High Performance Linear Algebra

Lecture 2.2: Projection methods

Ph.D. program in High Performance Scientific Computing

Salvatore Filippone   Pasqua D'Ambra   Fabio Durastante

March 12 2026 — 13.00:15.00

Table of Contents

Iterative solvers: Steepest descent

Given

$$Ax = b$$

Given

$$Ax = b$$

let us transform it into a minimization problem

$$\min_x \phi(x) = \frac{1}{2}x^T A x - x^T b, \tag{1}$$

with A symmetric positive definite.

The minimization problem is equivalent:

$$\min_x \phi(x) \Rightarrow \nabla \phi(x_c) = 0$$

The minimization problem is equivalent:

$$\min_x \phi(x) \Rightarrow \nabla\phi(x_c) = 0$$

$$\Rightarrow Ax - b = 0$$

The minimization problem is equivalent:

$$\min_{x} \phi(x) \Rightarrow \nabla\phi(x_c) = 0$$

$$\Rightarrow Ax - b = 0$$

$$\Rightarrow Ax = b$$

The minimization problem is equivalent:

$$\min_x \phi(x) \Rightarrow \nabla\phi(x_c) = 0$$

$$\Rightarrow Ax - b = 0$$

$$\Rightarrow Ax = b$$

To build an iterative method, at each step we have to:

1. Choose a search direction p_k;
2. Move along the search direction with a step α_k such that

$$\phi(x_{k+1}) = \phi(x_k + \alpha_k p_k) < \phi(x_k)$$

First idea: move along the direction where $\phi(x)$ changes most rapidly. Therefore:

$$-\nabla\phi(x) = b - Ax = r,$$

since we want to have a reduction in ϕ. If the residual r is nonzero, there exists α such that

$$\phi(x_i + \alpha r_i) < \phi(x_i).$$

First idea: move along the direction where $\phi(x)$ changes most rapidly. Therefore:

$$-\nabla\phi(x) = b - Ax = r,$$

since we want to have a reduction in ϕ. If the residual r is nonzero, there exists α such that

$$\phi(x_i + \alpha r_i) < \phi(x_i).$$

This is a one-dimensional problem; minimizing explicitly we have

$$\phi(x_i + \alpha r_i) = \phi(x_i) - \alpha r_i^T r_i + \frac{1}{2}\alpha^2 r_i^T A r_i,$$

therefore

$$\alpha = r_i^T r_i / r_i^T A r_i.$$

An easy implication is that:

$$r_k - \alpha A r_k = r_{k+1} \perp r_k.$$

$x_0 =$ initial guess
$r_0 = b - Ax_0$
$k = 0$
while $r_k \neq 0$ **do**
  $k \leftarrow k + 1$
  $\alpha_k \leftarrow r_{k-1}^T r_{k-1} / r_{k-1}^T A r_{k-1}$
  $x_k \leftarrow x_{k-1} + \alpha_k r_{k-1}$
  $r_k \leftarrow b - Ax_k$
end while

Main problem: convergence of steepest descent can be very slow. Why? It depends on:

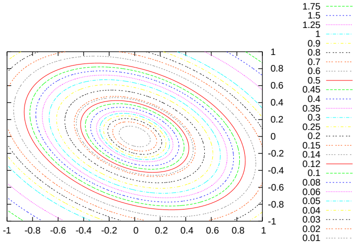
$$\kappa_2(A) = \lambda_1(A)/\lambda_n(A).$$

What does it mean?

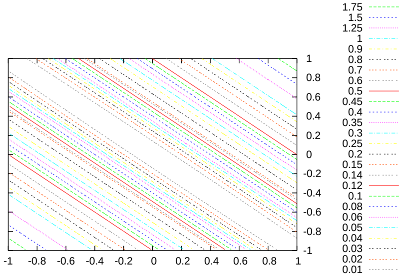
Main problem: convergence of steepest descent can be very slow. Why? It depends on:

$$\kappa_2(A) = \lambda_1(A)/\lambda_n(A).$$

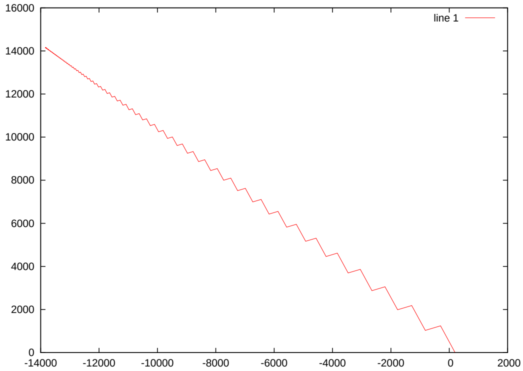
What does it mean? Let's look at a "good" case:



Now, for a “bad” case:



Behaviour over 350 iterations:



The main problem of steepest descent is: successive search directions are mutually orthogonal

$$r_{k+1}^T r_k = 0.$$

The main problem of steepest descent is: successive search directions are mutually orthogonal

$$r_{k+1}^T r_k = 0.$$

Let us then modify our choice of search directions

$$p_1, p_2, \ldots, p_k \neq r_k$$

in such a way that:

1. $p_k^T r_{k-1} \neq 0$;
2. The p_k are linearly independent.

The main problem of steepest descent is: successive search directions are mutually orthogonal

$$r_{k+1}^T r_k = 0.$$

Let us then modify our choice of search directions

$$p_1, p_2, \dots, p_k \neq r_k$$

in such a way that:

1. $p_k^T r_{k-1} \neq 0$;
2. The p_k are linearly independent.

Now, x_k has to solve

$$\min_{x \in x_0 + \mathit{span}\{p_1,\dots,p_k\}} \phi(x)$$

Therefore:

$$x_k = x_0 + P_{k-1}y + \alpha p_k,$$

where

$$P_{k-1} = [p_1, \ldots, p_{k-1}], \qquad y \in \mathbb{R}^{k-1}.$$

Therefore:

$$x_k = x_0 + P_{k-1}y + \alpha p_k,$$

where

$$P_{k-1} = [p_1, \ldots, p_{k-1}], \qquad y \in \mathbb{R}^{k-1}.$$

Substituting, we have

$$\phi(x_k) = \phi(x_0 + P_{k-1}y) + \alpha y^T P_{k-1}^T A p_k + \frac{\alpha^2}{2} p_k^T A p_k - \alpha p_k^T r_0.$$

Therefore:

$$x_k = x_0 + P_{k-1} y + \alpha p_k,$$

where

$$P_{k-1} = [p_1, \ldots, p_{k-1}], \qquad y \in \mathbb{R}^{k-1}.$$

Substituting, we have

$$\phi(x_k) = \phi(x_0 + P_{k-1} y) + \alpha y^T P_{k-1}^T A p_k + \frac{\alpha^2}{2} p_k^T A p_k - \alpha p_k^T r_0.$$

If at all possible, we'd like

$$p_k \in span\{Ap_1, \ldots, Ap_{k-1}\}^\perp,$$

as this would allow a "greedy" minimization strategy.

A characterization is:

$$p_k = r_{k-1} - AP_{k-1} z_{k-1},$$

where

$$z_{k-1} = \min_{z \in \mathbb{R}^{k-1}} \| r_{k-1} - AP_{k-1} z \|_2 .$$

A characterization is:

$$p_k = r_{k-1} - AP_{k-1}z_{k-1},$$

where

$$z_{k-1} = \min_{z \in \mathbb{R}^{k-1}} \|r_{k-1} - AP_{k-1}z\|_2.$$

It follows that

$$p^T AP_{k-1} = 0,$$

i.e. the p_k are A-conjugate: the residual for the best approximation from a subspace (projection) is orthogonal to the subspace.

If

$$p_k \in span\{Ap_1, \dots, Ap_{k-1}\}^{\perp},$$

then we can easily find

$$\alpha = p_k^T r_{k-1} / p_k^T A p_k;$$

moreover

$$p_k^T r_{k-1} = p_k^T (b - A(x_0 + P_{k-1} y_{k-1})) = p_k^T r_0.$$

If

$$p_k \in span\{Ap_1, \ldots, Ap_{k-1}\}^{\perp},$$

then we can easily find

$$\alpha = p_k^T r_{k-1} / p_k^T A p_k;$$

moreover

$$p_k^T r_{k-1} = p_k^T (b - A(x_0 + P_{k-1} y_{k-1})) = p_k^T r_0.$$

How do we find one such p_k in practice?

If

$$p_k \in span\{Ap_1, \ldots, Ap_{k-1}\}^\perp,$$

then we can easily find

$$\alpha = p_k^T r_{k-1} / p_k^T A p_k;$$

moreover

$$p_k^T r_{k-1} = p_k^T (b - A(x_0 + P_{k-1} y_{k-1})) = p_k^T r_0.$$

How do we find one such p_k in practice? Let's try with

$$p_k = r_{k-1} + \beta_k p_{k-1}$$

If

$$p_k \in span\{Ap_1, \dots, Ap_{k-1}\}^\perp,$$

then we can easily find

$$\alpha = p_k^T r_{k-1} / p_k^T A p_k;$$

moreover

$$p_k^T r_{k-1} = p_k^T (b - A(x_0 + P_{k-1} y_{k-1})) = p_k^T r_0.$$

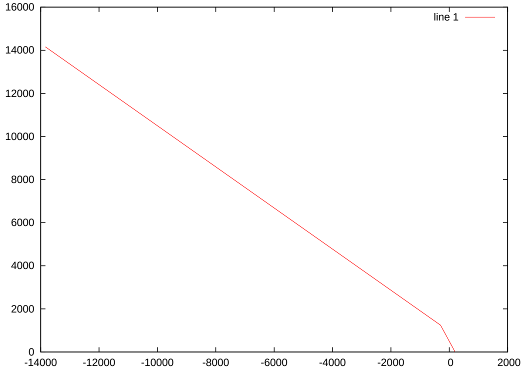
How do we find one such p_k in practice? Let's try with

$$p_k = r_{k-1} + \beta_k p_{k-1}$$

hence, multiplying by $(Ap_{k-1})^T$,

$$\beta_k = -\frac{p_{k-1}^T A r_{k-1}}{p_{k-1}^T A p_{k-1}}$$

Compute $r^{(0)} \leftarrow b - Ax^{(0)}$
while $r_i \neq 0$ **do**
  $i \leftarrow i + 1$
  if $i = 1$ **then**
    $p^{(1)} \leftarrow r^{(0)}$
  else
    $\beta_i \leftarrow -p_{i-1}^T A r_{i-1} / p_{i-1}^T A p_{i-1}$
    $p^{(i)} \leftarrow r^{(i-1)} + \beta_i p^{(i-1)}$
  end if
  $\alpha_i \leftarrow p_i^T r_{i-1} / p_i^T A p_i$
  $x^{(i)} \leftarrow x^{(i-1)} + \alpha_i p^{(i)}$
  $r^{(i)} \leftarrow r^{(i-1)} - \alpha_i A p_i$
end while



Note that since the p_k are A-conjugate and A is SPD, the CG method produces the solution in at most n iterations, which caused it to be mistaken for a direct method.

CG was derived by Hestenes and Stiefels a method to solve the equivalent problem

$$\min_x \phi(x) = \frac{1}{2}x^T Ax - x^T b, \tag{2}$$

That A is SPD implies that minimum exists and is unique.

Lemma

Let $\mathcal{S} \subset \mathbb{R}^N$. Then x_k minimizes ϕ over $\mathcal{S}$, if and only if it also minimizes $\|x - x_*\|_A = \|r\|_{A^{-1}}$ over $\mathcal{S}$.

Proof.

$$\|x - x_*\|_A^2 = x^T A x - x^T A x_* - x_*^T A x + x_*^T A x_*. \tag{3}$$

A SPD and $Ax_* = b$ implies $-x^T A x_* - x_*^T A x = -2x^T b$, hence

$$\|x - x_*\|_A^2 = 2\phi(x) + x_*^T A x_*, \tag{4}$$

$x_*^T A x_*$ is a constant, then the two problems are equivalent (in theory). Moreover

$$\begin{aligned} \|x - x_*\|_A^2 &= (x - x_*)^T A (x - x_*) = (A(x - x_*))^T A^{-1} (A(x - x_*)) \\ &= \|b - Ax\|_{A^{-1}}^2 \end{aligned}$$

□

Theorem

The Conjugate Gradient method produces a sequence of iterates x_k minimizing $\|x - x_*\|_A$ for all $x \in x_0 + \mathcal{K}_k$.

For a proof see [2, 4]. Now, we have

$$x_* - w = x_* - x_0 - \sum_{j=0}^{k-1} \gamma_j A^j r_0 .$$

Since $Ax_* = b$, we have $r_0 = b - Ax_0 = A(x_* - x_0)$ and therefore

$$x_* - w = x_* - x_0 - \sum_{j=0}^{k-1} \gamma_j A^{j+1}(x_* - x_0) = p(A)(x_* - x_0),$$

where the polynomial $p(x) = 1 - \sum_{j=0}^{k-1} \gamma_j z^{j+1}$ is of degree k and satisfies $p(0) = 1$.

Hence we obtain:

$$\|x_* - x\|_A = \min_{p \in \mathcal{P}_k} \|p(A)(x_* - x_0)\|_A \tag{5}$$

i.e. CG minimizes the A-norm of the error over polynomials p such that $p(0) = 1$.
A SPD implies

$$A = U\Lambda U^T \Rightarrow p(A) = Up(\Lambda)U^T$$

but $\Lambda > 0$, hence

$$A^{1/2} = U\Lambda^{1/2}U^T,$$

and

$$\|x\|_A^2 = x^T Ax = \|A^{1/2}x\|_2^2.$$

For any $x \in \mathbb{R}^n$ we then have

$$\|p(A)x\|_A = \|A^{1/2}p(A)x\|_2 \leq \|p(A)\|_2\|A^{1/2}x\|_2 = \|p(A)\|_2\|x\|_A;$$

Therefore:

$$\|x_* - x\|_A \leq \|(x_* - x_0)\|_A \min_{p\in\mathcal{P}_k} \max_{z\in\sigma(A)} |p(z)|$$

Now $\sigma(A) = [\lambda_1, \lambda_N]$; the Chebichev polynomials guarantee $|T_k| \leq 1$ on $[\lambda_1, \lambda_N]$; normalizing to $p(0) = 1$ we have

$$\tilde{P}_k(\lambda) = T_k(\frac{\lambda_N + \lambda_1 - 2\lambda}{\lambda_N - \lambda_1})\left(T_k(\frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1})\right)^{-1};$$

therefore

$$\|x_* - x_k\|_A \leq \|(x_* - x_0)\|_A \cdot (T_k(\frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1}))^{-1},$$

But we also have

$$T_k\left(\frac{\alpha+1}{\alpha-1}\right)=\frac{1}{2}\left(\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^k+\left(\frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}\right)^k\right)>\frac{1}{2}\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^k$$

hence we have the bound

$$||e_j||_A \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^j \cdot \|e_0\|_A$$

where $\kappa(A)=\lambda_N/\lambda_1$ is the 2-norm condition number of the matrix A; the number of iterations to achieve a tolerance of ϵ is bounded as

$$\eta(\epsilon) \leq \frac{1}{2}\sqrt{\kappa(A)}\log(2/\epsilon)+1.$$

Refinement of convergence bounds: suppose m and b such that

$$\lambda_{N-m} \leq b \leq \lambda_{N-m+1}, \quad 1 < m < N$$

and define

$$S = S_1 \cup S_2, \quad S_1 = [\lambda_1, b], \quad S_2 = \bigcup_{i=N-m+1}^{N} \lambda_i$$

Define the polynomial

$$\tilde{P}_k(\lambda) = \left(\prod_{i=N-m+1}^{N} \left(1 - \frac{\lambda}{\lambda_i}\right) \right) \frac{T_{k-m}\left(\frac{b+\lambda_1-2\lambda}{b-\lambda_1}\right)}{T_{k-m}\left(\frac{b+\lambda_1}{b-\lambda_1}\right)},$$

Then the bounds on the iterations becomes

$$p(\epsilon) \leq \frac{1}{2}\sqrt{\frac{b}{\lambda_1}} \log(2/\epsilon) + m + 1,$$

A brief timeline:

- **1952** Lanczos, Hestenes & Stiefel: CG and Lanczos methods. Largely misunderstood at the time.
- **1970-1972** Paige, Reid, Godunov & Prokopov: Reintroduction of CG and Lanczos in practical application:
- **1980-1986** Axelsson, Saad & Schultz, Cullum, Greenbaum, Sonneveld: BCG, CGS, GMRES.
- **1990-1995** Freund, van der Vorst: TFQMR, BiCGSTAB. Simoncini & Gallopoulos: block GMRES
- **1975-1995** Saad, Paige, Cullum, Golub, Greenbaum, Faber & Manteuffel, Nachtigal & Reddy & Threfethen: Convergence analysis in approximate arithmetic, theoretical limits to recurrence relations

Table of Contents

2 Projection Methods

A different view: solve $Ax = b$ with a projection method:

$$x_m \quad \in \quad \{x_0 + \mathcal{K}_m\}$$

A different view: solve $Ax = b$ with a projection method:

$$\begin{aligned} x_m &\in \{x_0 + \mathcal{K}_m\} \\ b - Ax_m &\perp \mathcal{L}_m \end{aligned}$$

A different view: solve $Ax = b$ with a projection method:

$$\begin{aligned} x_m &\in \{x_0 + \mathcal{K}_m\} \\ b - Ax_m &\perp \mathcal{L}_m \end{aligned}$$

where we use a Krylov subspace

$$\mathcal{K}_m(A, r_0) = span\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}$$

A different view: solve $Ax = b$ with a projection method:

$$\begin{array}{rcl} x_m & \in & \{x_0 + \mathcal{K}_m\} \\ b - Ax_m & \perp & \mathcal{L}_m \end{array}$$

where we use a Krylov subspace

$$\mathcal{K}_m(A, r_0) = \mathit{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}$$

The Conjugate Gradient method is a specific member of this class, corresponding to one particular choice for building $\mathcal{L}_m$ and an orthonormal base of $\mathcal{K}_m$.

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

equivalent to

$$\text{Find} \quad \tilde{x} \in x_0 + \mathcal{K}, \qquad \text{such} \quad \text{that} \quad r_0 - A\delta \perp \mathcal{L}$$

$$\begin{aligned} \tilde{x} &= x_0 + \delta, & \delta &\in \mathcal{K}, \\ (r_0 - A\delta, \omega) &= 0, & \forall \omega &\in \mathcal{L} \end{aligned}$$

$V = [v_1, \ldots, v_m]$ e $W = [w_1, \ldots, w_m]$ bases (orthonormal?) of $\mathcal{K}$ and $\mathcal{L}$

$$\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, A^2r_0, \dots, A^{m-1}r_0\}$$

Common variants:

1. $\mathcal{L} = \mathcal{K}$ (also $\mathcal{L} = A\mathcal{K}$) : CG, MINRES, GMRES
2. $\mathcal{L} = \mathcal{K}(A^T, r_0)$: BiCG, BiCGSTAB

Properties

- A is (possibly implicitly) reduced to Hessenberg or tridiagonal form;
- The vectors in the bases W and V satisfy specific recurrence relations;
- The solution has some sort of “optimality” property (in exact arithmetic)

A projector is an idempotent operator:

$$P^2 = P, \qquad \mathrm{Ran}(P) = M$$

For all projectors we have

$$\mathrm{Ker}(P) = \mathrm{Ran}(I - P) = S$$

and the two subspaces M and S uniquely determine P. Define $L = S^{\perp}$; we then have

$$u = Px \in M, \qquad x - u \perp L$$

Conversely, if $M \bigcap L^{\perp} = \{0\}$ we have a projection along M orthogonal to L.

Given biorthogonal bases V and W for spaces M and L

$$(v_i, w_j) = \delta_{ij} \qquad \Rightarrow W^H V = I$$

consider $Px = Vy$; then $x - Px \perp L$, hence

$$((x - Vy), w_j) = 0 \qquad \Rightarrow W^H(x - Vy) = 0 \qquad \Rightarrow W^H x = y.$$

Therefore

$$Px = VW^H x \quad \forall x \qquad \Rightarrow P = VW^H$$

The adjoint projector P^H

$$(P^H x, y) = (x, Py) \quad \forall x, \forall y$$

is indeed a projector

$$((P^H)^2 x, y) = (P^H, Py) = (x, P^2 y) = (x, Py) = (P^H x, y)$$

and we have

$$\mathrm{Ker}(P^H) = \mathrm{Ran}(P)^\perp$$

The condition

$$M = L \quad \Rightarrow \mathrm{Ker}(P) = \mathrm{Ran}(P)^{\perp}$$

or

$$P = P^H$$

implies the decomposition

$$R^n = M \oplus M^{\perp}$$

Therefore, given a basis V of M we have

$$P = VV^H$$

which obviously satisfies $P = P^H$.

If P is a projector on M then

$$(I - P)$$

is a projector on $M^\perp$; obviously

$$x = Px + (I - P)x$$

but P is orthogonal, therefore

$$\|x\|_2^2 = \|Px\|_2^2 + \|(I - P)x\|_2^2$$

We finally have the characterization

$$\min_{y \in M} \|x - y\|_2 = \|x - Px\|_2$$

Let us now formulate a general projection method; if

$$V = [v_1, \ldots, v_m], \qquad W = [w_1, \ldots, w_m]$$

are bases of the subspaces $\mathcal{K}$ and $\mathcal{L}$, then we have

$$\begin{aligned} x &= x_0 + Vy \\ W^T A V y &= W^T r_0 \end{aligned}$$

hence the most general formulation:

repeat
  Select subspaces $\mathcal{K}$ and $\mathcal{L}$
  Choose bases $V = [v_1, \ldots, v_m]$, $W = [w_1, \ldots, w_m]$
  $r \leftarrow b - Ax$
  $y \leftarrow (W^T A V)^{-1} W^T r$
  $x \leftarrow x + Vy$
until convergence

Question: we have assumed that $W^T AV$ is nonsingular. No guarantee in general, but:

Theorem

Let $A, \mathcal{K}, \mathcal{L}$ satisfy either:

1. *A SPD and $\mathcal{L} = \mathcal{K}$, or*
2. *A nonsingular and $\mathcal{L} = A\mathcal{K}$,*

then $W^T AV$ is nonsingular for any choice of basis in $\mathcal{K}, \mathcal{L}$.

Proof, part 1.

If W and V are two basis of the same space, then there exists G non singular such that $W = VG$, hence

$$W^T AV = G^T V^T AV$$

but $V^T AV$ is SPD (since Vx is nonzero for any nonzero x, and A is SPD), hence nonsingular. □

Proof, part 2.

We can write $W = AVG$ for some nonsingular G, hence

$$W^T AV = G^T V^T A^T AV$$

but $A^T A$ is SPD since A is nonsingular, hence the projection matrix is nonsingular. □

In a one-dimentional case

$$\mathcal{K} = span\{v\} \qquad \mathcal{L} = span\{w\}$$

we have $x \leftarrow x + \alpha v$ and the Petrov-Galerkin orthogonality condition

$$r - A\delta \perp w$$

maps into

$$\alpha = \frac{(r, w)}{(Av, w)}$$

In particular, the steepest descent method is given by $v = r$ and $w = r$

$$r \leftarrow b - Ax$$
$$\alpha \leftarrow (r, r)/(Ar, r)$$
$$x \leftarrow x + \alpha v$$

GMRES Builds V explicitly;

GMRES Builds V explicitly;

CG For A SPD, builds V *implicitly*;

Indeed, the approximate solution can be written in the form:

$$A^{-1}b \approx x_m = x_0 + q_{m-1}(A)r_0.$$

The basic idea of CG is to build recursively $q_m(A)$ with a three-term recurrence, keeping A-orthogonality (cfr orthogonal polynomials).

A projection method using $\mathcal{K} = \mathcal{K}_m$, $\mathcal{L} = A\mathcal{K}_m$

$$x = x_0 + V_m y$$

Consider $\mathcal{K}_m = span\{v_1, Av_1, \ldots, A^{m-1}v_1\}$;

$$AV_m = V_m H_m + w_m e_m^T = V_{m+1}\bar{H}_m$$

$$V_m^T A V_m = H_m$$

with H_m a Hessenberg matrix. Rewriting in the new basis a $x = x_0 + V_m y$

$$b - Ax = b - A(x_0 + V_m y) = r_0 - AV_m y = \beta v_1 - V_{m+1}\bar{H}_m$$

$$b - Ax = V_{m+1}(\beta e_1 - \bar{H}_m y)$$

then we find the solution of the auxiliary problem

$$\min \|\beta e_1 - \bar{H}_m y\|_2$$

GMRES

1: Compute $r_0 \leftarrow b - Ax_0$, $\beta \leftarrow \|r_0\|$ and $v_1 \leftarrow r_0/\beta$
2: **for** $j = 1, \ldots, m$ **do**
3:     Compute $w_j \leftarrow Av_j$
4:     **for** $i = 1, \ldots, j$ **do**
5:         $h_{ij} \leftarrow (w_j, v_i)$
6:         $w_j \leftarrow w_j - h_{ij}v_i$
7:     **end for**
8:     $h_{j+1,j} \leftarrow \|w_j\|_2$ If $h_{j+1,j} = 0$ then $m \leftarrow j$ and go to 11.
9:     $v_{j+1} \leftarrow w_j/h_{j+1,j}$
10: **end for**
11: Define the $(m+1) \times m$ Hessenberg matrix $H_m = \{h_{ij}\}_{1 \le y \le m+1, 1 \le j \le m}$
12: Compute y_m the minimizer of $\|\beta e_1 - H_m y\|_2$ and set $x_m \leftarrow x_0 + V_m y_m$

To be defined

- How to solve the auxiliary least squares;
- How many steps we can afford (m).

For the first, apply an orthogonal reduction of H with Givens' rotations:

$$G(\theta) = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$$

Defining

$$y = G(i, k\theta)x$$

we have

$$y_j = \begin{cases} cx_i - sx_k & j = i \\ sx_i + cx_k & j = k \\ x_j & j \neq i, k \end{cases}$$

and $y_k = 0$ is obtained by

$$c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}} \qquad s = \frac{-x_k}{\sqrt{x_i^2 + x_k^2}}$$

For numerically stable formulae see [1], algorithm 5.1.3.

Main issue with GMRES:

We need to store explicitly the base V and apply orthogonalization to its vectors.

Hence:

- The amount of work per iteration grows with the iteration count (quadratic cost);
- The amount of storage grows: if $nnz(A) = kn$, then every k iterations we are adding as much space as for A.

Solution: Restarted GMRES $RGMRES(m)$

- Perform m steps of GMRES;
- If not converged, let $x_0 \leftarrow x_m$ and start over.

Lemma

Let x_m be the approximate solution obtained at the m-th step of GMRES, and let $r_m = b - Ax_m$. Then we have

$$x_m = x_0 + q_m(A)r_0$$

and

$$\|r_m\|_2 = \|(I - Aq_m(A))r_o\|_2 = \min_{q \in \mathbb{P}_{m-1}} \|(I - Aq(A))r_0\|_2.$$

Proof.

The result follows by construction because the vector x_m is explicitly built to minimize the 2-norm of the residual in the affine subspace $x_0 + \mathcal{K}_m$, and because $\mathcal{K}_m$ is precisely the set of all vectors that can be expressed as $x_0 + q(A)r_0$ where q is a polynomial of degree not exceeding $m - 1$. □

Theorem

Let A be a diagonalizable with $A = X\Lambda X^{-1}$; define moreover

$$\epsilon^{(m)} = \min_{p\in\mathcal{P}_m} \max_{i=1,\ldots,N} |p(\lambda_i)|.$$

Then we have the bound

$$\|\mathbf{r}_m\|_2 \leq \kappa_2(X)\epsilon^{(m)}\|\mathbf{r}_0\|_2,$$

where $\kappa_2(X) = \|X\|_2\|X^{-1}\|_2$. where $\mathcal{P}_m$ is the set of polynomials of degree not exceeding m which satisfy the constraint $p(0) = 1$.

Proof.

To any $p \in \mathcal{P}_m$ we may associate $x \in \mathcal{K}_m$ by the relation $b - Ax = p(A)r_0$; hence

$$\|b - Ax\|_2 = \|Xp(\Lambda)X^{-1}r_0\|_2 \leq \|X\|_2\|X^{-1}\|_2\|r_0\|_2\|p(\Lambda)\|_2.$$

Given that Λ is diagonal, we have $\|p(\Lambda)\|_2 = \max_{i=1,\ldots,N} |p(\lambda_i)|$; by construction x_m minimizes the residual, therefore we have

$$\begin{aligned} \|b - Ax_m\|_2 &\leq \|b - Ax\|_2 = \|Xp(\Lambda)X^{-1}r_0\|_2 \\ &\leq \|X\|_2\|X^{-1}\|_2\|r_0\|_2 \max_{i=1,\ldots,N} |p(\lambda_i)| = \kappa_2(X)\|r_0\|_2 \max_{i=1,\ldots,N} |p(\lambda_i)|, \end{aligned}$$

for *any* polynomial $p \in \mathcal{P}_m$. Using the definition of $\epsilon^{(m)}$ we obtain the thesis. □

When A is symmetric we have

$$V_m^T A V_m = H_m = T_m = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \beta_3 & \\ & \beta_3 & \alpha_3 & \beta_4 \\ & & \beta_4 & \alpha_5 \end{pmatrix}$$

tridiagonal. Factoring T:

$$T_m = L_m U_m = \begin{pmatrix} 1 & & \\ \lambda_2 & 1 & \\ & \lambda_3 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 & \beta_2 & \\ & \eta_2 & \beta_3 \\ & & \eta_3 \end{pmatrix}$$

and defining

$$P_m = V_m U_m^{-1} \qquad z_m = L_m^{-1} \beta e_1$$

we have

$$x_m = x_0 + P_m z_m$$

$$p_m = \eta_m^{-1}(v_m - \beta_m p_{m-1})$$

where

$$\begin{aligned}
\lambda_m &= \frac{\beta_m}{\eta_{m-1}} \\
\eta_m &= \alpha_m - \lambda_m \beta_m \\
\zeta_m &= -\lambda_m \zeta_{m-1} \\
x_m &= x_{m-1} + \zeta_m p_m
\end{aligned}$$

This is called the D-Lanczos algorithm

Theorem

Let r_m and p_m, $m = 0, 1, \ldots$ be respectively the residual vectors and the search directions produced by the D-Lanczos algorithm. Then

1. *Each residual r_m is such that $r_m = \sigma_m v_{m+1}$ where σ_m is a scalar. As a result, the residual vectors are orthogonal to each other.*
2. *The search directions form an A-conjugate set, i.e., $(Ap_i, p_j) = 0$ if $i \neq j$.*

Proof, part 1.

We have the following relations

$$\begin{aligned} r_m = b - Ax_m &= b - A(x_0 + V_m y_m) \\ &= r_0 - AV_m y_m \\ &= \|r_0\|_2 v_1 - V_m T_m y_m - \beta_{m+1} e_m^T y_m v_{m+1}; \end{aligned}$$

by definition we have $T_m y_m = \|r_0\|_2 e_1$, therefore $\|r_0\|_2 v_1 = V_m T_m y_m$ and thus

$$r_m = -\beta_{m+1} e_m^T y_m v_{m+1}. \tag{6}$$

i.e. r_m is in the same direction as v_{m+1}. □

Proof, part 2.

First note that $P_m^T A P_m$ is a symmetric matrix; moreover

$$
\begin{aligned}
P_m^T A P_m &= U_m^{-T} V_m^T A V_m U_m^{-1} \\
&= U_m^{-T} T_m U_m^{-1} \\
&= U_m^{-T} L_m.
\end{aligned}
$$

Thus, $U_m^{-T} L_m$ is a lower triangular matrix which is also symmetric, i.e. it is a diagonal matrix; but this is equivalent to the thesis □

Choose a starting guess x_0
Set $r_0 \leftarrow b - Ax_0$ and $p_0 \leftarrow r_0$
for $j = 0, 1, \ldots$ until convergence **do**
    $\alpha_j \leftarrow (r_j, r_j)/(Ap_j, p_j)$
    $x_{j+1} \leftarrow x_j + \alpha_j p_j$
    Check convergence
    $r_{j+1} \leftarrow r_j - \alpha_j Ap_j$
    $\beta_j \leftarrow (r_{j+1}, r_{j+1})/(r_j, r_j)$
    $p_{j+1} \leftarrow r_{j+1} + \beta_j p_j$
end for

Let us recap: when the matrix A is SPD we have

$$V_m^T A V_m = H_m = T_m = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \beta_3 & \\ & \beta_3 & \alpha_3 & \beta_4 \\ & & \beta_4 & \alpha_5 \end{pmatrix}$$

which is good, because T being tridiagonal implies we can work with just three vectors. When A is non-symmetric we have

$$V_m^T A V_m = H_m,$$

and potentially we must keep all the vectors. Is there a way to get a tridiagonal matrix from a non-symmetric A?

Krylov methods: biorthogonalization

For non-symmetric system matrix A, build *two* biorthogonal bases:

$$\begin{aligned}\mathcal{K}_m(A, v_1) &= span\{v_1, Av_1, \ldots, A^{m-1}v_1\} \\ \mathcal{K}_m(A^T, w_1) &= span\{w_1, A^T w_1, \ldots, (A^T)^{m-1}w_1\}\end{aligned}$$

i.e.

$$(v_i, w_j) = \delta_{ij}$$

Then we have

$$W_m^T A V_m = T_m$$

with T tridiagonal.

But the price to pay is:

- We have two orthogonal transformations instead of one;
- We (may) need to perform products with A^T;
- The method may exhibit *breakdowns* , i.e. $(v_i, w_i) = 0$

Choose v_1 and w_1 such that $(v_1, w_1) = 1$
Set $\beta_1 = \delta_1 = 0$ and $v_0, w_0 = 0$
for $j = 1, 2, \ldots, m$ **do**

$\alpha_i \leftarrow (Av_j, w_j)$
$\hat{v}_{j+1} \leftarrow Av_j - \alpha_j v_j - \beta_j v_{j-1}$
$\hat{w}_{j+1} \leftarrow A^T w_j - \alpha_j w_j - \delta_j w_{j-1}$
$\delta_{j+1} \leftarrow |(\hat{v}_{j+1}, \hat{w}_{j+1})|^{1/2}$ If $\delta_{j+1} = 0$ stop.
$\beta_{j+1} \leftarrow (\hat{v}_{j+1}, \hat{w}_{j+1})/\delta_{j+1}$
$w_{j+1} \leftarrow \hat{w}_{j+1}/\beta_{j+1}$
$v_{j+1} \leftarrow \hat{v}_{i+1}/\delta_{j+1}$

end for

N.B.: any choice such that

$$\delta_{j+1}\beta_{j+1} = (\hat{v}_{j+1}, \hat{w}_{j+1})$$

would be fine.

Compute $r^{(0)} \leftarrow b - Ax^{(0)}$; choose $\tilde{r}_0 : \ (r_0, \tilde{r}_0) \neq 0$
Set $p_0 = r_0$; $\tilde{p}_0 = \tilde{r}_0$; $i = 0$;
for $i < imax$, until convergence **do**

$\alpha_i \leftarrow (r_i, \tilde{r}_i)/(Ap_i, \tilde{p}_i)$
$x_{i+1} \leftarrow x_i + \alpha_i p_i$
$r_{i+1} \leftarrow r_i - \alpha_i A p_i$
$\tilde{r}_{i+1} \leftarrow \tilde{r}_i - \alpha_i A^T \tilde{p}_i$
$\beta_i \leftarrow (r_{i+1}, \tilde{r}_{i+1})/(r_i, \tilde{r}_i)$
$p_{i+1} \leftarrow r_{i+1} + \beta_i p_i$
$\tilde{p}_{i+1} \leftarrow \tilde{r}_{i+1} - \beta_i \tilde{p}_i$

end for

The biorthogonality condition is expressed as follows:

$$(r_i, \tilde{r}_j) = (Ap_i, \tilde{p}_j) = 0 \quad \text{for} \quad i \neq j.$$

All methods based on biconjugation may suffer from

Breakdown

When the denominator of an expression such as

$$\alpha_i \leftarrow (r_i, \tilde{r}_i)/(Ap_i, \tilde{p}_i)$$

or

$$\beta_i \leftarrow (r_{i+1}, \tilde{r}_{i+1})/(r_i, \tilde{r}_i)$$

is zero, the iteration cannot proceed.

A common strategy is to set

$$x_0 = x_i,$$

and *restart* the iteration.

Residuals and search directions can be written as

$$r_j = \phi_j(A)r_0 \qquad p_j = \pi_j(A)r_0 \tag{7}$$

$$\tilde{r}_j = \phi_j(A^T)\tilde{r}_0 \qquad \tilde{p}_j = \pi_j(A^T)\tilde{r}_0 \tag{8}$$

From this we can derive

$$\alpha_j = \frac{(\phi_j(A)r_0, \phi_j(A^T)\tilde{r}_0)}{(A\pi_j(A)r_0, \pi_j(A^T)\tilde{r}_0)} = \frac{(\phi_j^2(A)r_0, \tilde{r}_0)}{(A\pi_j^2(A)r_0, \tilde{r}_0)} \tag{9}$$

Therefore an alternative is to search for

$$r_j' = \phi_j^2(A)r_0$$

so that only producs by A are involved.

Compute $r^{(0)} \leftarrow b - Ax^{(0)}$; choose $\tilde{r}_0$
Set $p_0 = u_0 = r_0$;
for $i = 1, 2, \ldots imax$, until convergence **do**

$\alpha_i \leftarrow (r_i, \tilde{r}_0)/(Ap_i, \tilde{r}_0)$

$q_i \leftarrow u_i - \alpha_i A p_i$

$x_{i+1} \leftarrow x_i + \alpha_i(u_i + q_i)$

$r_{i+1} \leftarrow r_i - \alpha_i A(u_i + q_i)$

$\beta_i \leftarrow (r_{i+1}, \tilde{r}_0)/(r_i, \tilde{r}_0)$

$u_{i+1} \leftarrow r_{i+1} + \beta_i q_i$

$\tilde{p}_{i+1} \leftarrow u_{i+1} + \beta_i(q_i + \beta p_i)$

end for

CGS convergence history is often erratic.
Therefore, modify by computing

$$r_j' = \psi_j(A)\phi_j(A)r_0$$

where ψ is built to smooth out variations

$$\psi_{j+1}(t) = (1 - \omega_j t)\psi_j(t)$$

Krylov methods: BiCGSTAB

Compute $r^{(0)} \leftarrow b - Ax^{(0)}$; choose $\tilde{r}_0$
Set $p_0 = r_0$; $i = 0$;
for $i = 1, 2, \ldots imax$, until convergence **do**
  $\alpha_i \leftarrow (r_i, \tilde{r}_i)/(Ap_i, \tilde{r}_i)$
  $s_i \leftarrow r_i - \alpha_i A p_i$
  $\omega_i \leftarrow (As_i, s_i)/(As_i, As_i)$
  $x_{i+1} \leftarrow x_i + \alpha_i p_i + \omega_i s_i$
  $r_{i+1} \leftarrow s_i - \omega_i A s_i$
  $\beta_i \leftarrow (\alpha_i/\omega_i) \times (r_{i+1}, \tilde{r}_0)/(r_i, \tilde{r}_0)$
  $p_{i+1} \leftarrow r_{i+1} + \beta_i(p_i - \omega_i A p_i)$
end for

BiCGSTAB: BiConjugate Gradient Stabilized.

Is there a "universally good" method?

Is there a "universally good" method? Unfortunately, no.

Indeed, there are good reasons to think no such method can exist: in [3], a set of methods is applied to some matrices constructed in a specific way, such that each and every method is sometimes the best, sometimes the worst, sometimes average.

In approximate arithmetic, many of the properties we have stated break down, because of rounding errors

References

[1] G. H. Golub and C. F. Van Loan. *Matrix computations*. Fourth. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2013, pp. xiv+756. ISBN: 978-1-4214-0794-4.

[2] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, 1995.

[3] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen. "How Fast are Nonsymmetric Matrix Iterations?" In: *SIAM Journal on Matrix Analysis and Applications* 13.3 (1992), pp. 778–795. DOI: 10.1137/0613049. eprint: https://doi.org/10.1137/0613049. URL: https://doi.org/10.1137/0613049.

[4] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Second. Society for Industrial and Applied Mathematics, 2003. DOI: 10.1137/1.9780898718003. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898718003. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898718003.