Inverse Parameter Estimation

Ronald DeVore

Collaborators: Andrea Bonito, Albert Cohen, Guergana Petrova, Gerrit Welper

Motivation

- Parametric PDEs are used to model complex physical systems
- Uncertainty Quantification: We may have uncertainty in the parameters (or even the model)
- However we have some information (measurements) of the state (solution to the pde)
- Using this information, what can we say about the parameters giving rise to this state?
- This talk concerns rigorous theory to answer this question

The Setting

- We limit ourselves to the friendly setting of Parametric Elliptic PDEs
- $D \subset \mathbb{R}^d$ is a Lipschitz domain and $\mathcal{A}$ is the collection of diffusion coefficients $a \in L_\infty(D)$ that satisfy the Uniform Ellipticity Assumption

$$\mathrm{UEA}: \quad 0 < r \le a(x) \le R, \quad x \in D, \quad \text{for all } a \in \mathcal{A}$$

- For each $a \in \mathcal{A}$ we are interested in the solution u_a to

$$\begin{aligned} -\mathrm{div}(a(x)\nabla u_a(x)) &= f(x), \quad && x \in D, \\ u_a(x) &= 0, && x \in \partial D \end{aligned}$$

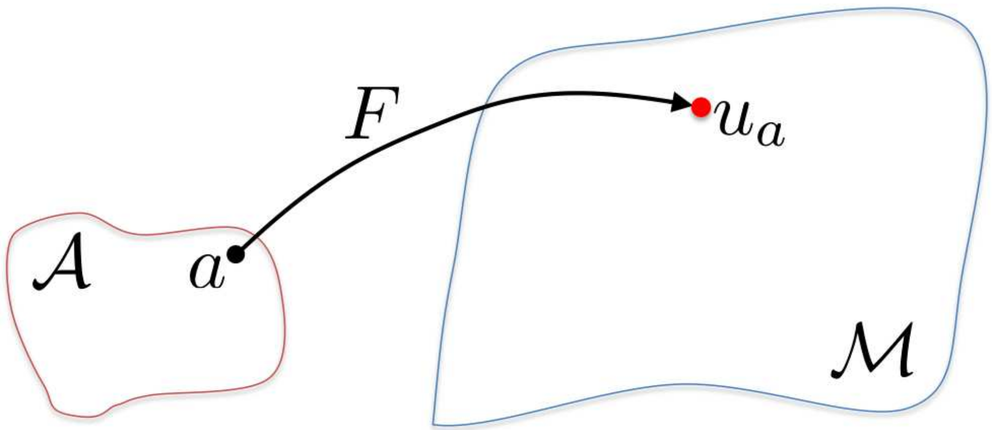
- Let $\mathcal{M} := \mathcal{M}(f, \mathcal{A}) = \{u_a : \ a \in \mathcal{A}\}$ be the solution manifold and $F : a \mapsto u_a$ the solution map

Additional Structure

- Usually we work with subsets $\mathcal{A}_0 \subset \mathcal{A}$ which impose additional structure on the diffusion coefficients
- The affine model: a satisfies UEA, i.e. $a \in \mathcal{A}$ and
  - $a(x,y) = \bar{a}(x) + \sum_{j=1}^{\infty} y_j \psi_j(x)$, $y_j \in [-1,1]$, $j = 1, 2, \dots$
  - Notation: $\mathcal{P} := [-1,1]^{\mathcal{N}}$ the set of parameters and $u(x,y) := u_a(x)$
  - We typically impose further restrictions on the affine decomposition such as decay for the $\|\psi_j\|_{L_\infty(D)}$, for example $(\|\psi_j\|_{L_\infty(D)})_{j\geq 1} \in \ell_p$ with $p < 1$
- Second example: $\mathcal{A}_s := \mathcal{A}_s(M) := \{a \in \mathcal{A} : \|a\|_{H^s} \leq M\}$
  - Note in this example we still have the condition that $a \in L_\infty(D)$

Parameter Identification

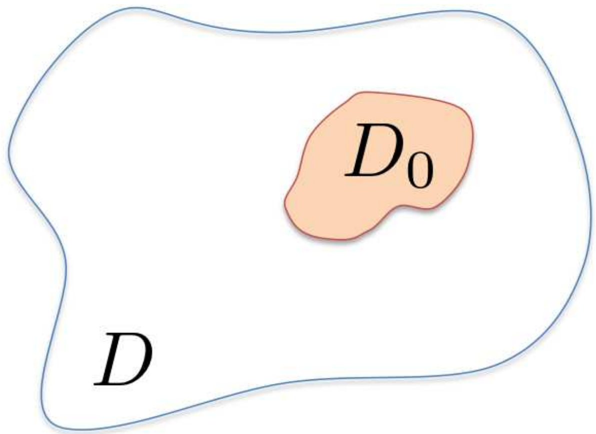
- First Question: Does u_a determine a?
- We fix f and ask whether the solution map $F : a \to u_a$ is invertible



- The answer depends on f.

Parameter Identification

- If for some $a \in \mathcal{A}$ we have ∇u_a vanishes on an open subset $D_0 \subset D$ then for any b which agrees with a outside of D_0, we have $u_a = u_b$ and therefore there is no uniqueness



- To avoid this, we assume always that $f \in L_\infty(D)$ and $f > 0$ on D
- Problem 1: Does this guarantee unique invertibility of F

The forward map

- Before analyzing the inverse map we recall results about the smoothness of the forward map F
  - The usual estimate is
    $\|u_a - u_b\|_{H^1(D)} \le \frac{\|f\|_{H^{-1}}}{r^2} \|a-b\|_{L_\infty(D)}$
  - The above is not useful when a, b have jump discontinuities that do not match
  - Improved estimates (Bonito-DeVore-Nochetto): If $p \ge 2$ and $q := \frac{2p}{p-2}$ then
    $$\|u_a - u_b\|_{H^1(D)} \le r^{-1} \|\nabla u_a\|_{L_p(D)} \|a-b\|_{L_q(D)}, \quad q = \frac{2p}{p-2}$$
  - Note that since $a \in L_\infty(D)$, we obtain
    $$\|u_a - u_b\|_{H^1(D)} \le C \|\nabla u_a\|_{L_p(D)} \|a-b\|^{\theta}_{L_2(D)}, \quad \theta = 2/q$$

Sufficient conditions

- The previous result requires $\nabla u_a \in L_p$
- Sufficient conditions on a which guarantee $\nabla u_a \in L_p$?
  - There is always a range of $p > 2$ (depending only on D), i.e. $2 \leq p < P$. where this is true for all $a \in \mathcal{A}$
    - Hence there is always a $\theta = \theta(D) > 0$ such that for all $a \in \mathcal{A}$ we have

$$\|u_a - u_b\|_{H^1(D)} \leq C \, \|a - b\|^{\theta}_{L_2(D)}$$

  - if $a \in \mathrm{VMO}$ then $\nabla u_a \in L_p(D)$ for all $p < \infty$
    - Hence for all $0 < \theta < 1$ and all $a \in \mathcal{A} \cap VMO$ we have

$$\|u_a - u_b\|_{H^1(D)} \leq C \, \|a - b\|^{\theta}_{L_2(D)}$$

    - Here C depends on p and the VMO modulus of a

Inverse Map 1D

- If $D = [0,1]$ and $f \geq c > 0$ is in $L_\infty(D)$ the analysis is simple
  - For any $a, b \in \mathcal{A}$ we have $\|u_a - u_b\|_{H_1} \leq C\|a-b\|_{L_2[0,1]}$
  - For any $a, b \in \mathcal{A}$ we have $\|a-b\|_{L_2[0,1]} \leq C\|u_a - u_b\|_{H_1}^{1/3}$
  - The exponent $1/3$ cannot be improved
  - Notice that these results hold with no additional assumptions on a other than UEA, i.e. $a \in \mathcal{A}$

Higher Dimensions

- In higher space dimension $d \geq 2$ the situation is more complex and the results are not as complete
- We assume $f \geq c > 0$ and $f \in L_\infty(D)$, with D Lipschitz
  - In this setting, I do not even know if u_a uniquely determines $a \in \mathcal{A}$ (see Problem 1)
- We can show unique determination of a and smoothness for the inverse map $u_a \mapsto a$ provided we impose extra conditions on the diffusion coefficients a
- In Bonito-Cohen-DeVore-Petrova-Welper we prove results of the type

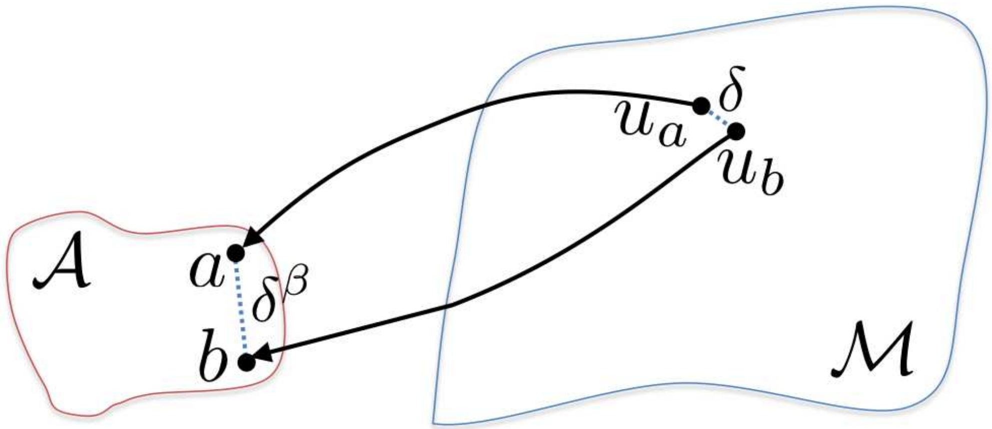
$$\|a - b\|_{L_2(D)} \leq C\|u_a - u_b\|_{H^1(D)}^{\beta}$$

  - In otherwords, we prove that the inverse map is $\mathrm{Lip}\ \beta$ under additional assumptions on the diffusion coeff.

Values of β

- Under the additional assumption that the diffusion coefficients are in $\mathcal{A}_1(D)$ we have $\beta = 1/6$
- Under the additional assumption that the diffusion coefficients are in $\mathcal{A}_s(D) \cap \mathrm{VMO}(\varphi)$ with $s > 1/2$, we can prove that there is $\beta = \beta(s) > 0$
- We can drop the VMO requirement provided $a, b \in \mathcal{A}_s(D)$ and $s > s^*$ with $s^* < 1$ depending only on D
- These results do not apply if a, b are piecewise constant. However, in this case we have the following:
  - Let $\mathcal{P}_n$ be the partition of $D = [0,1]^d$ into n^d cubes of equal side length $1/n$ and let $\mathcal{A}^n$ be the set of diffusion coefficients in $\mathcal{A}$ that are piecewise constant subordinate to $\mathcal{P}_n$
  - $\|a - b\|_{L_2} \leq Cn\|u_a - u_b\|_{H^1(D)}, \quad a, b \in \mathcal{A}^n$

Lip β smoothness of inverse map



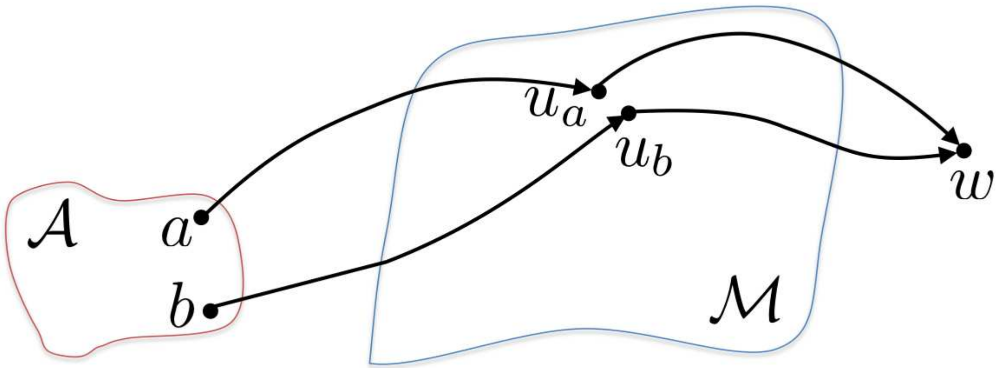
Summary

- Under moderate assumptions on the diffusion coeff.
  - $\|u_a - u_b\|_{H_0^1(D)} \leq C\|a-b\|_{L_2(D)}^{\alpha}$ for some $\alpha > 0$
  - $\|a-b\|_{L_2(D)} \leq C\|u_a - u_b\|_{H_0^1(D)}^{\beta}$ for some $\beta > 0$
  - If we observe the full state u_a this still does not tell us how to find a. In most settings, we do not observe the full state u_a but rather just partial information, namely, a finite number of measurements of the state
  - The remainder of this talk will address how well we can expect to recover a with this partial information. These are difficult questions- results are limited
  - In Uncertainty Quantification one assumes that parameters occur with an underlying probability distribution: the most closely related results are in Schwab-Stuart - IP 2012

The Numerical Setting

- We assume that we have a finite number of measurements $l_j(u_a) = w_j$, $j = 1, \dots, m$, of the state u_a
  - Since we are in a Hilbert space $\mathcal{H} := H_0^1(D)$ we can write $l_j = \langle \cdot, \omega_j \rangle$ with $\omega_j \in \mathcal{H}$
  - We let $W := \mathrm{span}\{\omega_1, \dots, \omega_m\rangle$
  - Then we can view the information we have as we are given $w = P_W(u_a)$
- Let $\mathcal{A}_0 \subset \mathcal{A}$ where membership in $\mathcal{A}_0$ may impose additional smoothness conditions on a. Once we have $\mathcal{A}_0$ fixed there is an α, β
- Notice that there are typically many $a \in \mathcal{A}_0$ for which $P_W(u_a) = w$ and so we need to clarify our goal.

Non-uniqueness of $M(u_a) = w$



$$M(u_a) = M(u_b) = w \in \mathbb{R}^m$$

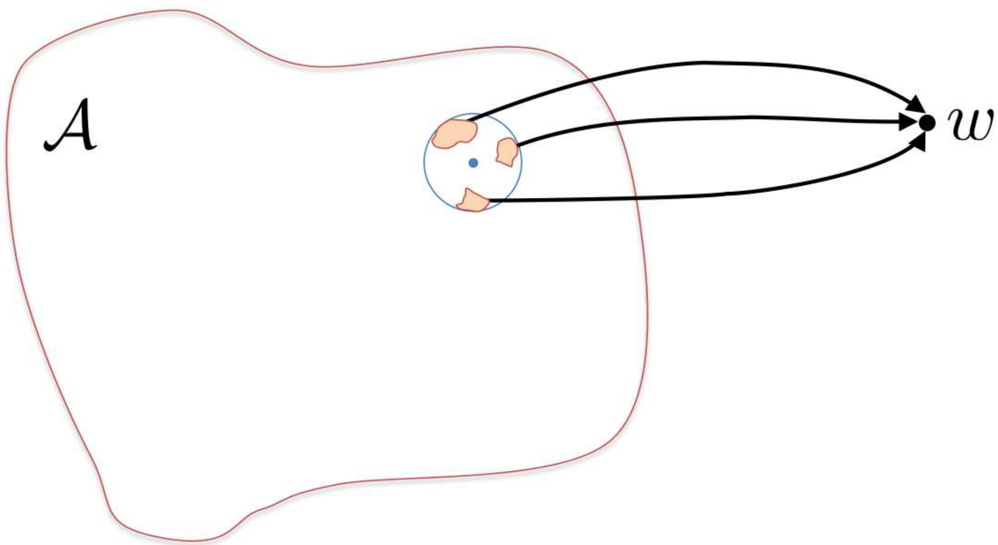
Goals

- Given any $w \in W$ (which may or may not be a measurement of some u_a, $a \in \mathcal{A}_0$) define the sets $S_w(\eta) := \{b \in \mathcal{A}_0 : \|P_W(u_b) - w\|_{L_2(D)} \leq \eta\}, \quad \eta \geq 0$
- Ideal Goal: Describe $\mathcal{S}_w(0)$
- This is too demanding for several reasons
  - Noise: If measurements are noisy, say we observe $\hat{w}$ then the a we seek is only in $\mathcal{S}_w(\delta)$ for some $\delta > 0$ depending on the noise level
  - Numerical issues: We cannot expect to compute an a in $\mathcal{S}_w(\eta)$ only an approximation to such an a
  - Computational resources: Decreasing η will eat up more and more computational resources eventually becoming unreasonable

Possible Goals: Smallest Ball

- The user provides a tolerance $\eta \geq 0$,
- Smallest Ball: Find a ball $B(a^*, R^*)$ in $L_2(D)$ such that $a^* \in \mathcal{A}_0$ and $\mathcal{S}_w(\eta) \subset B(a^*, R^*)$ with the ball as small as possible:
  - The smallest ball is the Chebyshev ball of $\mathcal{S}_w(\eta)$
  - a^* would give a (coarse) approximation to all possible a

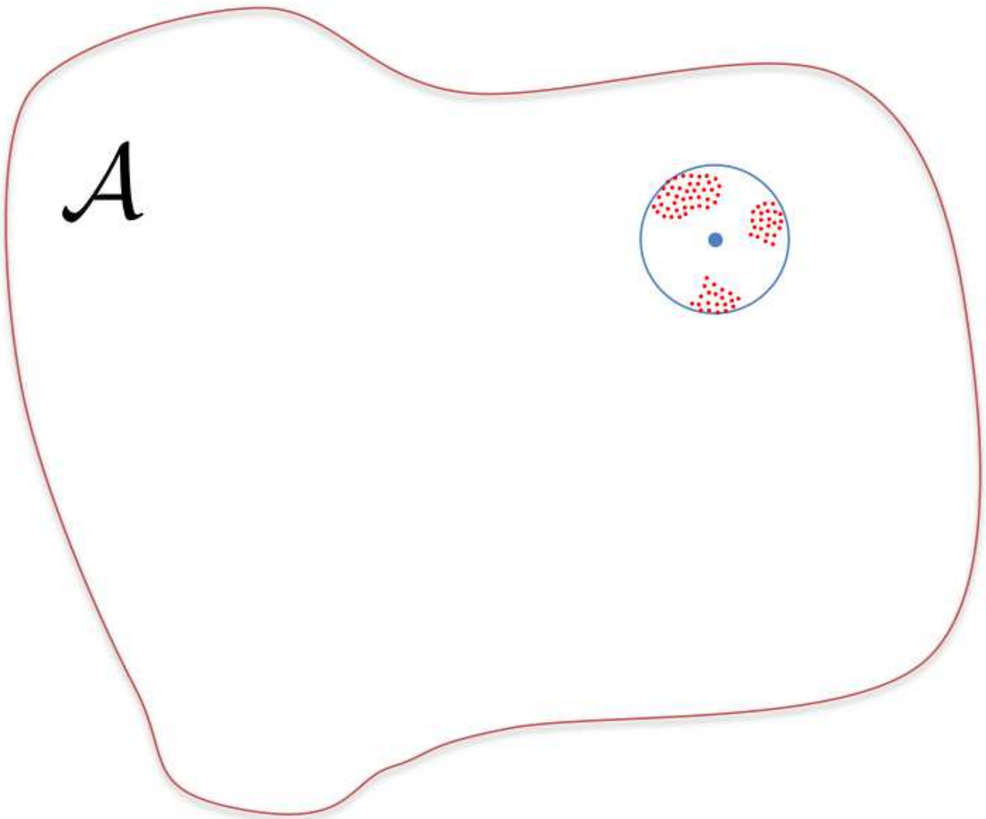
Smallest ball for $\mathcal{S}_w(0)$



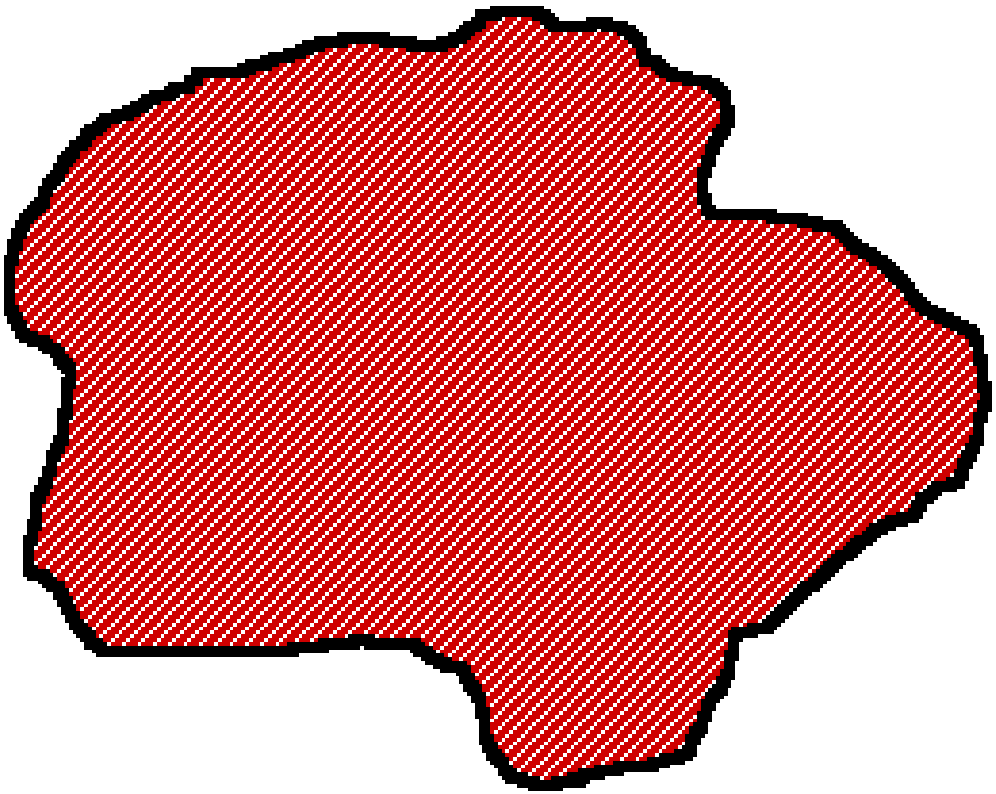
Possible Goals: Sketch

- The user provides a tolerance $\eta \geq 0$,
- Sketch: Find a small discrete set $\hat{S}$ that gives an ϵ net for $S_w(\eta)$
  - Smallest set is the entropy cover of $\mathcal{S}_w(\eta)$
  - Therefore we would like cardinality of $\hat{S}$ to be comparable to the covering number $N_\epsilon(\mathcal{S}_w(\eta))$

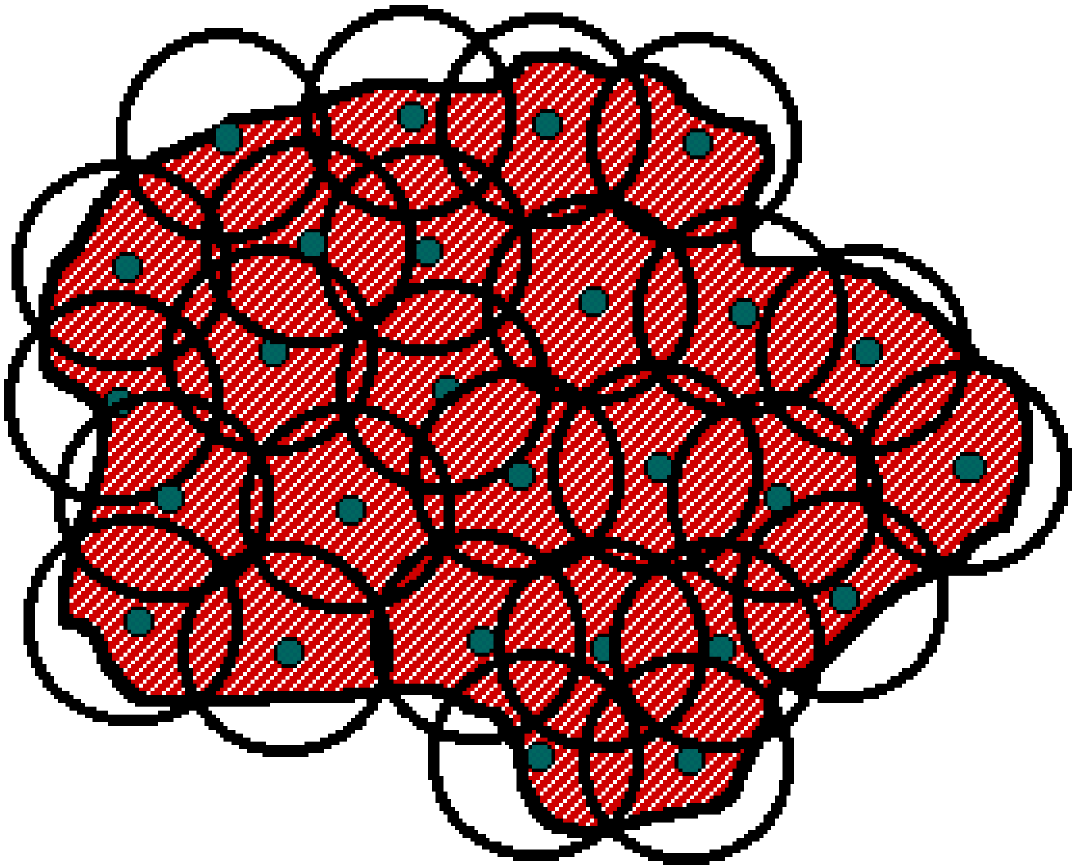
Sketch for $\mathcal{S}_w(\eta)$



ϵ net



Covering



Smallest Ball

- Algorithms for finding the smallest ball have three components
  - I. Use w to find $\hat{u} \in H^1(D)$ and $\hat{R}$ such that $B(\hat{u}, \hat{R})$ contains all u_a, $a \in \mathcal{A}_0$, such that $M(u_a) = w$
  - II. Find $b \in \mathcal{A}_0$ such that u_b approximates $\hat{u}$ at least to the precision $\hat{R}$
  - III. Use the smoothness of the inverse map and the knowledge of u_b to find a ball $B(b, \tilde{R})$ which contains all $a \in \mathcal{A}_0$ such that $M(u_a) = w$
- III. Our inverse theorem gives $\tilde{R} \leq C(2\hat{R})^\beta$. Indeed,

$$\|a - b\|_{L_2(D)} \leq C\|u_a - u_b\|^\beta_{H^1_0(D)} \leq C(2\hat{R})^\beta$$

- So Task III is easy once the other tasks are complete

I. by Reduced Modeling

- Task I is complicated by the fact that the solution manifold $\mathcal{M} := \{u_a : \ a \in \mathcal{A}_0\}$ is not easy to understand
- Strategy is to replace $\mathcal{M}$ by a reduced model
- Such models produce a low dimensional linear space $V \subset H^1(D)$ such that $\mathrm{dist}(\mathcal{M}, V)$ is small enough to complete Task I
- Two strategies for doing this
  - Greedy Algorithms
  - High dimensional polynomial expansions

Greedy algorithms

- These algorithms choose (through greedy selection) snapshots $v_1 = u_{a_1}, \dots, v_n := u_{a_n}$ so that $V_n := \mathrm{span}\{v_1, \dots, v_n\}$ is a good approximation to $\mathcal{M}$
- Greedy strategy introduced by Buffa-Maday-Patera-Prud'homme-Turinici chooses the k-th snapshot which is furthest from V_{k-1}
- Theorem (Binev-Cohen-Dahmen-DeVore-Petrova-Wojtaszczyk)
  - If there exist n dimensional spaces $Y_n \subset \mathcal{H}_0^1(D)$ such that $\mathrm{dist}(\mathcal{M}, Y_n) \le C n^{-\alpha}, \quad n = 1, \dots, N$ then $\mathrm{dist}(\mathcal{M}, V_n) \le C' n^{-\alpha}, \quad n = 1, \dots, N$
  - Almost optimal in terms of n widths
  - These algorithms have a very costly off-line implementation

Polynomial Expansions

- Cohen-DeVore-Schwab I,II,
  Chkifa-Cohen-DeVore-Schwab, +
- If the a have an affine expansion with
  $(\|\psi_j\|_{L_\infty(D)})_{j\geq 1} \in \ell_p, \quad p<1$
- Then $u(x,y) = \sum_\nu u_\nu(x) y^\nu$ with $(\|u_\nu\|_{H_0^1(D)}) \in \ell_p$
- It follows that for each $n \geq 1$, there is a set Λ_n such that
  - $\#(\Lambda_n) = n$
  - $\sup_{y\in\mathcal{P}} \|u(\cdot,y) - \sum_{\nu\in\Lambda_n} u_\nu y^\nu\|_{H_0^1(D)} \leq Cn^{-1/p+1}$
- This gives certifiable decay of n widths of $\mathcal{M}$
- u_ν found by recursively solving PDEs
- Finding Λ_n costly

Assimilating Data

- Take a reduced space $V = V_n$: what is a good choice for V will be uncovered as we proceed
- Let $\mathcal{N}$ be the null space of the measurement map M
- Define $\mu(\mathcal{N}, V) := \sup_{\eta \in \mathcal{N}} \frac{\|\eta\|_{H^1}}{\operatorname{dist}(\eta, V)_{H^1}}$
  - μ is the reciprocal of the angle between V and W
- Let $v^*(w) = \operatorname{Argmin}_{v \in V} \|w - M(v)\|_{\ell_2}$
- Then, Maday-Patera -Penn-Yano show that the ball $B(v^*(w), \hat{R})$, with $\hat{R} := 2\mu(\mathcal{N}, V^*)\operatorname{dist}(\mathcal{M}, V^*)_{H^1}$, contains all $u_a \in \mathcal{M}$ such that $M(u_a) = w$. So we take $\hat{u} := v^*(w)$
- The best choice V^* is one which minimizes $\mu(\mathcal{N}, V)\operatorname{dist}(\mathcal{M}, V)_{H^1}$ over all $V \subset H_0^1$. This would complete Task I with $\hat{R} = 2\mu(\mathcal{N}, V^*)\operatorname{dist}(\mathcal{M}, V^*)_{H^1}$

Task II

- We know $\hat{u} := v^*(w)$ and we want to find $b \in \mathcal{A}_0$ such that $\|v^*(w) - u_b\|_{H_0^1(D)} \leq C\hat{R}$
- As long as $C \geq 2$ we know there are such b
- One way to find such a b is to search over a (minimal) set $\mathcal{A}^n \subset \mathcal{A}_0$ such that $\mathcal{A}^n = \{a_j\}$ is an ϵ net for $\mathcal{A}_0$ with $\epsilon := (\frac{\hat{R}}{C})^{1/\alpha}$
  - Indeed, we know from our results on the forward map that $\|u_a - u_{a'}\|_{H^1} \leq C_0\|a - b\|^\alpha$
  - Hence, the u_{a_j} are an ϵ' net for $\mathcal{A}_0$ with $\epsilon' = C_0\hat{R}/C$
  - We use C so that we only have to approximately solve for u_a using the reduced space V^*
  - In fact, we never solve for u_a but rather use surrogate error estimators based on residuals- these are fast!

Post Mortem

- The bottlenecks in the above algorithm for finding a ball are
  - Finding the space V^*
    - Can we do this via greedy selection?
    - The usual greedy algorithms do not take into account $\mu(\mathcal{N}, V)$
  - The discretization of $\mathcal{A}^n$ - this manifests itself when the number of parameters is large
  - For an affine model of the parameters, we quantize the y_j with fine quantization when $\|\psi_j\|_{L_\infty(D)}$ is large and coarse quantization when it is small

Finding a sketch of $S_w(\eta)$

- A dream algorithm for sketching would be one which identifies an ϵ net for $S_w(\eta)$ whose size and computational costs are proportional to $N_\epsilon(\mathcal{S}_w(\eta))$
- We proceed to describe the main ingredients of such algorithms in the case of the affine model
- One constructs recursively
  - Discretizations $\mathcal{A}^1, \mathcal{A}^2, \dots$ of $\mathcal{A}_0$ using quantization of the y_j as described earlier
  - Reduced model spaces $V_1, V_2, \dots$ with control on $\mu(\mathcal{N}, V_n)\mathrm{dist}(\mathcal{M}, V_n)$
  - Using residual error estimators one can define cheap surrogates for computing $\|w - M(u_a)\|_{\ell_2}$

Testing points in $\mathcal{A}^n$

- Points in $\mathcal{A}^n$ can then be tested, i.e., one computes an approximation to $\|w - M(u_a)\|_{\ell_2}$ at the needed accuracy and thereby $\mathcal{A}^n$ can be decomposed into subsets
  - $\mathcal{A}^n(out)$: These are points in $\mathcal{A}^n$ which one can not only say these points can not be in $\mathcal{S}_w(\eta)$ but also regions of $\mathcal{A}_0$ near these points can be eliminated from firther consideration because the residual error is too large. Here one uses the direct and inverse estimates.
  - $\mathcal{A}^n(in)$: These are points in $\mathcal{A}^n$ that cannot be eliminated because the residual error estimate is not large enough
- The sets $\mathcal{A}^n(in)$ give finer and finer nets for $\mathcal{S}_w(\eta)$

Bottlenecks

- As before finding good educed model spaces $\mu(\mathcal{N}, V)\mathrm{dist}(\mathcal{M}, V)$
  - The usual greedy algorithms or polynomial basis selections do not pay attention to μ - naturally because they were not formulated with measurements in mind
  - Greedy algorithms are numerically intensive
- The cardinality of the sets $\mathcal{A}^n(in)$ grow exponentially in n limiting how large one can choose n
- This may lie in the nature of the problem since ϵ nets typically grow like $\epsilon^{-\tau}$ with τ moderately large
- It would be good to have a priori theoretical bounds for $N_\epsilon(\mathcal{S}_w(\eta))$