Data Assimilation in Reduced Modeling

Ronald DeVore

Collaborators: Peter Binev, Albert Cohen, Wolfgang Dahmen, Guergana Petrova, Przemek Wojtaszczyk

Parametric PDEs with Data

- Parametric PDEs are used to describe many physical processes
- In some settings, we know the governing PDEs but do not know the parameters of the solution we are trying to capture
- Think for example of groundwater modeling where the process is governed by Darcy's equations but the parameters are unknown to us - the unknown earth below us
- So we make some measurements by drilling
- How can we best utilize these measurements to find the state of the process?

Abstraction of the Problem

- Let $\mathcal{M} = \{u(\cdot, y) : \; y \in \mathcal{P}\}$ be the solution manifold of an parametric PDE
- Example (Elliptic PDE) $-\mathrm{div}(a\nabla u) = f$ + BC,
  $a = a(x,y) = \bar{a}(x) + \sum_{j=1}^{d} y_j \psi_j(x), \; y = (y_j) \in \mathcal{P} := [-1,1]^d$
- $\mathcal{M}$ compact in the energy Hilbert space $\mathcal{H}$,
- We wish to approximate an element $u \in \mathcal{M}$
- All we know about u are the observations $\ell_i(u)$, $i = 1, \dots, m$, where the ℓ_i are linear functionals on $\mathcal{H}$
- What is the best algorithm for approximating u from this information and what is the best error we can achieve?
- Problems of this type are known as optimal recovery

The Manifold

- We know that u lies on the solution manifold $\mathcal{M}$
- However, $\mathcal{M}$ is complex and acquiring information about it is (numerically) costly
- Reduced modeling (reduced bases and polynomial chaos) gives information that $\mathcal{M}$ is close to certain known nested linear spaces up to known tolerances ε_k:
  - $V_0 \subset V_1 \subset \cdots \subset V_n$ with $\dim(V_k) = k$
  - $\operatorname{dist}(u, V_k)_{\mathcal{H}} \le \varepsilon_k, \quad k = 0, 1, \ldots, n$
  - That is, for each $y \in \mathcal{P}$, $\operatorname{dist}(u(y), V_k)_{\mathcal{H}} \le \varepsilon_k$, $k = 0, 1, \ldots, n$
- So, a better model, is to replace the assumption $u \in \mathcal{M}$ by these weaker but better understood assumptions on approximability

Abstraction of Problem

- $\mathcal{H}$ a Hilbert space, $\ell_i(u) = \langle u, \omega_i \rangle$, $i = 1, \dots, m$, $W = \mathrm{span}\{\omega_1, \dots, \omega_m\}$, $W^\perp$ the null-space
- Data: measurements determine $P_W u = w \in W$
- Model class: $\mathcal{K}$ determined by
  - $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n, \quad \dim(V_k) = k$
  - $\varepsilon_0 \geq \varepsilon_1 \geq \cdots \geq \varepsilon_n$
- Two settings: one space or multi-space
  - $\mathcal{K} = \mathcal{K}^{\mathrm{one}} := \{u \in \mathcal{H} : \ \mathrm{dist}(u, V_n) \leq \varepsilon_n, \}$
  - $\mathcal{K} = \mathcal{K}^{\mathrm{mult}} := \{u \in \mathcal{H} : \ \mathrm{dist}(u, V_k) \leq \varepsilon_k, \ k = 0, 1, \dots, n\}$
- The knowledge we have about the target function is that it is in $\mathcal{K}_w := \{u \in \mathcal{K} : \ P_W u = w\}$
- Notice that $\mathcal{K}, \mathcal{K}_w$ are no longer necessarily compact but they are bounded sets

Algorithms

- In practice we will be given a measurement $w = P_W u$ and want to construct an approximation $\hat{u}(w)$ to u
- Algorithm: An algorithm is a mapping $A : W \to \mathcal{H}$
- Given $w = P_W u$, the element $\hat{u}(w) := A(P_W u)$ is the approximation to u
- Notice that all elements of $\mathcal{K}_w$ are assigned the same approximation $A(w)$
- Individual error: $\|u - A(P_W u)\|$
- Performance of an algorithm A on the class $\mathcal{K}_w$:

$$E_A(\mathcal{K}_w) := \sup_{u \in \mathcal{K}_w} \|u - A(P_W u)\|$$

- Performance of an algorithm A on the class $\mathcal{K}$:

$$E_A(\mathcal{K}) := \sup_{w \in W} \sup_{u \in \mathcal{K}_w} \|u - A(P_W u)\|$$

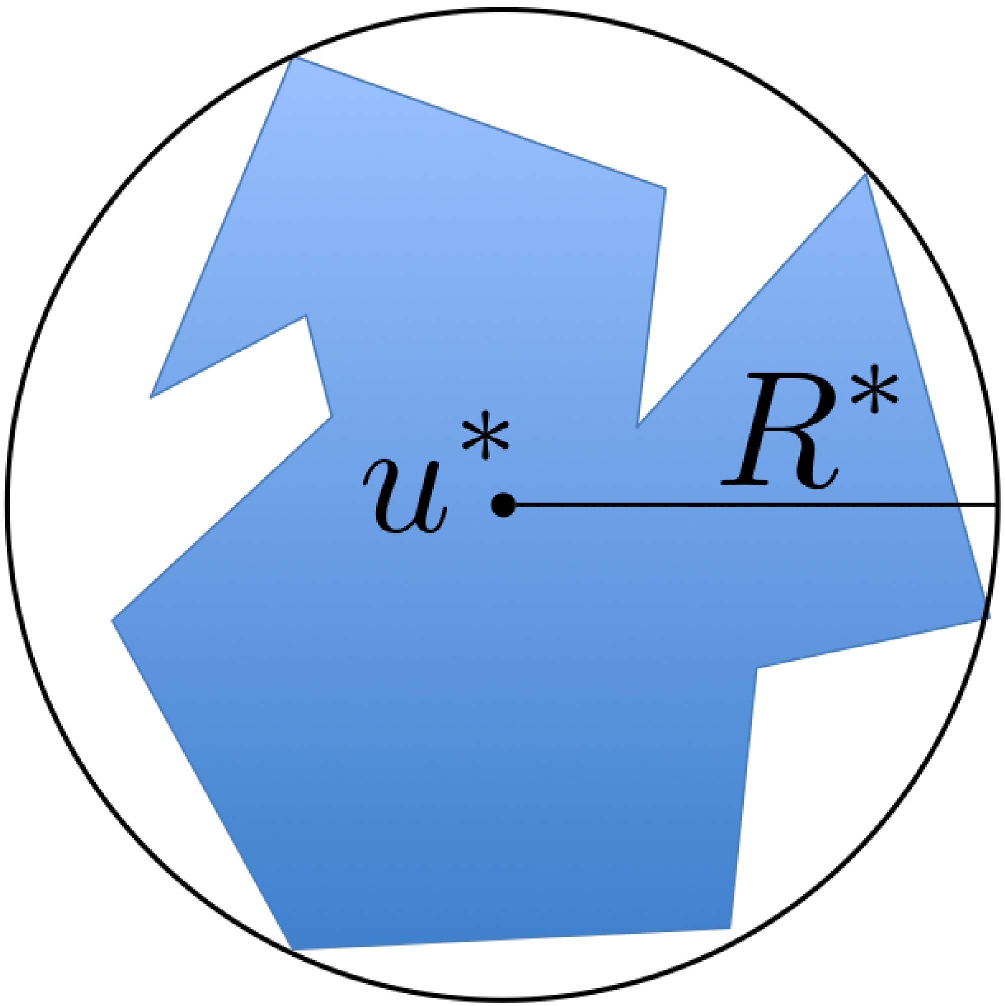
Best Algorithm

- Best Algorithm A^* for a class $\mathcal{K}$

$$A^* := \text{Argmin}_A\, E_A(\mathcal{K})$$

- The best algorithm for recovering any bounded set S in a Banach space is easy to describe
  - The Chebyshev ball of S is the smallest ball $B(u^*, R^*)$ that contains S
  - The best recovery is to choose the center u^* as the approximation and the error on the class S is R^*
  - u^* is called the Chebyshev center of S and R^* is the Chebyshev radius
  - Of course the problem is that it is generally non-trivial to find the Chebyshev ball of a given set S

Chebyshev ball of a set S



Best Algorithm in Our Case

- The best algorithm in our case is the mapping A^* which takes each $w \in W$ to the Chebyshev center $u^*(w)$ of $\mathcal{K}_w$

  $A^* : \; w \to A^*(w) = u^*(w)$
- This is true for both one-space or multi-space
- The performance of this best algorithm is then the Chebyshev radius

$$E^*(\mathcal{K}_w) = E_{A^*}(\mathcal{K}_w) = R^*(w), \quad w \in W$$

$$E^*(\mathcal{K}) = \sup_{w \in W} E_{A^*}(\mathcal{K}_w) = \sup_{w \in W} R^*(w)$$

One Space Case

- This case was initially studied by Maday-Patera-Penn-Yano
- Their algorithm A can be described as follows:
  - Read w and consider $\mathcal{H}_w := \{u \in \mathcal{H} : \; P_W u = w\}$
  - Determine $\bar{u}(w) \in \mathcal{H}_w$, $\bar{v}(w) \in V_n$ closest:

    $$\|\bar{u}(w) - \bar{v}(w)\| = \mathrm{dist}(\mathcal{H}_w, V_n)$$
  - Define $A(w) := \bar{u}(w)$
- MPPY provide some initial analysis of their algorithm
- The bounds in this analysis involve an inf-sup condition for the spaces V_n and W

The Inf Sup Condition

- For two closed subspaces $V, W \subset \mathcal{H}$, we define

$$\beta(V,W) := \inf_{v\in V} \sup_{w\in W} \frac{\langle v, w\rangle}{\|v\|\|w\|} = \inf_{v\in V} \frac{\|P_W v\|}{\|v\|}$$

- A useful relation is $\beta(V,W) = \beta(W^\perp, V^\perp)$ which gives

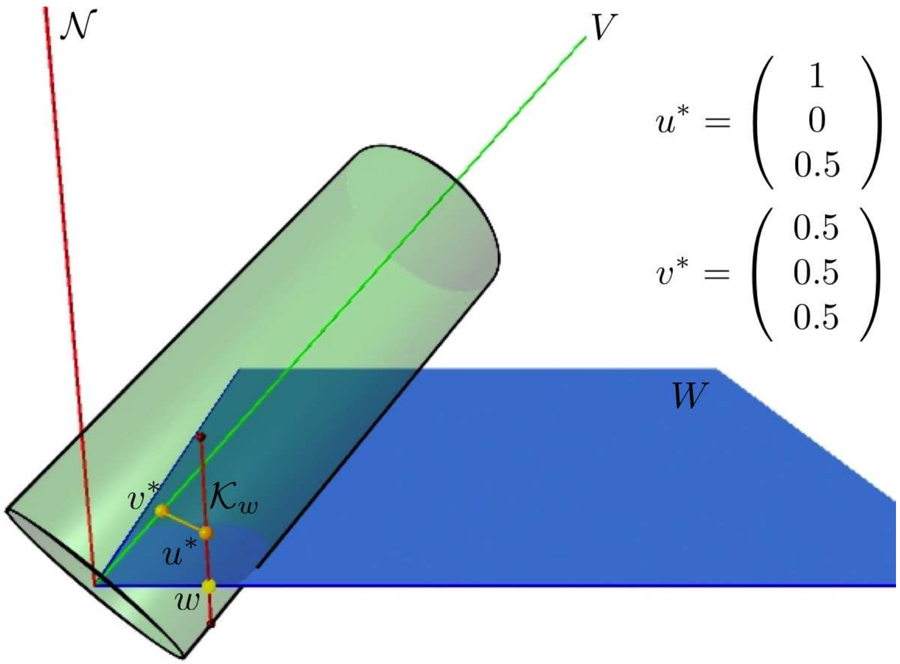
$$\mu(V,W) := \sup_{\eta\in W^\perp} \frac{\|\eta\|}{\|P_{V^\perp}\eta\|} = \beta^{-1}(V,W)$$

- Notice that $\mu(V,W) = \infty$ if $\dim(V) > \dim(W)$

Optimality in the One Space Case

- We prove:
  - The Chebyshev ball for $\mathcal{K}_w$ has center $u^*(w) = \bar{u}(w)$ where $\bar{u}(w)$ is given by the MPPY Algorithm
  - The Chebyshev radius for $\mathcal{K}_w$ is $R^*(w) := \mu(V_n, W)[\varepsilon_n^2 - \|u^*(w) - v^*(w)\|^2]^{1/2}$
- So $A^* : w \to u^*(w)$ is the best one-space algorithm
- $E(\mathcal{K}) = E_{A^*}(\mathcal{K}) = \mu(V_n, W)\varepsilon_n$
- Further Remarks:
  - MPPY are geniuses
  - Finding $u^*(w)$ does not require the knowledge of ε_n and is done by least squares
  - $\mathcal{K}_w$ is an ellipsoid (the intersection of the cylinder $\mathcal{K}$ with the affine space $\mathcal{H}_w := \{u : \ P_W u = w\}$)

One space geometry



$$u^* = \begin{pmatrix} 1 \\ 0 \\ 0.5 \end{pmatrix}$$

$$v^* = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

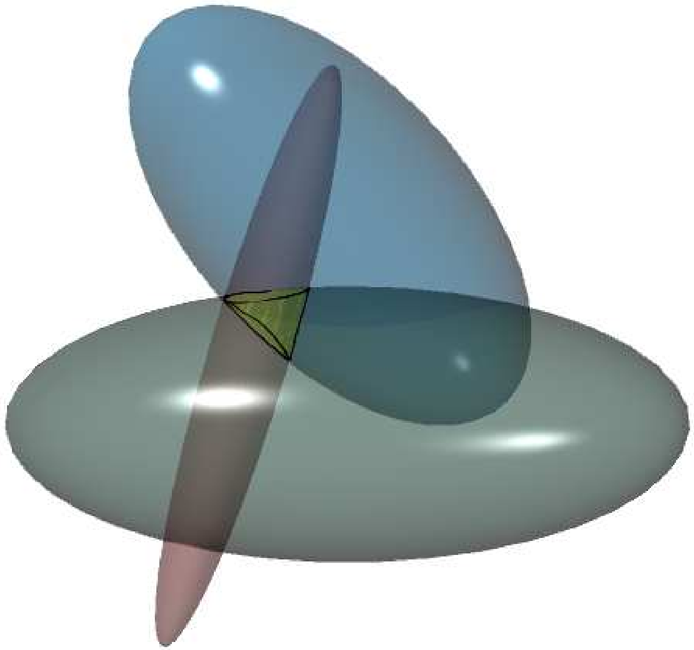
Initial Thoughts on Multi-Space Case

- $$\mathcal{K}_w := \mathcal{K}^{\text{mult}} = \bigcap_{0 \le k \le n} \mathcal{K}_w(V_k)$$
- Each one-space set $\mathcal{K}_w(V_k)$, $k = 0, \dots, n$ is an ellipsoid
- Finding Chebyshev ball for a general intersection of ellipsoids is known to be NP hard
- Poor Man's Algorithm: Choose best one-space: Gives an algorithm A with performance

$$E_A(\mathcal{K}_w) = \min_{k=0,\dots,n} \mu(V_k, W)\varepsilon_k$$

- Notice that the $\mu(V_k, W)$ increase with k but the ε_k decrease with k
- Simple examples in 2D or 3D where the Chebyshev radius of $\mathcal{K}_w$ is zero but Poor Man's Estimate is one

$\mathcal{K}_w$ - Intersection of ellipsoids



Our Results in the Multi-Space Case

- Numerical:
  - We give a numerical algorithm A for finding a point in the intersection
  - This algorithm is near optimal

$$E_A(\mathcal{K}_w) \leq 2E_{A^*}(\mathcal{K}_w)$$

- A Priori Bounds:
  - We provide a priori bounds for the Chebyshev radius in the multi-space case
  - These estimates show that the true Chebysev radius of the intersection may be arbitrarily smaller than the estimate in the Poor Man's Algorithm
  - Our a priori bounds identify a class of examples where the multi-space setting benefits

Numerical Multi-space Algorithms

- Finding a point in the intersection of a finite number of convex sets is a standard problem (known as convex feasibility, see Patrick Combettes' book
- We can use standard methods for this problem
- I will discuss only the method of Alternating Projections
- Two important observations
  - O1: If $\mathcal{K}_w$ is non-empty then it contains a point in the finite dimensional space $\tilde{\mathcal{H}} := V_n + W$
  - So we restrict ourselves to this finite dimensional problem
  - O2: For any point $u \in \tilde{\mathcal{H}}$, its projection $P_C u$ onto a convex set C is closer to each point $x \in C$ than u: $\|P_C u - x\| \leq \|u - x\|$

Numerical Multi-space Algorithms

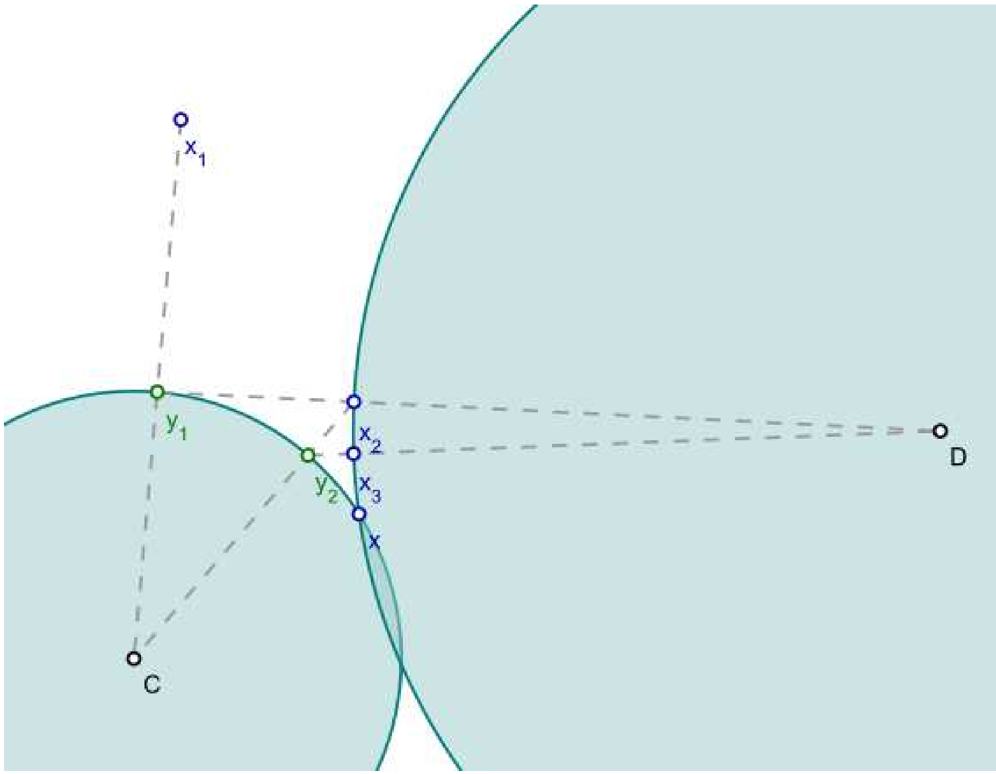
- Let $\mathcal{K}_j := \{u \in \mathcal{H} : \mathrm{dist}(u, V_j) \leq \varepsilon_j\}$
- $\mathcal{H}_w := \{u \in \mathcal{H} : \ P_w u = w\}$
- Each of the projections $P_{\mathcal{K}_j}, P_{H_w}$ is easy to implement
- Alternating Projections leads to the sequence

$$u^{k+1} := P_{\mathcal{K}_n} P_{\mathcal{K}_{n-1}} \cdots P_{\mathcal{K}_1} P_{\mathcal{K}_0} P_{\mathcal{H}_w} u^k.$$

- From general results on alternate projections onto convex sets, this sequence converges towards a point $u^* \in \mathcal{K}_w$ when $\mathcal{K}_w$ is not empty
- Notice that because of O2

$$\mathrm{dist}(u^{k+1}, \mathcal{K}_w) \leq \mathrm{dist}(u^k, \mathcal{K}_w)$$

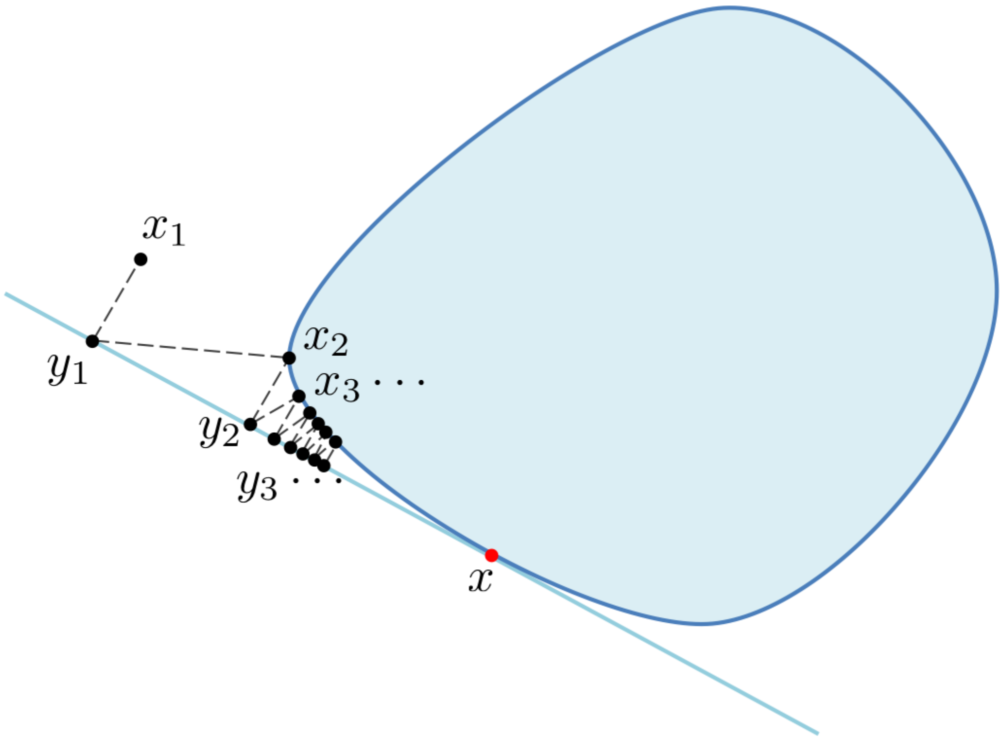
Alternating Projections



Performance Bounds

- In general convergence rates for convex feasibility can be arbitrarily slow

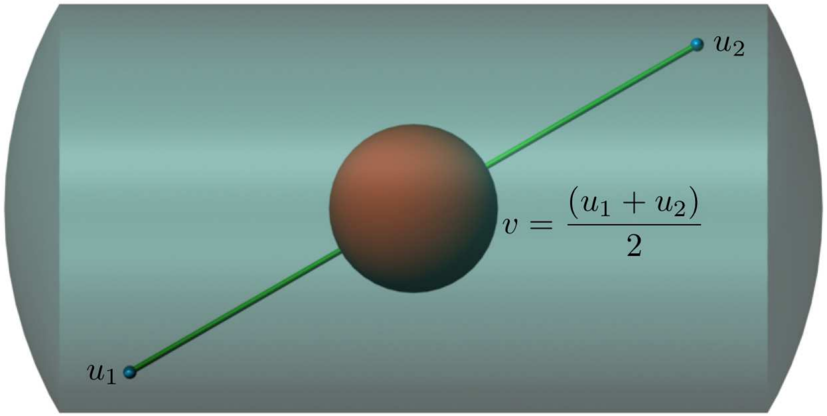
Possible Slow Convergence



Performance Bounds

- In general convergence rates for convex feasibility can be arbitrarily slow
- One can obtain convergence rates for Alternating Projections if the convex bodies are uniformly convex
- We do not have uniform convexity
- However, we have a restricted form:
  - Let $u_1, u_2 \in \mathcal{K}$ with $u_1 - u_2 \in W^\perp$
  - The ball $B := B(u_0, r)$ centered at $u_0 := \frac{1}{2}(u_1 + u_2)$ of radius $r := c_0 \min\left\{\delta, \delta^2\right\}$, $\delta := \|u_1 - u_2\|$, is completely contained in $\mathcal{K}$.
  - Here c_0 is an absolute constant depending on the ε_k, $\mu(V_k, W)$, $k = 0, \ldots, n$

Restricted Uniform Convexity



Two Applications

- We can exploit this restricted uniform convexity to prove the following two results:
  - u^k converges to a point in the intersection with rate (at worst) $O(k^{-1/2})$
  - For any $v \in \mathcal{K}$,

$$\mathrm{dist}(v, \mathcal{K}_w) \leq C[\mathrm{dist}(v, \mathcal{H}_w)]^{1/2}$$

  - Here C is an computable constant depending on the $\varepsilon_k, \mu(V_k, W)$, $k = 0, \ldots, n$, which we can compute
- We can use this second result for a Stopping Criteria in our algorithm by applying it to u^k since we know u^k is in $\mathcal{K}$

Favorable Bases

- The proofs of the above results rely on the geometry between W and the spaces V_k
- To see this geometry more clearly it is useful to choose special orthogonal bases for W and V_n
- We call these Favorable Bases

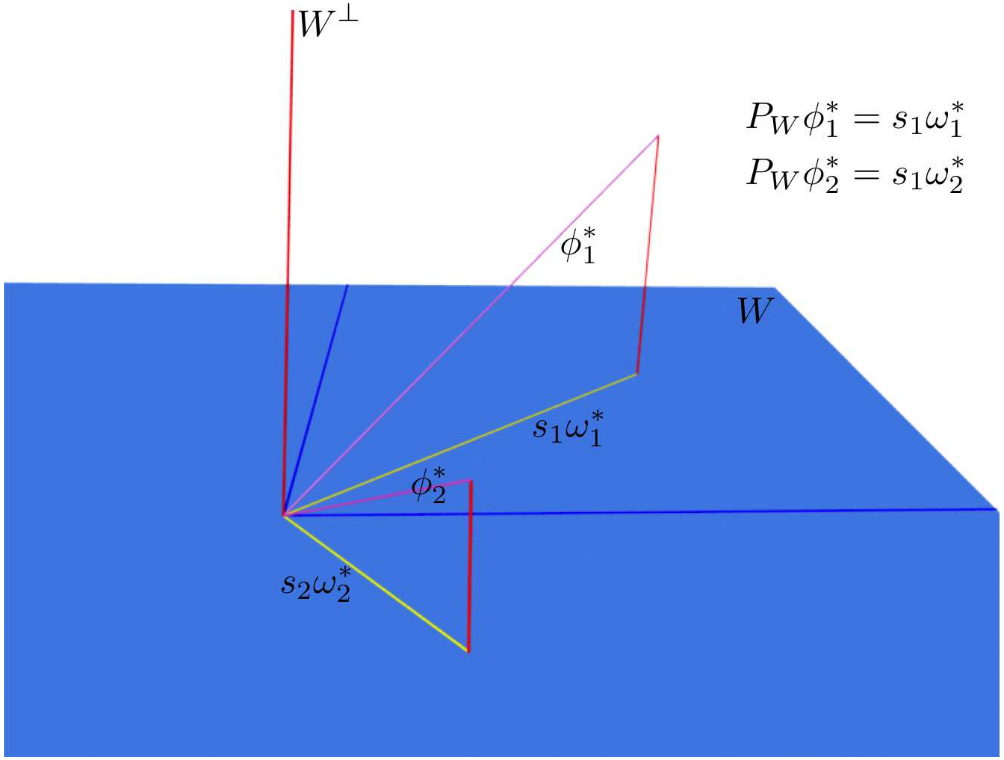
First basis for the V_k

- There is a natural basis for capturing the approximation properties of the V_k, $k = 0, \dots, n$ and the body $\mathcal{K}$
- Choose $\phi_1, \dots, \phi_n$ as an orthonormal system such that
- $V_k = \mathrm{span}\{\phi_1, \dots, \phi_k\}$
- This is a good basis for representing the body $\mathcal{K}$
  - $u \in \mathcal{K}$ if and only if $u = \sum_{j=1}^{n} \alpha_j \phi_j + e$
  - $e \in V_n^{\perp}$
  - $\sum_{j>k} \alpha_j^2 + \|e\|^2 \le \varepsilon_k, \quad k = 0, \dots, n$
- However, this basis is not good for seeing how W is oriented to V_n

Favorable bases for orientation

- Fix W, V_n with $n \leq m$
- Start with any basis $\omega_1, \ldots, \omega_m$ of W and do a SVD on the cross Grammian $G := (\langle \omega_i, \phi_j \rangle) = USV^t$
- This gives new orthonomal bases orthonormal bases $\{\phi_1^*, \ldots, \phi_n^*\}$ for V_n and $\{\omega_1^*, \ldots, \omega_m^*\}$ for W with the following properties:
  - They have diagonal cross Grammian: $\langle \omega_i^*, \phi_j^* \rangle = s_j \delta_{i,j}, \quad i = 1, \ldots, m \; ; \; j = 1, \ldots, n$
  - Perfect Alignment: $P_W(\phi_j^*) = s_j \omega_j^*, \; j = 1, \ldots, n,$
  - $\mu(W, V_n) = \max_{0 \leq j \leq n} s_j^{-1}$

Example of Favorable Bases



A Priori Estimates

- It is desireable to have easily computable a priori bounds for the radius of $\mathcal{K}^{\mathrm{mult}}(w)$
- Such bounds would tell us whether it is useful to implement the convex minimization and guide us to acceptable error
- A priori bounds can be derived through Favorable Bases
- Let $\Lambda = (\lambda_{i,j})$ be the change of basis matrix from (ϕ_j) to (ϕ_j^*)
- Introduce $\theta_i := \sum_{j=1}^{n} |\lambda_{i,j}|\varepsilon_{j-1}$
- These numbers are computable

The Bound

- Define k as the largest integer for which $\sum_{j=k}^{n} s_j^2 \theta_j^2 \geq \varepsilon_n^2$
- Define $E_n := \varepsilon_n^2 + \sum_{j=k}^{n} \theta_j^2$
- A priori Bound: $E(\mathcal{K}_w) \leq 2E_n$,
- Most favorable case: If the rows of the matrix Λ are a permutation of the identity matrix
  - In this case, the Poor Man's Bound can be one while E_n is arbitrarily small
  - More generally if the bases are coherent the multi-space algorithm has an a priori performance guarantee superior to the Poor Man's Algorithm