Solvers I — Basics

Lattice Practices, September 2024, Cyprus

Gustavo Ramirez-Hidalgo

Bergische Universität Wuppertal

(most of this material has been originally created by Karsten Kahl and Andreas Frommer)

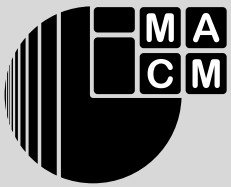


Table of Contents

Motivation and Notation
- Motivation
- Notations

Direct methods

Iterative solvers
- Introduction
- Optimal Krylov subspace methods

The Krylov Zoo
- The Zoo
- Summary

Some extra material

Linear systems of equations in Lattice QCD

In Lattice QCD codes roughly 85% of time is spent solving linear systems of equations of the type

$$\mathbf{D}\psi = \varphi \qquad (\star)$$

Hence it is of utmost importance to find efficient solvers!

Solving $(\star)$ is required in many situations, e.g.,

- in the calculation of Propagators
- in the hybrid Monte-Carlo process

Depending on the discretization and situation

- $\mathbf{D}$ is sparse (e.g., Wilson) or dense (e.g., Overlap)
- One has to solve only for one rhs or for many rhs

All this information should influence the choice of solver!

Discretizations of the Dirac operator

Discretizations by covariant finite-differences

$$\mathbf{d}_\mu \psi_x = a^{-1}(U^\mu_{x-a\mu}\psi_{x-a\mu} - (U^\mu_x)^\dagger \psi_{x+a\mu})$$

- Wilson discretization (stabilizing 2^{nd} order term)

$$\mathbf{D}_W = \sum_{\mu=1}^{4}(\gamma_\mu \otimes \mathbf{d}_\mu + a^{-1}\mathbf{d}_\mu^2) \quad \in \mathbb{C}^{12L_s^3L_t \times 12L_s^3L_t}$$

  Non-hermitian, sparse (next-neighbor), $(\gamma_5\mathbf{D})^\dagger = \gamma_5\mathbf{D}$

- Overlap discretization (Ginsparg-Wilson)

$$\mathbf{D}_O = I + \gamma_5 \,\mathrm{sign}(\gamma_5(D_W - m)) \quad \in \mathbb{C}^{12L_s^3L_t \times 12L_s^3L_t}$$

Properties of linear systems in Lattice QCD

Typical discretizations yield linear systems $\mathbf{D}\psi = \varphi$ where

- $\mathbf{D}$ is non-hermitian, yet $(\gamma_5\mathbf{D})^\dagger = \gamma_5\mathbf{D}$
- $\mathrm{spec}(\mathbf{D})$ lies in the right half-plane
- $\mathbf{D}$ is very large (on a $32^3 \times 64$ lattice $\approx$ 25M unknowns)
- $\mathbf{D}$ is sparse, i.e., contains only next-neighbor couplings

$$\approx 100 \quad \text{non-zeroes per row}$$

Matrix-Vector operations are cheap $\mathcal{O}(L_s^3 L_t) = \mathcal{O}(V)$

In implementations $\mathbf{D} \cdot x$ is often highly optimized

$\rightarrow$ use this in solvers for $\mathbf{D}\psi = \varphi$

Notations

- Linear system of equations $\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, n$

$$Ax = b, \quad A \in \mathbb{C}^{n\times n}, x \in \mathbb{C}^n, b \in \mathbb{C}^n$$

- Euclidean inner product

$$\langle x, y\rangle_2 = y^\dagger x = \sum_{i=1}^{n} \bar{y}_i x_i$$

- Adjoint $A^\dagger$ of A w.r.t. $\langle ., . \rangle_2$

$$\langle Ax, y\rangle_2 = \langle x, A^\dagger y\rangle_2$$

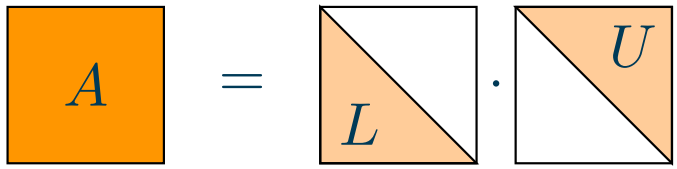
- A hermitian $\iff \quad A^\dagger = A$
- A hermitian positive definite

$$A^\dagger = A \quad \text{and} \quad x^\dagger Ax > 0, \quad x \neq 0$$

Direct methods

Idea: Solve $Ax = b$ by row-/column-manipulations

- Usually based on factorizing the system matrix A
- Methods based on Gaussian elimination
  - $A = LU$: LU factorization



  - $A = LDL^*$: Cholesky factorization (A hermitian)
- ⊕ No restrictions on applications
- ⊖ Expensive methods ($\mathcal{O}(n^3)$ for dense matrices)
- Methods exploiting sparsity exist, reducing complexity

Iterative solvers

Given: $Ax = b$ with solution $\hat{x}$, A sparse

Find: Approximations $x^{(k)}, k = 1, 2, \ldots$ s.t. $x^{(k)} \to \hat{x}$

1. How do we measure convergence $x^{(k)} \to \hat{x}$?
   - "Computable" measures ($\to$ stopping criteria)?
   - Monotone convergence in suitable norm possible?

2. How do we find iterates $x^{(k)}$ such that
   - the iterative process converges, i.e., $x^{(k)} \to \hat{x}$?
   - there is a "simple" update formula for $x^{(k+1)}$?
   - each iteration only requires the action of A on vector?

How do we measure convergence?

Given: Iterate $x^{(k)}$ in the k^{th} iteration

- Using the error $e^{(k)} = \hat{x} - x^{(k)}$

$$x^{(k)} \to \hat{x} \Longrightarrow ||e^{(k)}|| \to 0$$

In most cases the error is not readily computable!

- Using the residual $r^{(k)} = b - Ax^{(k)}$

$$x^{(k)} \to \hat{x} \Longrightarrow ||r^{(k)}|| \to 0$$

The residual is a computable quantity! Note that

$$r^{(k)} = b - Ax^{(k)} = A\hat{x} - Ax^{(k)} = Ae^{(k)}$$

In what follows we assume that $x^{(0)} = 0$

How do we find iterates $x^{(k)}$?

Task: Given b find x s.t. $Ax = b$ or

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, \ldots, n \qquad (*)$$

Idea: Solve for x_i in $(*)$ for each i

- Jacobi iteration for $i = 1, \ldots, n$

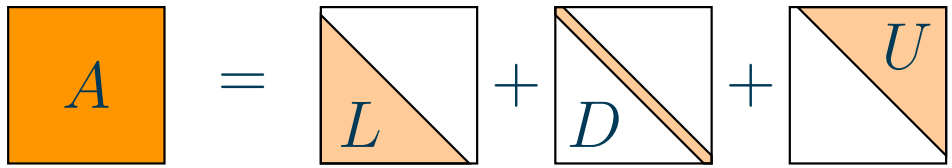
$$x_i^{(k+1)} = x_i^{(k)} + \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j^{(k)} \right)$$

- Gauss-Seidel iteration for $i = 1, \ldots, n$

$$x_i^{(k+1)} = x_i^{(k)} + \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij} x_j^{(k)} \right)$$

Splitting methods

Splitting methods use the additive decomposition of A



- Jacobi: $x^{(k+1)} = x^{(k)} + D^{-1} r^{(k)}$
- Gauss-Seidel: $x^{(k+1)} = x^{(k)} + (D + L)^{-1} r^{(k)}$
- SOR: $x^{(k+1)} = x^{(k)} + (\frac{1}{\omega} D + L)^{-1} r^{(k)}$

General splitting method: $A = M + N$

$$x^{(k+1)} = x^{(k)} + M^{-1} r^{(k)} \Longrightarrow e^{(k+1)} = e^{(k)} - M^{-1} A e^{(k)}$$

Convergent iff $\|I - M^{-1}A\| < 1$ for some norm $\|\cdot\|$

Often used as preconditioners ($\rightarrow$ Solvers II)

Linear Algebra (Minimal polynomial)

Let $p_m^\star$ be the polynomial of smallest degree m s.t.

$$p_m^\star(A) = 0 \quad \text{with} \quad p^\star(0) = 1 \;\Leftrightarrow\; p_m^\star(t) = 1 - tq_{m-1}^\star(t).$$

Consequence: $A^{-1} = q_{m-1}^\star(A)$, a polynomial in A!

$\Rightarrow$ Solution $\hat{x}$ of $Ax = b$ given by $q_{m-1}^\star(A)b$

Idea: Polynomial approximations $x^{(k)}$ of $\hat{x}$ by

$$x^{(k)} = q_k(A)b, \quad q_k(t) \in \Pi_k = \{p(t) = \sum_{\ell=0}^{k} \alpha_\ell t^\ell\}$$

Requirements: Computation of $x^{(k+1)}$ needs

- multiplication by A
- update of coefficients $\alpha_1, \ldots, \alpha_{k+1}$

Krylov subspace methods

Krylov subspace methods

Approximation $x^{(k)}$ of the solution $\hat{x}$ in Krylov subspace

$$\mathcal{K}_k(A, b) = \{p(A)b : p \in \Pi_{k-1}\} = \text{span}\{b, Ab, \ldots, A^{k-1}b\}$$

Polynomial connection:

- $x^{(k)} = q_{k-1}(A)b, \quad \deg q_{k-1} \leq k-1$
- $r^{(k)} = p_k(A)b, \;\; e^{(k)} = p_k(A)e^{(0)}, \quad p_k = 1 - tq_{k-1}$

One-to-one correspondence: Any sequence p_k with $p_k(0) = 1$ defines $r^{(k)} = p_k(A)b$, $x^{(k)} = q_{k-1}(A)b$.

Categories:

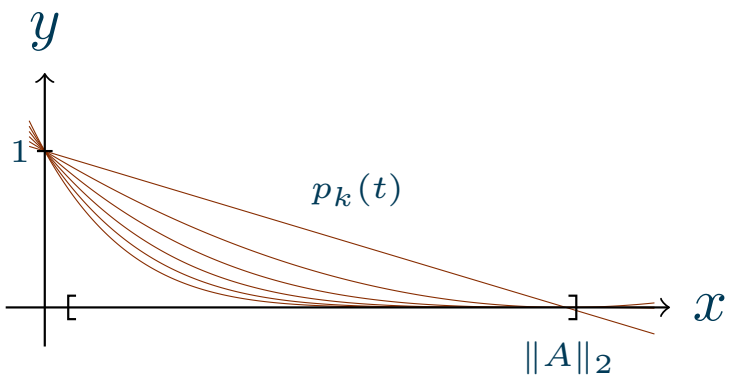
- stationary (e.g. Richardson, Chebyshev): p_k indept of b
- non-stationary (e.g. CG, GMRES, ...): p_k adapts to b

Example: Richardson iteration

For A hermitian positive definite, i.e., $\mathrm{spec}(A) \subseteq \mathbb{R}^+$

Richardson Iteration

b given, $x^{(0)} = 0$, $\alpha > \|A\|_2/2$
for $k = 0, 1, 2, \ldots$ **do**
$x^{(k+1)} = (I - \alpha^{-1}A)x^{(k)} + \alpha^{-1}b$
end for



We have

$$
\begin{aligned}
r^{(k)} &= p_k(A)b \quad \in \mathcal{K}_{k+1}(A,b) \text{ with } p_k(t) = (1-\alpha^{-1}t)^k \\
e^{(k)} &= p_k(A)e^{(0)} \\
x^{(k)} &= q_{k-1}(A)b \quad \in \mathcal{K}_k(A,b) \text{ with } p_k(t) = 1 - tq_{k-1}(t)
\end{aligned}
$$

a stationary Krylov subspace method!

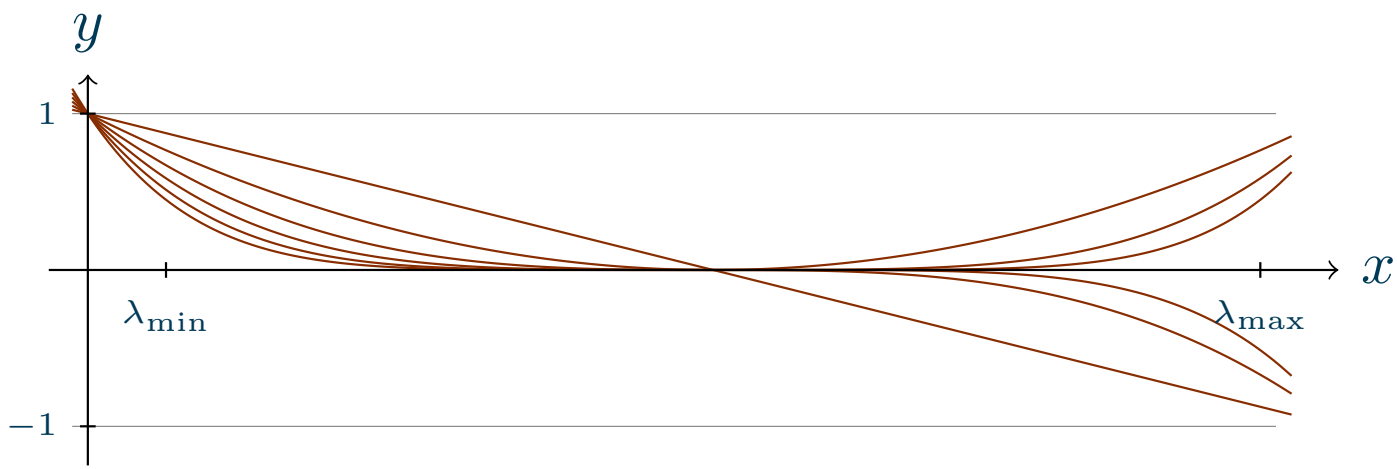
Krylov subspace methods — Theory

Does Richardson iteration converge? We have

$$\|r^{(k)}\|_2 = \|(I - \alpha^{-1}A)^k b\|_2 \leq \underbrace{\|(I - \alpha^{-1}A)\|_2}_{<1}{}^k \|b\|_2 \longrightarrow 0$$

Best choice for α :

$$\alpha = \frac{\lambda_{\max} + \lambda_{\min}}{2} \quad \Rightarrow \quad \|(I - \alpha^{-1}A)\|_2 = \frac{\kappa - 1}{\kappa + 1}, \quad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

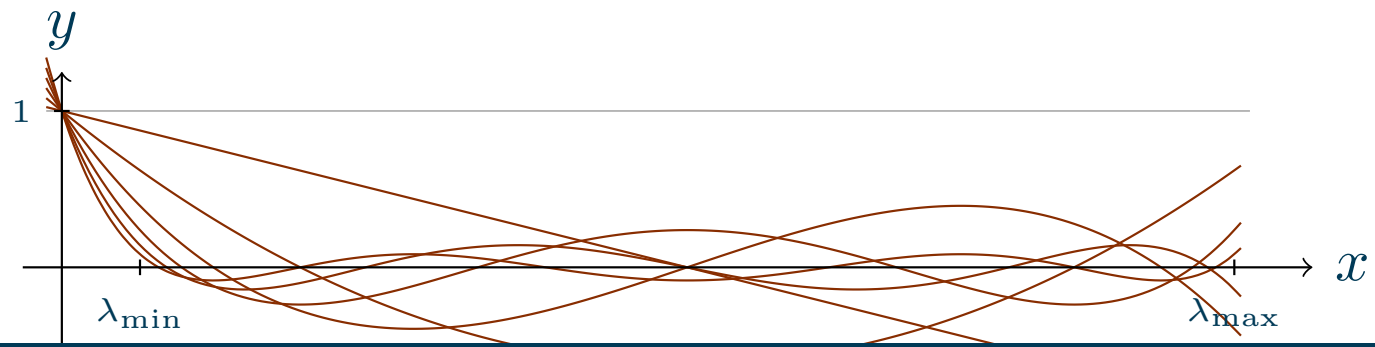


Krylov subspace methods — Theory

$$r^{(k)} = p_k(A)b, \quad e^{(k)} = p_k(A)e^{(0)}$$

Can we aim for optimality?

- Minimize $\|p_k(A)e^{(0)}\| = \|e^{(k)}\|$ over all p_k
- Without knowing $e^{(0)}$ **????**
- Works for the A-norm
  $$\|x\|_A = \langle x, x\rangle_A^{\frac{1}{2}} \text{ with } \langle x, y\rangle_A = \langle Ax, y\rangle$$
- Method of conjugate gradients (CG)



Optimal Krylov subspace methods I — Conjugate Gradients

Conjugate Gradients

$r^{(0)} = b, p^{(0)} = r^{(0)}$

for $k = 1, 2, \ldots$ **do**

$$\alpha_{k-1} = \frac{\langle r^{(k-1)}, r^{(k-1)} \rangle_2}{\langle Ap^{(k-1)}, p^{(k-1)} \rangle_2}$$

$$x^{(k)} = x^{(k-1)} + \alpha_{k-1} p^{(k-1)}$$

$$r^{(k)} = r^{(k-1)} - \alpha_{k-1} Ap^{(k-1)}$$

$$\beta_{k-1} = \frac{\langle r^{(k)}, r^{(k)} \rangle_2}{\langle r^{(k-1)}, r^{(k-1)} \rangle_2}$$

$$p^{(k)} = r^{(k)} + \beta_{k-1} p^{(k-1)}$$

end for

Minimization of the functional

$$\mathcal{L}(x) = \tfrac{1}{2}\langle x, x \rangle_A - \langle x, b \rangle_2 = \tfrac{1}{2}\left(\|e\|_A^2 - \|\hat{x}\|_A^2 \right)$$

- $p^{(k)}$ conjugate gradients of $\mathcal{L}$

Conjugate Gradients — Properties

1. A must be hermitian and positive definite
2. Minimal error in $\|.\|_A$ for $x^{(k)} \in \mathcal{K}_k(A, b)$

$$\begin{aligned} \|e^{(k)}\| = \min_{p_k \in \overline{\Pi}_k} \|p_k(A)e^{(0)}\|_A &\leq \frac{1}{\cosh\left(k \ln \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)} \|e^{(0)}\|_A \\ &\leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \|e^{(0)}\|_A \end{aligned}$$

3. Variational property $r^{(k)} \perp \mathcal{K}_k(A, b)$
4. Minimization of the functional

$$\mathcal{L}(x) = \frac{1}{2}\langle x, x\rangle_A - \langle x, b\rangle_2, \quad x \in \mathcal{K}_k(A, b)$$

5. Short recurrence, i.e., $x^{(k+1)}$ requires only $x^{(k)}$

Optimal Krylov subspace methods II — CR and MINRES

For A hermitian find the Krylov subspace method with

$$\min_{x^{(k)} \in \mathcal{K}_k(A,b)} \|b - Ax^{(k)}\|_2 = \min_{x^{(k)} \in \mathcal{K}_k(A,b)} \|r^{(k)}\|_2$$

The optimal method w.r.t. $\|.\|_2$ is known as conjugate residuals (CR)

Similar to CG, CR introduces search directions $p^{(k)}$

- The residuals are conjugate, i.e. $\langle r^{(k)}, Ar^{(\ell)} \rangle = 0$ for $k \neq \ell$
- The $Ap^{(k)}$ are mutually orthogonal
- Short recurrence: $x^{(k+1)}$ requires only $x^{(k)}$

Conjugate Residuals (CR)

$r^{(0)} = b, p^{(0)} = r^{(0)}$

for $k = 1, 2, \ldots$ **do**

$$\alpha_{k-1} = \frac{\langle r^{(k-1)}, Ar^{(k-1)} \rangle_2}{\langle Ap^{(k-1)}, Ap^{(k-1)} \rangle_2}$$

$$x^{(k)} = x^{(k-1)} + \alpha_{k-1} p^{(k-1)}$$

$$r^{(k)} = r^{(k-1)} - \alpha_{k-1} Ap^{(k-1)}$$

$$\beta_{k-1} = \frac{\langle r^{(k)}, Ar^{(k)} \rangle_2}{\langle r^{(k-1)}, Ar^{(k-1)} \rangle_2}$$

$$p^{(k)} = r^{(k)} + \beta_{k-1} p^{(k-1)}$$

Compute $Ap_k = Ar_k + \beta_{k-1} Ap_{k-1}$

end for

- CR can break down (division by 0) if A is indefinite
- MINRES is a break-down free, short-recurrence realization of CR (works for any hermitian A)

Optimal Krylov subspace methods III — GMRES

What if A is not hermitian? Optimality w.r.t. $\|.\|_2$ possible

$$\min_{x^{(k)} \in \mathcal{K}_k(A,b)} \|b - Ax^{(k)}\|_2 = \min_{x^{(k)} \in \mathcal{K}_k(A,b)} \|r^{(k)}\|_2$$

Idea: For orthonormal basis $v_1, \ldots, v_k$ of $\mathcal{K}_k(A, b)$

$$x \in \mathcal{K}_k(A,b) \Longrightarrow x = \sum_{\ell=1}^{k} v_\ell y_\ell = [v_1 \mid \ldots \mid v_k] y = V_k y$$

Hence we find

$$\min_{x^{(k)} \in \mathcal{K}_k(A,b)} \|b - Ax^{(k)}\|_2 = \min_{y} \|b - AV_k y\|_2$$

The Arnoldi Iteration

Compute orthonormal basis $\{v_1, \ldots, v_k\}$ of $\mathcal{K}_k(A, b)$

Arnoldi Iteration

$\beta = \|b\|_2, v_1 = \beta^{-1} b$
for $k = 1, 2, \ldots$ **do**
  $q = Av_k$
  for $j = 1, \ldots, k$ **do**
    $h_{j,k} = \langle v_j, q \rangle_2$
    $q = q - h_{j,k} v_j$
  end for
  $h_{k+1,k} = \|q\|_2$
  $v_{k+1} = h_{k+1,k}^{-1} q$
end for

With $V_k = [v_1 \mid \ldots \mid v_k]$ and

$$H_{k+1,k} =$$

we have the Arnoldi relation

$$AV_k = V_{k+1} H_{k+1,k}$$

with $V_k^\dagger V_k = I$

Full GMRES

Using the Arnoldi relation in $(\star)$ we find

$$\min_y \|b - AV_k y\|_2 = \min_y \|b - V_{k+1} H_{k+1,k} y\|_2$$

Since V_{k+1} has orthonormal columns and $v_1 = \|b\|_2^{-1} b$

$$\min_y \|b - V_{k+1} H_{k+1,k} V_k y\|_2 = \min_y \|\|b\|_2 e_1 - H_{k+1,k} y\|_2$$

Naïve GMRES

for $k = 1, 2, \ldots$ **do**
  Compute $v_k, H_{k+1,k}$ *(Arnoldi)*
  Solve $\operatorname{argmin}_y \|\|b\|_2 e_1 - H_{k+1,k} y\|_2$
  $x^{(k)} = V_k y$
end for

Restarted GMRES

Although an optimal method, GMRES has severe drawbacks:

- The computation of $x^{(k)}$ requires $V_k = [v_1 \mid \ldots \mid v_k]$
  - $\Rightarrow$ Storage requirements grow with k
  - $\Rightarrow$ Computation time in Arnoldi grows with k
- Least-Squares solution requires $\mathcal{O}(k^3)$ operations

Idea: Restart GMRES every m-iterations ($\Rightarrow$ GMRES(m))

GMRES(m)

for $\ell = 0, 1, \ldots$ **do**

$r^{(0)} = b - Ax^{(0)}, \beta = \|r^{(0)}\|_2, v_1 = \beta^{-1} r^{(0)}$

Compute $V_m, H_{m+1,m}$ *(Arnoldi)*

$y_m = \operatorname{argmin}_y \|\beta e_1 - H_{m+1,m} y\|_2$

$x^{(0)} = x^{(0)} + V_m y_m$

end for

Optimal Krylov subspace methods — Summary

	requirements	optimality	recurrence
CG	$A = A^\dagger$ $\langle x, x \rangle_A > 0, x \neq 0$	$\|.\|_A$	short
MINRES*	$A = A^\dagger$	$\|.\|_2$	short
GMRES†	none	$\|.\|_2$	long

- What are the requirements for short recurrence?
- Do non-optimal methods exist with short recurrence?
  (see the extra material slides on Faber-Manteuffel and Barth-Manteuffel)

See also the two extra-material slides on BiCGstab (a non-optimal Krylov subspace method – used in the LQCD community for a long time).

*mathematically equivalent, but possibly unstabe: CR

†mathematically equivalent, but possibly unstabe: GCR

List of Methods

	requirements	optimal	recurrence	
CG	hpd	$\|.\|_A$	short	
MINRES	hermitian	$\|.\|_2$	short	
GMRES	none	$\|.\|_2$	long	$\rightarrow$ restarts
CGN	none	$\|.\|_{A^\dagger A}$	short	$A^\dagger A x = A^\dagger b$
BCG	none	no	short	similar to CG unstable
QMR	none	no	short	similar to GMRES
BiCGstab	none	no	short	breakdowns
SUMR	shifted unitary	$\|.\|_2$	short	multiple recursion

More on Krylov subspace methods: [3, 6].

Krylov subspace methods are all-duty solvers

- require only multiplication by $A\cdot$ and inner products
  - easy to implement (especially if $A\cdot$ is already done)
  - easy to parallelize ($\log(p)$-scaling due to inner products)
- Whenever short-term recurrence is possible
  - constant cost per iteration
  - constant memory consumption
- If only long-term recurrence is possible
  - restarts limit amount of work
  - deflated-restarts are even more efficient ($\rightarrow$ Solvers II)
- Convergence speed depends on the spectrum of A
  - separation from the origin important (since $p_k \in \overline{\Pi}_k$)
  - preconditioning improves performance ($\rightarrow$ Solvers II)

[1] T. Barth and T. Manteuffel.
Multiple recursion conjugate gradient algorithms. I. sufficient conditions.
SIAM J. Matrix Anal. Appl., 21, 2000.

[2] V. Faber and T. Manteuffel.
Necessary and sufficient conditions for the existence of a conjugate gradient method.
SIAM J. Numer. Anal., 21, 1984.

[3] A. Greenbaum.
Iterative Methods for Solving Linear Systems, volume 17 of *Frontiers in Applied Mathematics*.
Society for Industrial and Applied Mathematics, 1997.

[4] M. Hestenes and E. Stiefels.
Methods of conjugate gradients for solving linear systems.
Journal of Research of the National Bureau of Standards, Section B, 49, 1952.

[5] C. Jagels and L. Reichel.
A fast minimal residual algorithm for shifted unitary matrices.
Numer. Linear Algebra Appl., 1, 1994.

[6] Y. Saad.
Iterative Methods for Sparse Linear Systems.
Society for Industrial and Applied Mathematics, 2nd edition, 2003.

Some extra material

Krylov subspace methods — Theory

(particularly useful right after Richardson's convergence analysis)

$$r^{(k)} = p_k(A)b, \quad e^{(k)} = p_k(A)e^{(0)}$$
$$\implies \quad \|r^{(k)}\| \leq \|p_k(A)\|\|b\|, \quad \|e^{(k)}\| \leq \|p_k(A)\|\|e^{(0)}\|$$

Notation: $\overline{\Pi}_k = \{p \in \Pi_k, p(0) = 1\}$

Better than Richardson:

- $\|p_k(A)\|_2 = \max_{\lambda \in \text{spec}(A)} |p_k(\lambda)|$
- "best": $p_k = \text{argmin } \left\{\|\tilde{p}_k(A)\|_2,\ \tilde{p}_k \in \overline{\Pi}_k\right\}$
- $\|p_k\|_{[\lambda_{\min}, \lambda_{\max}]} = \max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |p_k(\lambda)| \ \geq \ \|p_k(A)\|_2$
- "second best": $p_k = \text{argmin } \left\{\|\tilde{p}_k(A)\|_{[\lambda_{\min}, \lambda_{\max}]},\ \tilde{p}_k \in \overline{\Pi}_k\right\}$
- Solution for "second best" is known:
  Chebyshev polynomials $\Rightarrow$ Chebyshev iteration

Faber-Manteuffel Theorem

The Faber-Manteuffel Theorem

There exists an optimal method with $(s+2)$-term recurrence iff A is s-normal, i.e., $A^\dagger = p(A), \quad p \in \Pi_s$

- A normal $\Longrightarrow A^\dagger = p(A), \quad p \in \Pi_{n-1}$
- A hermitian, $A^\dagger = A \longrightarrow$ (3-term recurrence)
- A anti-hermitian $A^\dagger = -A \longrightarrow$ (3-term recurrence)
  - Chiral operator $\gamma_5 D = -D\gamma_5 \Longrightarrow D^\dagger = -D$

Barth-Manteuffel Theorem — Ginsparg-Wilson relation

The Barth-Manteuffel Theorem

There exists an optimal method with $(s+2,t)$-term recurrence iff A is (s,t)-normal, i.e.,

$$A^\dagger = \frac{p(A)}{q(A)}, \quad p \in \Pi_s, q \in \Pi_t$$

- Methods have multiple recursions
- Occurrence in Lattice QCD: Ginsparg-Wilson relation

$$\begin{aligned} D\gamma_5 + \gamma_5 D = aD\gamma_5 D &\iff \gamma_5(I-aD)D^\dagger = -\gamma_5 D \\ &\iff D^\dagger = (I-aD)^{-1}D \end{aligned}$$

- D fulfills Ginsparg-Wilson $\iff$ D is $(1,1)$-normal

Non-optimal Krylov subspace methods — BCG

What if no optimal short recurrence method exists for A?

Ansatz: Throw optimality over board!

- Instead of building one Krylov subspace build two

$$\mathcal{K}_k(A, r^{(0)}) \quad \text{and} \quad \mathcal{K}_k(A^\dagger, \tilde{r}^{(0)})$$

  - bi-orthogonalization
- Similar to CG, the residuals of BCG fulfill

$$r^{(k)} \perp \mathcal{K}_k(A^\dagger, \tilde{r}^{(0)})$$

- Not optimal in any norm
  - erratic convergence behaviour ($\rightarrow$ excercises)
  - breakdowns can occur, i.e., convergence not guaranteed
- BiCGstab is a stabilized variant of BCG

BiCGstab

$r^{(0)} = b, \beta_0 = 0$

$\hat{r} = r$ shadow residual $\langle r, \hat{r} \rangle_2 \neq 0$

for $k = 0, 1, \ldots$ **do**

$\rho_k = \langle r^{(k)}, \hat{r} \rangle_2$

$\beta_k = \frac{\rho_k}{\rho_{k-1}} \cdot \frac{\alpha_{k-1}}{\omega_{k-1}}$

$p^{(k)} = r^{(k)} + \beta_k (p^{k-1} - \omega_{k-1} v^{(k-1)})$

$\alpha_k = \frac{\rho_k}{\langle Ap^{(k)}, \hat{r} \rangle_2}$

$x^{(k+\frac{1}{2})} = x^{(k)} + \alpha_k p^{(k)}$

$s^{(k)} = r^{(k)} - \alpha_k A p^{(k)}$ $s^{(k)} \equiv r^{(k+\frac{1}{2})}$

$\omega_k = \frac{\langle s^{(k)}, As^{(k)} \rangle_2}{\langle As^{(k)}, As^{(k)} \rangle_2}$

$x^{(k+1)} = x^{(k+\frac{1}{2})} + \omega_k s^{(k)}$

$r^{(k+1)} = s^{(k)} - \omega_k A s^{(k)}$

end for