I$.

Our analysis relies on a probabilisty control of $\|\mathbf{G} - I\|$, where $\|M\|$ is the spectral norm of a matrix, or equivalently of the condition number $\kappa(\mathbf{G})$.

Stable sampling : note that

$$\|\mathbf{G} - I\| \leq \delta \iff (1-\delta)\|v\|^2 \leq \|v\|_n^2 \leq (1+\delta)\|v\|^2, \quad v \in V_m$$

By convention, we set $u_W = 0$ in the event where $\|\mathbf{G} - I\| \geq \frac{1}{2}$ and retain it otherwise.

We want to compare the least-square approximation error $\|u - u_W\|$ with the best approximation error in the $L^2(X, \rho)$ norm

$$e_m(u) := \inf_{v \in V_m} \|u - v\|,$$

This comparison is tied to the stability of the weighted least-squares method.

If $L_1, \dots, L_m$ is an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm, the Grammian matrix

$$\mathbf{G} = (G_{k,j}) := \Big( \frac{1}{n} \sum_{i=1}^{n} w(x_i) L_k(x_i) L_j(x_i) \Big), .$$

involved in the normal equations satisfies $\mathbb{E}(\mathbf{G}) = I$.

Our analysis relies on a probabilisty control of $\|\mathbf{G} - I\|$, where $\|M\|$ is the spectral norm of a matrix, or equivalently of the condition number $\kappa(\mathbf{G})$.

Stable sampling : note that

$$\|\mathbf{G} - I\| \le \delta \iff (1-\delta)\|v\|^2 \le \|v\|_n^2 \le (1+\delta)\|v\|^2, \quad v \in V_m$$

By convention, we set $u_W = 0$ in the event where $\|\mathbf{G} - I\| \ge \frac{1}{2}$ and retain it otherwise.

## The key ingredient to our analysis

Let $L_1, \dots, L_m$ be an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm. We introduce

$$k_{m,w}(x) := w(x) \sum_{j=1}^{m} |L_j(x)|^2,$$

and

$$K_{m,w} := \|k_{m,w}\|_{L^\infty} = \sup_{x \in X} w(x) \sum_{j=1}^{m} |L_j(x)|^2.$$

Both are independent on the choice orthonormal basis : only depends on $(V_m, \rho, w)$.

Since $\int_X k_{m,w} d\mu = \sum_{j=1}^m \|L_j\|^2 = m$, one has

$$K_{m,w} \geq m.$$

In the case $w = 1$, we obtain the Christoffel function $k_m(x) := \sum_{j=1}^m |L_j(x)|^2$, which is the diagonal of the orthogonal projection kernel onto $V_m$, and such that

$$K_m := \|k_m\|_{L^\infty} = \max_{v \in V_m} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

## The key ingredient to our analysis

Let $L_1, \dots, L_m$ be an orthonormal basis of $V_m$ for the $L^2(X,\rho)$ norm. We introduce

$$k_{m,w}(x) := w(x) \sum_{j=1}^{m} |L_j(x)|^2,$$

and

$$K_{m,w} := \|k_{m,w}\|_{L^\infty} = \sup_{x \in X} w(x) \sum_{j=1}^{m} |L_j(x)|^2.$$

Both are independent on the choice orthonormal basis : only depends on $(V_m, \rho, w)$.

Since $\int_X k_{m,w} d\mu = \sum_{j=1}^m \|L_j\|^2 = m$, one has

$$K_{m,w} \geq m.$$

In the case $w = 1$, we obtain the Christoffel function $k_m(x) := \sum_{j=1}^m |L_j(x)|^2$, which is the diagonal of the orthogonal projection kernel onto $V_m$, and such that

$$K_m := \|k_m\|_{L^\infty} = \max_{v \in V_m} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

## The key ingredient to our analysis

Let $L_1, \dots, L_m$ be an orthonormal basis of $V_m$ for the $L^2(X, \rho)$ norm. We introduce

$$k_{m,w}(x) := w(x) \sum_{j=1}^{m} |L_j(x)|^2,$$

and

$$K_{m,w} := \|k_{m,w}\|_{L^\infty} = \sup_{x \in X} w(x) \sum_{j=1}^{m} |L_j(x)|^2.$$

Both are independent on the choice orthonormal basis : only depends on $(V_m, \rho, w)$.

Since $\int_X k_{m,w} d\mu = \sum_{j=1}^m \|L_j\|^2 = m$, one has

$$K_{m,w} \geq m.$$

In the case $w = 1$, we obtain the Christoffel function $k_m(x) := \sum_{j=1}^m |L_j(x)|^2$, which is the diagonal of the orthogonal projection kernel onto $V_m$, and such that

$$K_m := \|k_m\|_{L^\infty} = \max_{v \in V_m} \frac{\|v\|_{L^\infty}^2}{\|v\|^2}.$$

## Sample result in the noiseless case (Cohen-Migliorati 2017, Doostan-Hampton 2015)

Let $0 < \varepsilon < 1$ be arbitrary. Under the condition

$$K_{m,w} \leq c \frac{n}{\log(2m/\varepsilon)}, \quad c := \frac{1-\log 2}{2},$$

the weighted least-squares approximation is

(i) stable : one has the deviation bound

$$\Pr\left\{\|G - I\| \geq \frac{1}{2}\right\} \leq \varepsilon.$$

(ii) accurate : one has

$$\mathbb{E}(\|u - u_W\|^2) \leq (1 + \delta(n)) e_m(u)^2 + \varepsilon \|u\|^2, \quad \delta(n) := \frac{c}{\log(2m/\varepsilon)}.$$

Variant to these results : error bounds in probability, noisy case.

Typical choice : $\varepsilon = m^{-r}$ for $r > 0$ larger than approximation rate.

Gives stability condition $K_{m,w} \lesssim \frac{n}{\log m}$, which imposes at least that $n \gtrsim m \log m$.

It can be much more demanding if $K_{m,w} >> m$.

## Sample result in the noiseless case (Cohen-Migliorati 2017, Doostan-Hampton 2015)

Let $0 < \varepsilon < 1$ be arbitrary. Under the condition

$$K_{m,w} \leq c \frac{n}{\log(2m/\varepsilon)}, \quad c := \frac{1-\log 2}{2},$$

the weighted least-squares approximation is

(i) stable : one has the deviation bound

$$\Pr\left\{\|G - I\| \geq \frac{1}{2}\right\} \leq \varepsilon.$$

(ii) accurate : one has

$$\mathbb{E}(\|u - u_W\|^2) \leq (1 + \delta(n))e_m(u)^2 + \varepsilon\|u\|^2, \quad \delta(n) := \frac{c}{\log(2m/\varepsilon)}.$$

Variant to these results : error bounds in probability, noisy case.

Typical choice : $\varepsilon = m^{-r}$ for $r > 0$ larger than approximation rate.

Gives stability condition $K_{m,w} \lesssim \frac{n}{\log m}$, which imposes at least that $n \gtrsim m \log m$.

It can be much more demanding if $K_{m,w} >> m$.

## Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $m \times m$ rank one random matrix

$$\mathbf{X} = w(x)(L_k(x)L_j(x))_{j,k=1,\dots,m},$$

which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|X\| \le K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \ge \delta\right\} \le 2m\exp\left(-\frac{nc(\delta)}{K}\right),$$

where $c(\delta) := \delta + (1-\delta)\log(1-\delta) > 0$ (in particular $c(\frac{1}{2}) := c = \frac{1-\log 2}{2}$).

Here $K = \sup_{x\in X} w(x)\sum_{j=1}^{m} |L_j(x)|^2 = K_{m,w}$.

Therefore $K_{m,w} \le c\frac{n}{\log(2m/\varepsilon)} \implies \Pr\{\|G - I\| \ge \frac{1}{2}\} \le \varepsilon$.

## Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $m \times m$ rank one random matrix

$$\mathbf{X} = w(x)(L_k(x)L_j(x))_{j,k=1,\dots,m},$$

which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|X\| \leq K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \geq \delta\right\} \leq 2m \exp\left(-\frac{nc(\delta)}{K}\right),$$

where $c(\delta) := \delta + (1-\delta)\log(1-\delta) > 0$ (in particular $c(\frac{1}{2}) := c = \frac{1-\log 2}{2}$).

Here $K = \sup_{x \in X} w(x) \sum_{j=1}^{m} |L_j(x)|^2 = K_{m,w}$.

Therefore $K_{m,w} \leq c \frac{n}{\log(2m/\varepsilon)} \implies \Pr\{\|G - I\| \geq \frac{1}{2}\} \leq \varepsilon$.

## Where does the stability condition comes from

We may write

$$\mathbf{G} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i,$$

where $\mathbf{X}_i$ are i.i.d. copies of the $m \times m$ rank one random matrix

$$\mathbf{X} = w(x)(L_k(x)L_j(x))_{j,k=1,\dots,m},$$

which has expectation $\mathbb{E}(\mathbf{X}) = \mathbf{I}$.

Matrix Chernoff bound (Ahlswede-Winter 2000, Tropp 2011) : if $\|X\| \leq K$ a.s., then

$$\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mathbb{E}(\mathbf{X})\right\| \geq \delta\right\} \leq 2m\exp\left(-\frac{nc(\delta)}{K}\right),$$

where $c(\delta) := \delta + (1-\delta)\log(1-\delta) > 0$ (in particular $c(\frac{1}{2}) := c = \frac{1-\log 2}{2}$).

Here $K = \sup_{x\in X} w(x)\sum_{j=1}^{m}|L_j(x)|^2 = K_{m,w}$.

Therefore $K_{m,w} \leq c\frac{n}{\log(2m/\varepsilon)} \implies \Pr\{\|G-I\| \geq \frac{1}{2}\} \leq \varepsilon$.

## The unweighted case $w = 1$

The stability regime is described by the condition $K_m = \|k_m\|_{L^\infty} \lesssim \frac{n}{\log m}$.

We can estimate the Christoffel function $k_m(x) = \sum_{j=1}^m |L_j(x)|^2$ in cases of practical interest.

A simple example : $X = [-1, 1]$ and $V_m = \mathbb{P}_{m-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dx}{\pi\sqrt{1-x^2}}$ : the $L_j$ are the Chebychev polynomials and $K_m = 2m + 1$. Up to log factors, the stability regime is $n \gtrsim m$.

(ii) Uniform distribution $\rho = \frac{dx}{2}$ : the $L_j$ are normalized Legendre polynomials and $K_m = \sum_{j=1}^m (2j-1) = m^2$. Up to log factors, the stability regime is $n \gtrsim m^2$.

These regimes are confirmed numerically.

## The unweighted case $w = 1$

The stability regime is described by the condition $K_m = \|k_m\|_{L^\infty} \lesssim \frac{n}{\log m}$.

We can estimate the Christoffel function $k_m(x) = \sum_{j=1}^m |L_j(x)|^2$ in cases of practical interest.

A simple example : $X = [-1, 1]$ and $V_m = \mathbb{P}_{m-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dx}{\pi\sqrt{1-x^2}}$ : the $L_j$ are the Chebychev polynomials and $K_m = 2m + 1$. Up to log factors, the stability regime is $n \gtrsim m$.

(ii) Uniform distribution $\rho = \frac{dx}{2}$ : the $L_j$ are normalized Legendre polynomials and $K_m = \sum_{j=1}^m (2j-1) = m^2$. Up to log factors, the stability regime is $n \gtrsim m^2$.

These regimes are confirmed numerically.

## The unweighted case $w = 1$

The stability regime is described by the condition $K_m = \|k_m\|_{L^\infty} \lesssim \frac{n}{\log m}$.

We can estimate the Christoffel function $k_m(x) = \sum_{j=1}^m |L_j(x)|^2$ in cases of practical interest.

A simple example : $X = [-1, 1]$ and $V_m = \mathbb{P}_{m-1}$ the univariate polynomials.

(i) Distribution $\rho = \frac{dx}{\pi\sqrt{1-x^2}}$ : the $L_j$ are the Chebychev polynomials and $K_m = 2m + 1$. Up to log factors, the stability regime is $n \gtrsim m$.

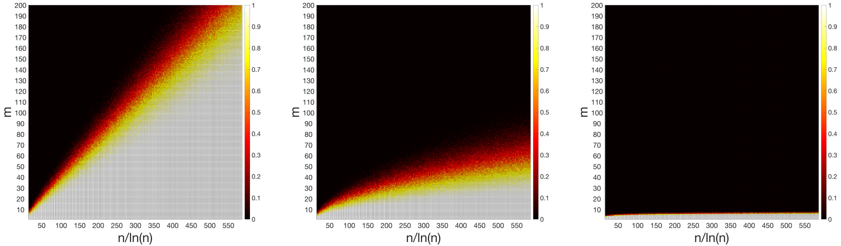
(ii) Uniform distribution $\rho = \frac{dx}{2}$ : the $L_j$ are normalized Legendre polynomials and $K_m = \sum_{j=1}^m (2j - 1) = m^2$. Up to log factors, the stability regime is $n \gtrsim m^2$.

These regimes are confirmed numerically.

## Illustration

Regime of stability : probability that $\kappa(\mathbf{G}) \leq 3$, white if 1, black if 0.

Left : for $\rho = \frac{dx}{\pi\sqrt{1-x^2}}$. Center : for $\rho = \frac{dx}{2}$.




Right : the gaussian case $X = \mathbb{R}$ and $\rho = g(x)dx$, where $g(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, for which the $L_j$ are the Hermite polynomials.

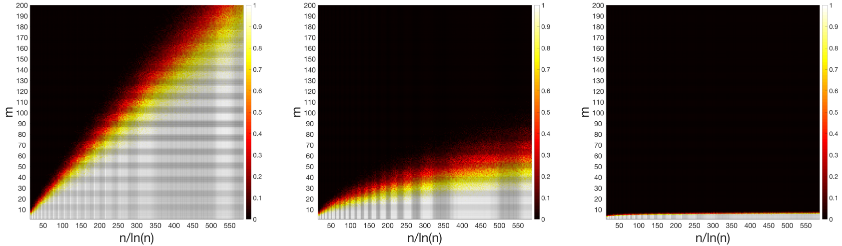
The unweighted theory cannot handle this case since $K_m = \infty$

A more ad-hoc analysis shows that stability holds if $n \gtrsim \exp(cm)$ and this regime is observed numerically.

## Illustration

Regime of stability : probability that $\kappa(\mathbf{G}) \leq 3$, white if 1, black if 0.

Left : for $\rho = \frac{dx}{\pi\sqrt{1-x^2}}$. Center : for $\rho = \frac{dx}{2}$.




Right : the gaussian case $X = \mathbb{R}$ and $\rho = g(x)dx$, where $g(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, for which the $L_j$ are the Hermite polynomials.

The unweighted theory cannot handle this case since $K_m = \infty$

A more ad-hoc analysis shows that stability holds if $n \gtrsim \exp(cm)$ and this regime is observed numerically.

## High dimensions : parametric PDE's

Prototype example : elliptic PDE's on some domain $D \subset \mathbb{R}^2$ or $\mathbb{R}^3$ with affine parametrization of the diffusion function by $x = (x_1, \dots, x_d) \in X = [-1,1]^d$

$$-\mathrm{div}(a\nabla u) = f, \quad a = \bar{a} + \sum_{j=1}^{d} x_j \psi_j,$$

with ellipticity assumption $0 < r < a < R$ for all $x \in X$, so $x \mapsto u(x) \in V = H_0^1(D)$.

With $\Lambda \subset \mathbb{N}^d$, approximation by multivariate polynomial space

$$V_\Lambda := \left\{ \sum_{\nu\in\Lambda} v_\nu x^\nu, \quad v_\nu \in V \right\} = V \otimes \mathbb{P}_\Lambda,$$

where $x^\nu = x_1^{\nu_1} \cdots x_d^{\nu_d}$.

We only consider downward closed index sets : $\nu \in \Lambda$ and $\mu \leq \nu \Rightarrow \mu \in \Lambda$.

Basis of $\mathbb{P}_\Lambda$ : tensorized orthogonal polynomials $L_\nu(x) = \prod_{j=1}^{d} L_{\nu_j}(x_j)$ for $\nu \in \Lambda$.

## High dimensions : parametric PDE's

Prototype example : elliptic PDE's on some domain $D \subset \mathbb{R}^2$ or $\mathbb{R}^3$ with affine parametrization of the diffusion function by $x = (x_1, \dots, x_d) \in X = [-1, 1]^d$

$$-\mathrm{div}(a\nabla u) = f, \quad a = \bar{a} + \sum_{j=1}^{d} x_j \psi_j,$$

with ellipticity assumption $0 < r < a < R$ for all $x \in X$, so $x \mapsto u(x) \in V = H_0^1(D)$.

With $\Lambda \subset \mathbb{N}^d$, approximation by multivariate polynomial space

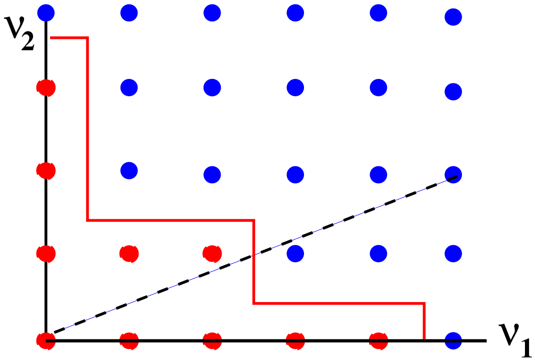
$$V_\Lambda := \left\{ \sum_{\nu \in \Lambda} v_\nu x^\nu, \ \ v_\nu \in V \right\} = V \otimes \mathbb{P}_\Lambda,$$

where $x^\nu = x_1^{\nu_1} \cdots x_d^{\nu_d}$.

We only consider downward closed index sets : $\nu \in \Lambda$ and $\mu \leq \nu \Rightarrow \mu \in \Lambda$.

Basis of $\mathbb{P}_\Lambda$ : tensorized orthogonal polynomials $L_\nu(x) = \prod_{j=1}^{d} L_{\nu_j}(x_j)$ for $\nu \in \Lambda$.

## Downward closed multivariate polynomials



Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2016) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_m \ldots$, with $m := \#(\Lambda_m)$ such that

$$\inf_{v \in V_m} \|u - v\|_{L^2(X,V,\rho)} \leq C m^{-s},$$

with $V_m := V_{\Lambda_m}$, where $\rho$ is any tensorized Jacobi measures. The exponent $s > 0$ is robust with respect to the dimension $d$.

Chkifa-Cohen-Nobile-Tempone (M2AN, 2014) : estimate $K_m$ for $\mathbb{P}_{\Lambda_m}$.

With $\rho = \otimes^d(\frac{dx}{2})$ the uniform distribution over $X$, one has $K_m \leq m^2$ for all downward closed sets $\Lambda_m$ such that $\#(\Lambda_m) = m$. Up to log factors, the stability regime is $n \gtrsim m^2$.

With the tensor-product Chebychev measure, improvement $K_m \leq m^\alpha$ with $\alpha := \frac{\log 3}{\log 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{i=1}^{d} x_j \psi_j, \quad x_i \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $X = \mathbb{R}^d$.

Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2016) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_m \ldots$, with $m := \#(\Lambda_m)$ such that

$$\inf_{v \in V_m} \|u - v\|_{L^2(X,V,\rho)} \leq Cm^{-s},$$

with $V_m := V_{\Lambda_m}$, where $\rho$ is any tensorized Jacobi measures. The exponent $s > 0$ is robust with respect to the dimension $d$.

Chkifa-Cohen-Nobile-Tempone (M2AN, 2014) : estimate $K_m$ for $\mathbb{P}_{\Lambda_m}$.

With $\rho = \otimes^d(\frac{dx}{2})$ the uniform distribution over $X$, one has $K_m \leq m^2$ for all downward closed sets $\Lambda_m$ such that $\#(\Lambda_m) = m$. Up to log factors, the stability regime is $n \gtrsim m^2$.

With the tensor-product Chebychev measure, improvement $K_m \leq m^\alpha$ with $\alpha := \frac{\log 3}{\log 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{j=1}^{d} x_j \psi_j, \quad x_j \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $X = \mathbb{R}^d$.

## Breaking the curse of dimensionality

Cohen-DeVore-Schwab (2011) + Bachmayr-Migliorati (2016) : approximation results.

Under suitable summability conditions on $(|\psi_j|)_{j\geq 1}$, there exists a sequence of downward closed sets $\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_m \dots$, with $m := \#(\Lambda_m)$ such that

$$\inf_{v\in V_m} \|u - v\|_{L^2(X,V,\rho)} \leq Cm^{-s},$$

with $V_m := V_{\Lambda_m}$, where $\rho$ is any tensorized Jacobi measures. The exponent $s > 0$ is robust with respect to the dimension $d$.

Chkifa-Cohen-Nobile-Tempone (M2AN, 2014) : estimate $K_m$ for $\mathbb{P}_{\Lambda_m}$.

With $\rho = \otimes^d(\frac{dx}{2})$ the uniform distribution over $X$, one has $K_m \leq m^2$ for all downward closed sets $\Lambda_m$ such that $\#(\Lambda_m) = m$. Up to log factors, the stability regime is $n \gtrsim m^2$.

With the tensor-product Chebychev measure, improvement $K_m \leq m^\alpha$ with $\alpha := \frac{\log 3}{\log 2}$.

The theory and least-square method is not capable of handling lognormal diffusions :

$$a = \exp(b), \quad b = \sum_{i=1}^{d} x_j\psi_j, \quad x_i \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

which corresponds to the tensor product Gaussian measure over $X = \mathbb{R}^d$.

In the weighted least-square method, we sample according to $d\mu$ such that $d\rho = w d\mu$.

The stability condition is $K_{m,w} \lesssim \frac{n}{\log m}$, where $K_{m,w} := \sup_{x \in X} w(x) k_m(x)$.

Optimal choice : take

$$w(x) = w_m(x) = \frac{m}{k_m(x)} \iff d\mu = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^{m} |L_j|^2\Big) d\rho,$$

Then $d\mu$ is a probability measure and we have $k_{m,w} = m$.

Therefore, up to log factors, the stability regime is $n \gtrsim m$ independently of $\rho$.

In the weighted least-square method, we sample according to $d\mu$ such that $d\rho = w d\mu$.

The stability condition is $K_{m,w} \lesssim \frac{n}{\log m}$, where $K_{m,w} := \sup_{x\in X} w(x) k_m(x)$.

Optimal choice : take

$$w(x) = w_m(x) = \frac{m}{k_m(x)} \iff d\mu = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^m |L_j|^2\Big) d\rho,$$

Then $d\mu$ is a probability measure and we have $k_{m,w} = m$.

Therefore, up to log factors, the stability regime is $n \gtrsim m$ independently of $\rho$.

In the weighted least-square method, we sample according to $d\mu$ such that $d\rho = w d\mu$.

The stability condition is $K_{m,w} \lesssim \frac{n}{\log m}$, where $K_{m,w} := \sup_{x\in X} w(x) k_m(x)$.

Optimal choice : take

$$w(x) = w_m(x) = \frac{m}{k_m(x)} \iff d\mu = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^m |L_j|^2\Big) d\rho,$$

Then $d\mu$ is a probability measure and we have $k_{m,w} = m$.

Therefore, up to log factors, the stability regime is $n \gtrsim m$ independently of $\rho$.

## Optimal sampling for weighted least-squares

In the weighted least-square method, we sample according to $d\mu$ such that $d\rho = w d\mu$.

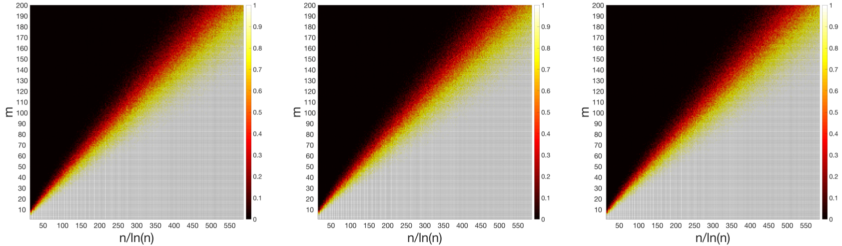
The stability condition is $K_{m,w} \lesssim \frac{n}{\log m}$, where $K_{m,w} := \sup_{x\in X} w(x)k_m(x)$.

Optimal choice : take

$$w(x) = w_m(x) = \frac{m}{k_m(x)} \iff d\mu = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^{m} |L_j|^2\Big) d\rho,$$

Then $d\mu$ is a probability measure and we have $k_{m,w} = m$.

Therefore, up to log factors, the stability regime is $n \gtrsim m$ independently of $\rho$.




Stability regime for univariate polynomials with $\rho$ Chebychev, uniform, and Gaussian.

## Sampling the optimal density

The optimal sampling measure $\mu$ now depends on $V_m$ :

$$d\mu = d\mu_m = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^m |L_j|^2\Big) d\rho.$$

In the case of parametric PDEs approximated with multivariate polynomials, $d\rho$ is a product measure (easy to sample), but $d\mu_m$ is not.

Sampling strategies in high dimension :

(i) Monte Carlo Markov Chain (MCMC) : generate by simple recursive rules a sample such that the the probability distribution asymptotically approaches $d\mu_m$.

(ii) Conditional sampling : obtains first component by sampling the marginal $d\mu_1(x_1)$, then the second component by sampling the conditional marginal probability $d\mu_{x_1}(x_2)$ for this choice of the first component, etc...

This second strategy is more efficient on our cases of interest.

## Sampling the optimal density

The optimal sampling measure $\mu$ now depends on $V_m$ :

$$d\mu = d\mu_m = \frac{k_m}{m} d\rho = \frac{1}{m}\Big(\sum_{j=1}^{m} |L_j|^2\Big) d\rho.$$

In the case of parametric PDEs approximated with multivariate polynomials, $d\rho$ is a product measure (easy to sample), but $d\mu_m$ is not.

Sampling strategies in high dimension :

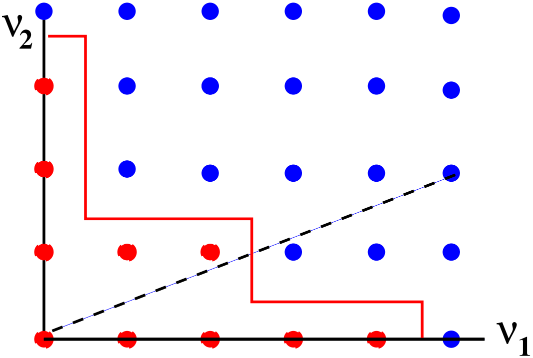
(i) Monte Carlo Markov Chain (MCMC) : generate by simple recursive rules a sample such that the the probability distribution asymptotically approaches $d\mu_m$.

(ii) Conditional sampling : obtains first component by sampling the marginal $d\mu_1(x_1)$, then the second component by sampling the conditional marginal probability $d\mu_{x_1}(x_2)$ for this choice of the first component, etc...

This second strategy is more efficient on our cases of interest.

## Adaptivity

Update adaptively the polynomial space $\Lambda_{m-1} \to \Lambda_m$, while increasing the amount of sample necessary for stability $n = n(m) \sim m \log m$.

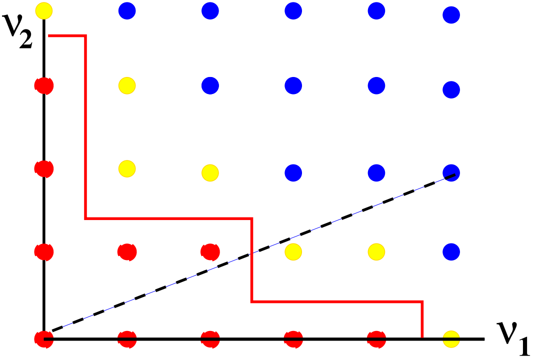



Problem : the optimal measure $\mu = \mu_m$ changes as we vary $m$. How should we recycle the previous samples ?

For certain simple cases $\mu_m \sim \mu^*$ as $m \to \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

# Adaptivity

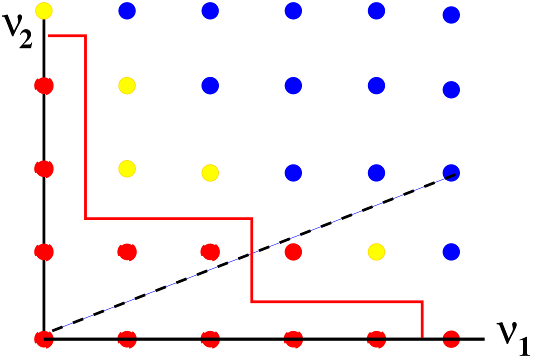
Update adaptively the polynomial space $\Lambda_{m-1} \to \Lambda_m$, while increasing the amount of sample necessary for stability $n = n(m) \sim m \log m$.




Problem : the optimal measure $\mu = \mu_m$ changes as we vary $m$. How should we recycle the previous samples ?

For certain simple cases $\mu_m \sim \mu^*$ as $m \to \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

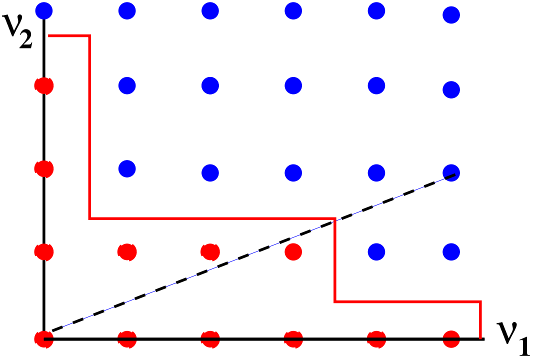
Update adaptively the polynomial space $\Lambda_{m-1} \to \Lambda_m$, while increasing the amount of sample necessary for stability $n = n(m) \sim m \log m$.




Problem : the optimal measure $\mu = \mu_m$ changes as we vary $m$. How should we recycle the previous samples ?

For certain simple cases $\mu_m \sim \mu^*$ as $m \to \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

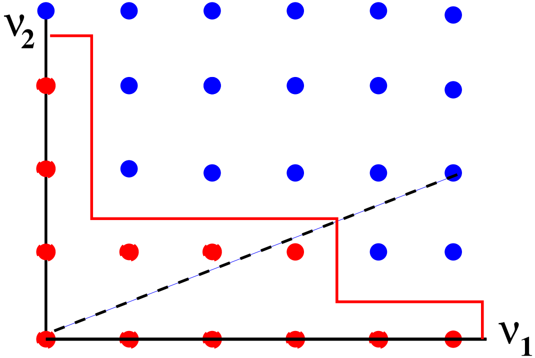
Update adaptively the polynomial space $\Lambda_{m-1} \to \Lambda_m$, while increasing the amount of sample necessary for stability $n = n(m) \sim m \log m$.




Problem : the optimal measure $\mu = \mu_m$ changes as we vary $m$. How should we recycle the previous samples ?

For certain simple cases $\mu_m \sim \mu^*$ as $m \to \infty$ (equilibrium measure for univariate polynomials on $[-1, 1]$). But no such asymptotic in general cases.

Update adaptively the polynomial space $\Lambda_{m-1} \to \Lambda_m$, while increasing the amount of sample necessary for stability $n = n(m) \sim m \log m$.




Problem : the optimal measure $\mu = \mu_m$ changes as we vary $m$. How should we recycle the previous samples ?

For certain simple cases $\mu_m \sim \mu^*$ as $m \to \infty$ (equilibrium measure for univariate polynomials on $[-1,1]$). But no such asymptotic in general cases.

## Sequencial sampling

Observe that

$$d\mu_m = \frac{1}{m}\Big(\sum_{j=1}^m |L_j|^2\Big) d\rho = \Big(1 - \frac{1}{m}\Big) d\mu_{m-1} + \frac{1}{m} d\nu_m \quad \text{where } d\nu_m = |L_m|^2 d\rho.$$

We use this mixture property to generate the sample in an incremental manner.

Assume that the sample $S_{m-1} = \{x_1, \dots, x_{n(m-1)}\}$ have been generated by independent draw according to the distribution $d\mu_{m-1}$.

Then we generate a new sample $S_m = \{x_1, \dots, x_{n(m)}\}$ as follows :

For each $i = 1, \dots, n(m)$, pick Bernoulli variable $b_i \in \{0, 1\}$ with probability $\{\frac{1}{m}, 1 - \frac{1}{m}\}$.

If $b_i = 0$, generate $x_i$ according to $d\nu_m$.

If $b_i = 1$, pick $x_i$ incrementally inside $S_{m-1}$. If $S_{m-1}$ has been exhausted generate $x_i$ according to $d\mu_{m-1}$.

Arras-Bachmayr-Cohen-Migliorati (2018) : the total number of sample $\tilde{n}(m)$ used at stage $m$ satisfies $\mathbb{E}(\tilde{n}(m)) \sim m \log(m)$ and $\tilde{n}(m) \lesssim m \log(m)$ with high probability for all values of $m$. With high probability, the matrix $\mathbf{G}$ satisfies $\kappa(\mathbf{G}) \leq 3$ for all values of $m$.

## Sequencial sampling

Observe that

$$d\mu_m = \frac{1}{m}\Big(\sum_{j=1}^m |L_j|^2\Big) d\rho = \Big(1-\frac{1}{m}\Big) d\mu_{m-1} + \frac{1}{m} d\nu_m \quad \text{where } d\nu_m = |L_m|^2 d\rho.$$

We use this mixture property to generate the sample in an incremental manner.

Assume that the sample $S_{m-1} = \{x_1, \dots, x_{n(m-1)}\}$ have been generated by independent draw according to the distribution $d\mu_{m-1}$.

Then we generate a new sample $S_m = \{x_1, \dots, x_{n(m)}\}$ as follows :

For each $i = 1, \dots, n(m)$, pick Bernoulli variable $b_i \in \{0,1\}$ with probability $\{\frac{1}{m}, 1-\frac{1}{m}\}$.

If $b_i = 0$, generate $x_i$ according to $d\nu_m$.

If $b_i = 1$, pick $x_i$ incrementally inside $S_{m-1}$. If $S_{m-1}$ has been exhausted generate $x_i$ according to $d\mu_{m-1}$.

Arras-Bachmayr-Cohen-Migliorati (2018) : the total number of sample $\tilde{n}(m)$ used at stage $m$ satisfies $\mathbb{E}(\tilde{n}(m)) \sim m\log(m)$ and $\tilde{n}(m) \lesssim m\log(m)$ with high probability for all values of $m$. With high probability, the matrix $\mathbf{G}$ satisfies $\kappa(\mathbf{G}) \leq 3$ for all values of $m$.

# Sequencial sampling

Observe that

$$d\mu_m = \frac{1}{m}\Big(\sum_{j=1}^{m} |L_j|^2\Big) d\rho = \Big(1-\frac{1}{m}\Big) d\mu_{m-1} + \frac{1}{m} d\nu_m \quad \text{where } d\nu_m = |L_m|^2 d\rho.$$

We use this mixture property to generate the sample in an incremental manner.

Assume that the sample $S_{m-1} = \{x_1, \dots, x_{n(m-1)}\}$ have been generated by independent draw according to the distribution $d\mu_{m-1}$.

Then we generate a new sample $S_m = \{x_1, \dots, x_{n(m)}\}$ as follows :

For each $i = 1, \dots, n(m)$, pick Bernoulli variable $b_i \in \{0, 1\}$ with probability $\{\frac{1}{m}, 1 - \frac{1}{m}\}$.

If $b_i = 0$, generate $x_i$ according to $d\nu_m$.

If $b_i = 1$, pick $x_i$ incrementally inside $S_{m-1}$. If $S_{m-1}$ has been exhausted generate $x_i$ according to $d\mu_{m-1}$.

Arras-Bachmayr-Cohen-Migliorati (2018) : the total number of sample $\tilde{n}(m)$ used at stage $m$ satisfies $\mathbb{E}(\tilde{n}(m)) \sim m \log(m)$ and $\tilde{n}(m) \lesssim m \log(m)$ with high probability for all values of $m$. With high probability, the matrix $\mathbf{G}$ satisfies $\kappa(\mathbf{G}) \le 3$ for all values of $m$.

## Conclusions

Optimal sampling yields stable least-squares method under the regime $n \sim m \log m$.

Applicable to any measure $\rho$ and spaces $V_m$, in any dimension.

Optimality can be preserved in a sequencial framework.

Perspective : adaptive weighted least-squares strategies for the selection of $\Lambda_m$.

Convergence results are in expectation or in probability. Deterministic sampling ?

## References

A. Cohen, M. Davenport and D. Leviatan, *On the stability and accuracy of least-squares approximation*, FoCM 2013.

A. Cohen and R. DeVore, *High dimensional approximation of parametric PDEs*, Acta Numerica 2015.

A. Doostan and J. Hampton, *Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression*, Computer Methods in Applied Mechanics and Engineering 2015.

A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*, SMAI J. of Comp. Math. 2017.

A. Doostan and M. Hadigol, *Least squares polynomial chaos expansion : A review of sampling strategies*, Computer Methods in Applied Mechanics and Engineering 2018.

## Conclusions

Optimal sampling yields stable least-squares method under the regime $n \sim m \log m$.

Applicable to any measure $\rho$ and spaces $V_m$, in any dimension.

Optimality can be preserved in a sequencial framework.

Perspective : adaptive weighted least-squares strategies for the selection of $\Lambda_m$.

Convergence results are in expectation or in probability. Deterministic sampling ?

## References


A. Cohen, M. Davenport and D. Leviatan, *On the stability and accuracy of least-squares approximation*, FoCM 2013.

A. Cohen and R. DeVore, *High dimensional approximation of parametric PDEs*, Acta Numerica 2015.

A. Doostan and J. Hampton, *Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression*, Computer Methods in Applied Mechanics and Engineering 2015.

A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*, SMAI J. of Comp. Math. 2017.

A. Doostan and M. Hadigol, *Least squares polynomial chaos expansion : A review of sampling strategies*, Computer Methods in Applied Mechanics and Engineering 2018.

## Conclusions

Optimal sampling yields stable least-squares method under the regime $n \sim m \log m$.

Applicable to any measure $\rho$ and spaces $V_m$, in any dimension.

Optimality can be preserved in a sequencial framework.

Perspective : adaptive weighted least-squares strategies for the selection of $\Lambda_m$.

Convergence results are in expectation or in probability. Deterministic sampling ?

## References


A. Cohen, M. Davenport and D. Leviatan, *On the stability and accuracy of least-squares approximation*, FoCM 2013.

A. Cohen and R. DeVore, *High dimensional approximation of parametric PDEs*, Acta Numerica 2015.

A. Doostan and J. Hampton, *Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression*, Computer Methods in Applied Mechanics and Engineering 2015.

A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*, SMAI J. of Comp. Math. 2017.

A. Doostan and M. Hadigol, *Least squares polynomial chaos expansion : A review of sampling strategies*, Computer Methods in Applied Mechanics and Engineering 2018.