SI231b: Matrix Computations

Lecture 22: State-of-the-art Iterative Methods: Krylov Subspace Methods

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 23, 2022

Start-of-the-art Iterative Methods

The Jacobi, Gauss-Seidel, and SOR iteration are all *stationary iteration*, i.e., they all have the form

$$\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{c},$$

where neither **B** nor **c** depends on k.

The convergence of stationary iteration cannot be guaranteed and often slow once converged (recall Lecture 21).

Start-of-the-art iterative methods belong to the category of **Krylov subspace methods**, where the approximate solution is searched in a **low-dimensional subspace**

$$\mathcal{K}_k(\mathbf{A},\ \mathbf{b}) = \operatorname{span}\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \cdots, \mathbf{A}^{k-1}\mathbf{b}\right\}.$$

Motivation: Cayley-Hamilton Theorem

For $\mathbf{A} \in \mathbb{R}^{n\times n}$, its characteristic polynomial is given by

$$p_A(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots\cdots + c_1\lambda + c_0,$$

where $c_0 = (-1)^n \det(\mathbf{A})$.

The Cayley-Hamilton Theorem states that

$$p_A(\mathbf{A}) = \mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots\cdots + c_1\mathbf{A} + c_0\mathbf{I} = 0.$$

For nonsingular $\mathbf{A}$, this in turn gives

$$\mathbf{A}^{-1} = -\frac{(-1)^n}{\det(\mathbf{A})}\left(\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots\cdots + c_1\mathbf{I}\right)$$

Therefore, the solution for $\mathbf{A}\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = -\frac{(-1)^n}{\det(\mathbf{A})}\left(c_1\mathbf{b} + c_2\mathbf{A}\mathbf{b} + c_3\mathbf{A}^2\mathbf{b} + \cdots\cdots + \mathbf{A}^{n-1}\mathbf{b}\right)$$

Krylov subspace methods compute the approximated solution from the low-dimensional subspace

$$\mathcal{K}_k(\mathbf{A},\ \mathbf{b}) = \text{span}\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \cdots, \mathbf{A}^{k-1}\mathbf{b}\right\}.$$

Krylov Subspace Methods

$$\mathcal{K}_1(\mathbf{A},\ \mathbf{b}) \subset \mathcal{K}_2(\mathbf{A},\ \mathbf{b}) \subset \mathcal{K}_3(\mathbf{A},\ \mathbf{b}) \subset \cdots\cdots \subset \mathcal{K}_n(\mathbf{A},\ \mathbf{b})$$

The Krylov subspace methods compute the iterative solution successive from $\mathcal{K}_1(\mathbf{A},\ \mathbf{b}),\ \mathcal{K}_2(\mathbf{A},\ \mathbf{b}),\ \cdots\cdots$ with better approximation of $\mathbf{A}^{-1}\mathbf{b}$.

Better or optimal approximation often refers to some sort of projection, Krylov subspace methods are also called Krylov projection methods.

Krylov subspace methods can be distinguished in four different classes,

- Ritz-Galerkin approach: construct $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ so that the residual $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \ \perp \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$;
- minimum residual norm approach: compute $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ so that the norm of the residual $\|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|_2$ is minimal over $\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$;
- Petrov-Galerkin approach: find $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ so that the residual $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$ is orthogonal to some other k-dimensional subspace;
- minimum error norm approach: determine $\mathbf{x}_k \in \mathbf{A}^T\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ so that $\|\mathbf{x}_k - \mathbf{x}\|_2$ is minimal.

Widely Used Krylov Subspace Methods

Widely used Krylov subspace methods include

- conjugate gradient (CG) [1952]:
  - for symmetric positive definite systems;
  - Ritz-Galerkin type.
- minimal residual (MINRES) [1975]:
  - for symmetric indefinite systems;
  - minimum residual norm approach.
- generalized minimal residual (GMRES) [1986]:
  - for non-symmetric systems;
  - minimum residual norm approach.

Conjugate Gradient

Motivation: for symmetric positive definite (SPD) matrix $\mathbf{A}$, the residual $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$ should be orthogonal to $\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$.

Some facts:

1. orthogonal residuals: $\mathbf{r}_i^T \mathbf{r}_k = 0$ for $i < k$
   - $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b}) \to \mathbf{r}_k \in \mathcal{K}_{k+1}(\mathbf{A},\ \mathbf{b})$;
   - $\mathbf{r}_i \in \mathcal{K}_{i+1}(\mathbf{A},\ \mathbf{b}) \subset \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$.
2. conjugate ($\mathbf{A}$-orthogonal) update directions: $(\mathbf{x}_i - \mathbf{x}_{i-1})^T \mathbf{A}(\mathbf{x}_k - \mathbf{x}_{k-1}) = 0$ for $i < k$.

   Can you show/prove this?

Conjugate Gradient Iteration Algorithms

With the key properties introduced before, we have the CG iteration

CG Iteration:

```
$\mathbf{x}_0 = 0,\ \mathbf{r}_0 = \mathbf{b},\ \mathbf{d}_0 = \mathbf{b}$
while $\|r_k\| >$ tol & $k <$ max_iter
        $\alpha_k = \frac{\mathbf{r}_{k-1}^T\mathbf{r}_{k-1}}{\mathbf{d}_{k-1}^T\mathbf{A}\mathbf{d}_{k-1}}$   step size of solution update
        $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k\mathbf{d}_{k-1}$
        $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k\mathbf{A}\mathbf{d}_{k-1}$
        $\beta_k = \frac{\mathbf{r}_k^T\mathbf{r}_k}{\mathbf{r}_{k-1}^T\mathbf{r}_{k-1}}$   step size of search direction update
        $\mathbf{d}_k = \mathbf{r}_k + \beta_k\mathbf{d}_{k-1}$   new search direction
        $k \leftarrow k+1$
end
```

Facts:

$$\mathcal{K}_k(\mathbf{A},\ \mathbf{b}) = \mathrm{span}\{\mathbf{x}_1,\ \mathbf{x}_2,\ \cdots\cdots,\ \mathbf{x}_k\} = \mathrm{span}\{\mathbf{d}_0,\ \mathbf{d}_1,\ \cdots\cdots,\ \mathbf{d}_{k-1}\}$$

Convergence of Conjugate Gradient

Optimality of CG

Let CG be applied to an SPD system $\mathbf{Ax} = \mathbf{b}$, then $\mathbf{x}_k$ is the unique point in $\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ that minimizes $\|\mathbf{x}_k - \mathbf{x}\|_{\mathbf{A}}$, and the convergence of CG is monotonic, i.e.,

$$\|\mathbf{x}_{k+1} - \mathbf{x}\|_{\mathbf{A}} \leq \|\mathbf{x}_k - \mathbf{x}\|_{\mathbf{A}}.$$

Here $\|\mathbf{z}\|_{\mathbf{A}} = \mathbf{z}^T\mathbf{A}\mathbf{z}$ for arbitrary $\mathbf{z}$.

Can you prove this?

Convergence rate of CG

Let CG be applied to an SPD system $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{A}$ has 2-norm condition number κ, then

$$\frac{\|\mathbf{x}_k - \mathbf{x}\|_{\mathbf{A}}}{\|\mathbf{x}_0 - \mathbf{x}\|_{\mathbf{A}}} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$$

Generalized Minimum Residual (GMRES)

Motivation:

CG only applies to symmetric positive definite problems, and the residual may not be monotonic decreasing.

At iteration k, we shall search $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$ such that $\|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|_2$ is minimal, i.e., we should solve

$$\min_{\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})} \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|_2.$$

Denote $\mathbf{K}_k = \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{b} & \cdots & \mathbf{A}^{k-1}\mathbf{b} \end{bmatrix}$ as the Krylov matrix, since $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A},\ \mathbf{b})$, then $\mathbf{x}_k = \mathbf{K}_k\mathbf{c}$ with $\mathbf{c} \in \mathbb{R}^k$, then $\mathbf{x}_k$ is given by solving the least square problem

$$\min_{\mathbf{c}} \|\mathbf{b} - \mathbf{A}\mathbf{K}_k\mathbf{c}\|_2.$$

However, this scheme is numerically unstable due to the ill-conditioning of $\mathbf{K}_k$ (recall the power method).

Mechanism of GMRES

Instead of using $\mathbf{b},\ \mathbf{Ab},\ \mathbf{A}^2\mathbf{b},\ \cdots$ as basis, we compute an orthonormal basis of $\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$, i.e.,

$$\mathcal{K}_k(\mathbf{A},\ \mathbf{b}) = \text{span}\{\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots,\ \mathbf{q}_k\}.$$

Let $\mathbf{Q}_k = [\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots,\ \mathbf{q}_k]$, then the least square problem becomes

$$\min_{\mathbf{c}} \|\mathbf{b} - \mathbf{AQ}_k\mathbf{c}\|_2.$$

To summarize, we need

1. compute the orthonormal basis $\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots,\ \mathbf{q}_k$ of $\mathcal{K}_k(\mathbf{A},\ \mathbf{b})$;
2. compute the QR factorization of $\mathbf{AQ}_k$ to solve the least square problem.

Arnoldi Iteration

Starting from $\mathbf{q}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|_2}$, the above two steps

1. QR factorization of the Krylov matrix $\mathbf{K}_k$;
2. QR factorization of $\mathbf{AQ}_k$.

can be done simultaneously by the **Arnoldi iteration** $\mathbf{AQ}_k = \mathbf{Q}_{k+1}\tilde{\mathbf{H}}_k$.

$$\mathbf{A}\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k\end{bmatrix} = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_{k+1}\end{bmatrix}\underbrace{\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & \vdots \\ & h_{32} & \ddots & \vdots \\ & & \ddots & \\ & & h_{k,k-1} & h_{k,k} \\ & & & h_{k+1,k}\end{bmatrix}}_{\tilde{\mathbf{H}}_k}$$

Arnoldi Iteration

The Arnoldi iteration can be computed in a stable manner using modified Gram-Schmidt method,

Arnoldi Iteration:

```
$\mathbf{q}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|_2}$
for k = 1, 2, ⋯
    $\mathbf{v} = \mathbf{A}\mathbf{q}_k$
    for j = 1, 2, ⋯, k
        $h_{j,k} = \mathbf{q}_j^T \mathbf{v}$
        $\mathbf{v} = \mathbf{v} - h_{j,k}\mathbf{q}_j$
    end
    $h_{k+1,k} = \|\mathbf{v}\|_2$
    $\mathbf{q}_{k+1} = \frac{\mathbf{v}}{h_{k+1,k}}$
end
```

Note: the Arnoldi iteration is a long-term recursion, i,e, the new vector should be orthogonal projected onto all previous basis vectors.

From Arnoldi to Lanczos

From the Arnoldi iteration $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\tilde{\mathbf{H}}_k$, we observe that

$$\mathbf{Q}_k^T\mathbf{A}\mathbf{Q}_k = \mathbf{H}_k,$$

where $\mathbf{H}_k$ is a $k \times k$ matrix obtained by removing the last row of $\tilde{\mathbf{H}}_k$.

The eigenvalues of $\mathbf{H}_k$ are called *Ritz values* or *Arnoldi eigenvalue estimates*, which are good approximation of the k largest eigenvalues of $\mathbf{A}$ [Trefthen & Bau 97].

When $\mathbf{A}$ is symmetric, the Arnoldi iteration then becomes the Lanczos iteration that gives

$$\mathbf{Q}_k^T\mathbf{A}\mathbf{Q}_k = \mathbf{T}_k,$$

where $\mathbf{T}_k$ is a tridiagonal matrix. Therefore, the Lanczos iteration is a *short-term* recurrence.

Efficient GMRES Implementation and Convergence

At step k, the least square problem

$$\begin{aligned}\min_{\mathbf{c}} \|\mathbf{b} - \mathbf{A}\mathbf{Q}_k\mathbf{c}\|_2 &= \min_{\mathbf{c}} \|\mathbf{b} - \mathbf{Q}_{k+1}\tilde{\mathbf{H}}_k\mathbf{c}\|_2 \\ &= \min_{\mathbf{c}} \|\mathbf{Q}_{k+1}^T\mathbf{b} - \tilde{\mathbf{H}}_k\mathbf{c}\|_2 \\ &= \min_{\mathbf{c}} \left\| \|\mathbf{b}\|_2\mathbf{e}_1 - \tilde{\mathbf{H}}_k\mathbf{c} \right\|_2,\end{aligned}$$

with $\mathbf{e}_1 = [1, 0, \cdots, 0]^T$.

The least square problem is of size $(k+1) \times k$, and solving this least square problem using QR factorization takes only $\mathcal{O}(k^2)$ flops. (how and why?)

The convergence of GMRES is complicated to analyze, we omit the details but just give two useful conclusions:

- The GMRES converges monotoniclly, i.e.,

$$\|\mathbf{r}_{k+1}\|_2 \leq \|\mathbf{r}_k\|_2$$

- GMRES gives exact solution (without rounding-off error) at most n iterations, i.e., $\|\mathbf{r}_n\|_2 = 0$

Readings

You are supposed to read

- Gene H. Golub and Charles F. Van Loan. Matrix Computations, *Johns Hopkins University Press*, 2013.

  Chapter 11.2, 11.3.

- Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra, SIAM, 1997.

  Lecture 33, 35, 38

Announcement

Our final exam takes places

- **when**: Dec. 19, 2022 from 15:00 to 17:00;
- **where**: Teaching Center 101;
- **what**: SI231b Matrix Computations;
- **how**: one page cheat-sheet of A4 size is allowed.

If you are interested to be a TA, you are welcome to contact me.

You are welcome to write useful suggestions/comments/feedback.

DO REMEMBER to submit your final report in time!