Projection methods

- Introduction to projection-type techniques
- Sample one-dimensional Projection methods
- Some theory and interpretation –
- See Chapter 5 of text for details.

Projection Methods

➤ The main idea of projection methods is to extract an approximate solution from a subspace.

➤ We define a subspace of approximants of dimension m and a set of m conditions to extract the solution

➤ These conditions are typically expressed by orthogonality constraints.

➤ This defines one basic step which is repeated until convergence (alternatively the dimension of the subspace is increased until convergence).

Example: Gauss-Seidel can be viewed as a sequence of projection steps.

Background on projectors

➤ A projector is a linear operator that is idempotent: $P^2 = P$

A few properties:

- P is a projector iff $I - P$ is a projector
- $x \in \mathrm{Ran}(P)$ iff $x = Px$ iff $x \in \mathrm{Null}(P)$
- This means that : $\mathrm{Ran}(P) = \mathrm{Null}(I - P)$.
- Any $x \in \mathbb{R}^n$ can be written (uniquely) as $x = x_1 + x_2$, $x_1 = Px \in \mathrm{Ran}(P)$ $x_2 = (I - P)x \in \mathrm{Null}(P)$ - So:

$$\mathbb{R}^n = \mathrm{Ran}(P) \oplus \mathrm{Null}(P)$$

✍ Prove the above properties

Background on projectors (Continued)

➤ The decomposition $\mathbb{R}^n = K \oplus S$ defines a (unique) projector P:

- From $x = x_1 + x_2$, set $Px = x_1$.
- For this P: $\mathrm{Ran}(P) = K$ and $\mathrm{Null}(P) = S$.
- Note: $dim(K) = m$, $dim(S) = n - m$.

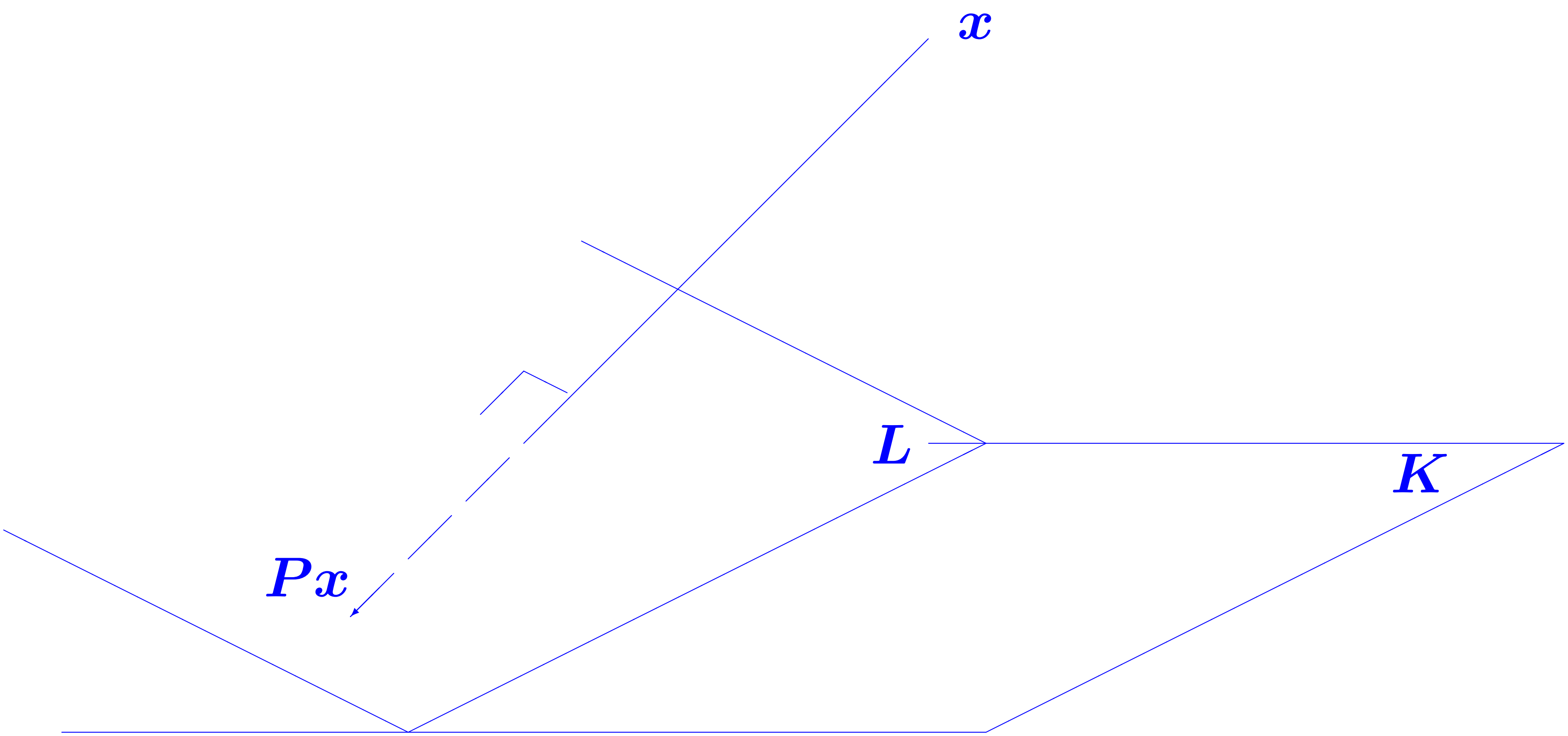
➤ Pb: express mapping $x \rightarrow u = Px$ in terms of K, S

➤ Note $u \in K$, $x - u \in S$

➤ Express 2nd part with m constraints: let $L = S^{\perp}$, then

$$u = Px \text{ iff } \begin{cases} u \in K \\ x - u \perp L \end{cases}$$

➤ Projection onto K and orthogonally to L



- Illustration: P projects onto K and orthogonally to L
- When $L = K$ projector is orthogonal.
- Note: $Px = 0$ iff $x \perp L$.

Projection methods

- Initial Problem: $\boldsymbol{b - Ax = 0}$

Given two subspaces $\boldsymbol{K}$ and $\boldsymbol{L}$ of $\mathbb{R}^{\boldsymbol{N}}$ define the approximate problem:

$$\text{Find } \tilde{\boldsymbol{x}} \in \boldsymbol{K} \text{ such that } \boldsymbol{b} - \boldsymbol{A}\tilde{\boldsymbol{x}} \perp \boldsymbol{L}$$

- Petrov-Galerkin condition
- $\boldsymbol{m}$ degrees of freedom $(\boldsymbol{K})$ + $\boldsymbol{m}$ constraints $(\boldsymbol{L}) \rightarrow$
- a small linear system ('projected problem')
- This is a basic projection step. Typically a sequence of such steps are applied

➤ With a nonzero initial guess x_0, approximate problem is

$$\text{Find} \quad \tilde{x} \in x_0 + K \quad \text{such that} \quad b - A\tilde{x} \perp L$$

Write $\tilde{x} = x_0 + \delta$ and $r_0 = b - Ax_0$. $\rightarrow$ system for δ:

$$\text{Find } \delta \ \in K \text{ such that } r_0 - A\delta \perp L$$

✍ Formulate Gauss-Seidel as a projection method -

✍ Generalize Gauss-Seidel by defining subspaces consisting of 'blocks' of coordinates $\mathbf{span}\{e_i, e_{i+1}, ..., e_{i+p}\}$

Matrix representation:

Let

- $V = [v_1, \ldots, v_m]$ a basis of K &
- $W = [w_1, \ldots, w_m]$ a basis of L

➤ Write approximate solution as $\tilde{x} = x_0 + \delta \equiv x_0 + Vy$ where $y \in \mathbb{R}^m$. Then Petrov-Galerkin condition yields:

$$W^T(r_0 - AVy) = 0$$

➤ Therefore,

$$\tilde{x} = x_0 + V[W^TAV]^{-1}W^Tr_0$$

Remark: In practice W^TAV is known from algorithm and has a simple structure [tridiagonal, Hessenberg,..]

Prototype Projection Method

Until Convergence Do:

1. Select a pair of subspaces K, and L;

2. Choose bases: $V = [v_1, \ldots, v_m]$ for K and $W = [w_1, \ldots, w_m]$ for L.

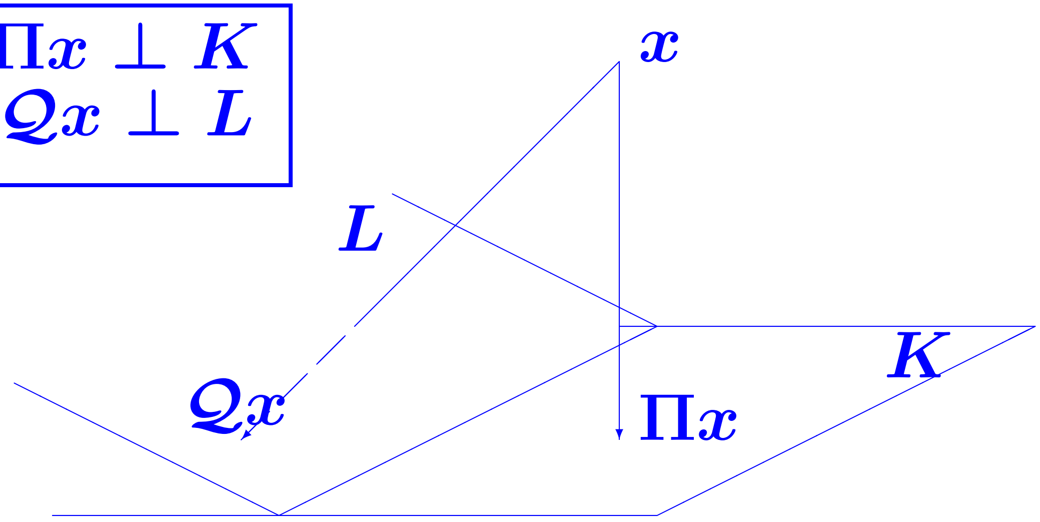
3. Compute :

$$r \leftarrow b - Ax,$$
$$y \leftarrow (W^T AV)^{-1} W^T r,$$
$$x \leftarrow x + Vy.$$

Projection methods: Operator form representation

➤ Let $\mathbf{\Pi}$ = the orthogonal projector onto $\boldsymbol{K}$ and $\boldsymbol{\mathcal{Q}}$ the (oblique) projector onto $\boldsymbol{K}$ and orthogonally to $\boldsymbol{L}$.

$$\boxed{\begin{array}{rcl} \mathbf{\Pi}\boldsymbol{x} & \in & \boldsymbol{K}, \quad \boldsymbol{x} - \mathbf{\Pi}\boldsymbol{x} \perp \boldsymbol{K} \\ \boldsymbol{\mathcal{Q}}\boldsymbol{x} & \in & \boldsymbol{K}, \quad \boldsymbol{x} - \boldsymbol{\mathcal{Q}}\boldsymbol{x} \perp \boldsymbol{L} \end{array}}$$



$\mathbf{\Pi}$ and $\boldsymbol{\mathcal{Q}}$ projectors

Assumption: no vector of $\boldsymbol{K}$ is $\perp$ to $\boldsymbol{L}$

In the case $x_0 = 0$, approximate problem amounts to solving

$$\mathcal{Q}(b - Ax) = 0, \quad x \quad \in K$$

or in operator form (solution is Πx)

$$\mathcal{Q}(b - A\Pi x) = 0$$

Question: what accuracy can one expect?

- Let $\boldsymbol{x}^*$ be the exact solution. Then

1) We cannot get better accuracy than $\|(\boldsymbol{I}-\boldsymbol{\Pi})\boldsymbol{x}^*\|_2$, i.e.,

$$\boxed{\|\tilde{\boldsymbol{x}}-\boldsymbol{x}^*\|_2 \geq \|(\boldsymbol{I}-\boldsymbol{\Pi})\boldsymbol{x}^*\|_2}$$

2) The residual of the exact solution for the approximate problem satisfies:

$$\boxed{\|\boldsymbol{b}-\boldsymbol{\mathcal{Q}}\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{x}^*\|_2 \leq \|\boldsymbol{\mathcal{Q}}\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{\Pi})\|_2\|(\boldsymbol{I}-\boldsymbol{\Pi})\boldsymbol{x}^*\|_2}$$

Two Important Particular Cases.

$\boxed{\mathbf{1.}\quad L = K}$

- When A is SPD then $\|x^* - \tilde{x}\|_A = \min_{z \in K} \|x^* - z\|_A$.
- Class of Galerkin or Orthogonal projection methods
- Important member of this class: Conjugate Gradient (CG) method

$\boxed{\mathbf{2.}\quad L = AK}$.

In this case $\|b - A\tilde{x}\|_2 = \min_{z \in K} \|b - Az\|_2$

- Class of Minimal Residual Methods: CR, GCR, ORTHOMIN, GMRES, CGNR, ...

One-dimensional projection processes

$$\boldsymbol{K = span\{d\}}$$
and
$$\boldsymbol{L = span\{e\}}$$

Then $\tilde{x} = x + \alpha d$. Condition $r - A\delta \perp e$ yields

$$\boxed{\alpha = \frac{(r,e)}{(Ad,e)}}$$

➤ Three popular choices:

(1) Steepest descent

(2) Minimal residual iteration

(3) Residual norm steepest descent

1. Steepest descent.

A is SPD. Take at each step $d = r$ and $e = r$.

$$\text{Iteration:} \quad \boxed{\begin{array}{l} r \leftarrow b - Ax, \\ \alpha \leftarrow (r, r)/(Ar, r) \\ x \leftarrow x + \alpha r \end{array}}$$

➤ Each step minimizes $f(x) = \|x - x^*\|_A^2 = (A(x - x^*), (x - x^*))$ in direction $-\nabla f$.

➤ Convergence guaranteed if A is SPD.

✍ As is formulated, the above algorithm requires 2 'matvecs' per step. Reformulate it so only one is needed.

Convergence based on the Kantorovitch inequality: Let B be an SPD matrix, λ_{max}, λ_{min} its largest and smallest eigenvalues. Then,

$$\frac{(Bx, x)(B^{-1}x, x)}{(x, x)^2} \leq \frac{(\lambda_{max} + \lambda_{min})^2}{4\,\lambda_{max}\lambda_{min}}, \quad \forall x \neq 0.$$

- This helps establish the convergence result

Let A an SPD matrix. Then, the A-norms of the error vectors $d_k = x_* - x_k$ generated by steepest descent satisfy:

$$\|d_{k+1}\|_A \leq \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \|d_k\|_A$$

- Algorithm converges for any initial guess x_0.

Proof: Observe $\|d_{k+1}\|_A^2 = (Ad_{k+1}, d_{k+1}) = (r_{k+1}, d_{k+1})$

- by substitution,

$$\|d_{k+1}\|_A^2 = (r_{k+1}, d_k - \alpha_k r_k)$$

- By construction $r_{k+1} \perp r_k$ so we get $\|d_{k+1}\|_A^2 = (r_{k+1}, d_k)$. Now:

$$\begin{aligned} \|d_{k+1}\|_A^2 &= (r_k - \alpha_k A r_k, d_k) \\ &= (r_k, A^{-1} r_k) - \alpha_k (r_k, r_k) \\ &= \|d_k\|_A^2 \left(1 - \frac{(r_k, r_k)}{(r_k, A r_k)} \times \frac{(r_k, r_k)}{(r_k, A^{-1} r_k)}\right). \end{aligned}$$

Result follows by applying the Kantorovich inequality. ■

2. Minimal residual iteration.

A positive definite ($A + A^T$ is SPD). Take at each step $d = r$ and $e = Ar$.

Iteration: $\boxed{\begin{array}{l} r \leftarrow b - Ax, \\ \alpha \leftarrow (Ar, r)/(Ar, Ar) \\ x \leftarrow x + \alpha r \end{array}}$

- Each step minimizes $f(x) = \|b - Ax\|_2^2$ in direction r.
- Converges under the condition that $A + A^T$ is SPD.

✍ As is formulated, the above algorithm would require 2 'matvecs' at each step. Reformulate it so that only one matvec is required

Convergence

Let A be a real positive definite matrix, and let

$$\mu = \lambda_{min}(A + A^T)/2, \quad \sigma = \|A\|_2.$$

Then the residual vectors generated by the Min. Res. Algorithm satisfy:

$$\|r_{k+1}\|_2 \leq \left(1 - \frac{\mu^2}{\sigma^2}\right)^{1/2} \|r_k\|_2$$

- In this case Min. Res. converges for any initial guess x_0.

Proof: Similar to steepest descent. Start with

$$\begin{aligned}\|r_{k+1}\|_2^2 &= (r_k - \alpha_k A r_k, r_k - \alpha_k A r_k) \\ &= (r_k - \alpha_k A r_k, r_k) - \alpha_k (r_k - \alpha_k A r_k, A r_k).\end{aligned}$$

By construction, $r_{k+1} = r_k - \alpha_k A r_k$ is $\perp$ Ar_k. ➤ $\|r_{k+1}\|_2^2 = (r_k - \alpha_k A r_k, r_k)$. Then:

$$\begin{aligned}\|r_{k+1}\|_2^2 &= (r_k - \alpha_k A r_k, r_k) \\ &= (r_k, r_k) - \alpha_k (A r_k, r_k) \\ &= \|r_k\|_2^2 \left(1 - \frac{(A r_k, r_k)}{(r_k, r_k)} \frac{(A r_k, r_k)}{(A r_k, A r_k)}\right) \\ &= \|r_k\|_2^2 \left(1 - \frac{(A r_k, r_k)^2}{(r_k, r_k)^2} \frac{\|r_k\|_2^2}{\|A r_k\|_2^2}\right).\end{aligned}$$

Result follows from the inequalities $(Ax, x)/(x, x) \geq \mu > 0$ and $\|A r_k\|_2 \leq \|A\|_2 \; \|r_k\|_2$. ■

3. Residual norm steepest descent.

A is arbitrary (nonsingular). Take at each step $d = A^T r$ and $e = Ad$.

Iteration: $\boxed{\begin{array}{l} r \leftarrow b - Ax, d = A^T r \\ \alpha \leftarrow \|d\|_2^2 / \|Ad\|_2^2 \\ x \leftarrow x + \alpha d \end{array}}$

- Each step minimizes $f(x) = \|b - Ax\|_2^2$ in direction $-\nabla f$.
- Important Note: equivalent to usual steepest descent applied to normal equations $A^T Ax = A^T b$.
- Converges under the condition that A is nonsingular.