# Sketching Krylov Subspace Methods
# – Linear Systems –
# Lecture 23

F. M. Faulstich

04/19/2024

# Conjugate Gradient

- Method of steepest descent

# Conjugate Gradient

- Method of steepest descent
  $\Rightarrow$ Rewrite the problem as minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

  for $\mathbf{A} \in \mathbb{H}_n$ and $\mathbf{A} \succ \mathbf{0}$.

# Conjugate Gradient

- Method of steepest descent
  $\Rightarrow$ Rewrite the problem as minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

  for $\mathbf{A} \in \mathbb{H}_n$ and $\mathbf{A} \succ \mathbf{0}$.
  $\Rightarrow$ "walk" in the direction of the steepest descent
  $\Rightarrow$ Very intuitive but – in general – not optimal

# Conjugate Gradient

- Method of steepest descent
  ⇒ Rewrite the problem as minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

  for $\mathbf{A} \in \mathbb{H}_n$ and $\mathbf{A} \succ \mathbf{0}$.
  ⇒ "walk" in the direction of the steepest descent
  ⇒ Very intuitive but – in general – not optimal
- Different idea:

# Conjugate Gradient

- Method of steepest descent
  $\Rightarrow$ Rewrite the problem as minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

  for $\mathbf{A} \in \mathbb{H}_n$ and $\mathbf{A} \succ \mathbf{0}$.
  $\Rightarrow$ "walk" in the direction of the steepest descent
  $\Rightarrow$ Very intuitive but – in general – not optimal
- Different idea:
  **A**-conjugate the gradient w.r.t. the previous search directions
  $\Rightarrow$ conjugate gradient (CG) method

# Conjugate Gradient

- Method of steepest descent
  $\Rightarrow$ Rewrite the problem as minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

  for $\mathbf{A} \in \mathbb{H}_n$ and $\mathbf{A} \succ \mathbf{0}$.
  $\Rightarrow$ "walk" in the direction of the steepest descent
  $\Rightarrow$ Very intuitive but – in general – not optimal
- Different idea:
  $\mathbf{A}$-conjugate the gradient w.r.t. the previous search directions
  $\Rightarrow$ conjugate gradient (CG) method
- CG is a Krylov subspace method, i.e., $\mathbf{x}_k \in \mathcal{K}_k$
- The search direction at k$th$ iteration is optimal

$$\|\mathbf{x}_* - \mathbf{x}_k\|_{\mathbf{A}} = \min_{\mathbf{x}\in\mathcal{K}_k} \|\mathbf{x}_* - \mathbf{x}\|_{\mathbf{A}}$$

  and $\|\mathbf{x}_* - \mathbf{x}_\ell\|_{\mathbf{A}} = 0$ for some $\ell \leq n$

# Generalized Minimal Residual Method

Question: What if $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a general matrix?

# Generalized Minimal Residual Method

- Following the idea of Krylov subspace methods we seek

$$\min_{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k(\mathbf{A},\mathbf{r}_0)} \|\mathbf{b}-\mathbf{A}\mathbf{x}\|$$

  where $\mathcal{K}_k(\mathbf{A},\mathbf{r}_0) = \text{span}(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, ..., \mathbf{A}^{k-1}\mathbf{r}_0)$
  (Note that if $\mathbf{x}_0 = \mathbf{0}$, we have $\mathbf{r}_0 = \mathbf{b}$)

- Assume we have an orthonormal basis $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ of $\mathcal{K}_k(\mathbf{A},\mathbf{r}_0)$. Then

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{V}_k\mathbf{y}$$

  for some $\mathbf{y}\in\mathbb{R}^k$, with $\mathbf{V}_k = [\mathbf{v}_1|...|\mathbf{v}_k]\in\mathbb{R}^{n\times k}$ and

$$\|\mathbf{b}-\mathbf{A}\mathbf{x}\| = \|\mathbf{b}-\mathbf{A}(\mathbf{x}_0+\mathbf{V}_k\mathbf{y})\| = \|\mathbf{r}_0-\mathbf{A}\mathbf{V}_k\mathbf{y}\|$$

  hence

$$\min_{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k(\mathbf{A},\mathbf{r}_0)} \|\mathbf{b}-\mathbf{A}\mathbf{x}\| = \min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{r}_0-\mathbf{A}\mathbf{V}_k\mathbf{y}\|$$

  $\Rightarrow$ Ordinary least-squares problem!

# Generalized Minimal Residual Method

- NOTE: This method needs an orthonormal basis of $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$
  $\Rightarrow$ Vulnerable to an imperfect basis caused by computational errors

Different approaches to computing this basis lead to different "flavors" of GMRES

# GMRES with Arnoldi (Gram-Schmidt)

- Arnoldi process:

  $\mathbf{r}_0 \leftarrow$ Initial residual

  $\mathbf{v}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|$

  $\mathbf{V}_1 = [\mathbf{v}_1]$

  For $p = 2$ to $k$:

  $$\mathbf{w}_p = (\mathbf{I} - \mathbf{V}_{p-1}\mathbf{V}_{p-1}^\top)\mathbf{A}\mathbf{v}_{p-1}$$
  $$\mathbf{v}_p = \mathbf{w}_p / \|\mathbf{w}_p\|$$
  $$\mathbf{V}_p = [\mathbf{V}_{p-1} | \mathbf{v}_p]$$

# GMRES with Arnoldi (Gram-Schmidt)

- This yields

$$\mathbf{AV}_k = \mathbf{V}_{k+1}\mathbf{H}_k$$

with

$$\mathbf{H}_k = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,k} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,k} \\ 0 & h_{3,2} & h_{3,3} & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & & h_{k,k-1} & h_{k,k} \\ 0 & & 0 & h_{k+1,k} \end{bmatrix}$$

# GMRES with Arnoldi (Gram-Schmidt)

- Note that

$$\mathbf{r}_0 = \|\mathbf{r}_0\|\mathbf{v}_1 =: \beta\mathbf{v}_1$$

- Hence

$$\|\mathbf{r}_0 - \mathbf{A}\mathbf{V}_k\mathbf{y}\| = \|\mathbf{r}_0 - \mathbf{V}_{k+1}\mathbf{H}_k\mathbf{y}\| = \|\mathbf{V}_{k+1}(\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y})\| = \|\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}\|$$

- Thus we seek to solve

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}\|$$

Using QR factorization of $\mathbf{H}_k = \mathbf{Q}_k\mathbf{R}_k$ we get

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{Q}_k(\beta\mathbf{Q}_k^\top\mathbf{e}_1 - \mathbf{R}_k\mathbf{y})\| = \min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{g}_k - \mathbf{R}_k\mathbf{y}\|$$

The minimum is obtained at $[\mathbf{R}_k\mathbf{y}]_{i=1}^k = [\mathbf{g}_k]_{i=1}^k$. Hence

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{g}_k - \mathbf{R}_k\mathbf{y}\| = [\mathbf{g}_k]_{k+1}$$

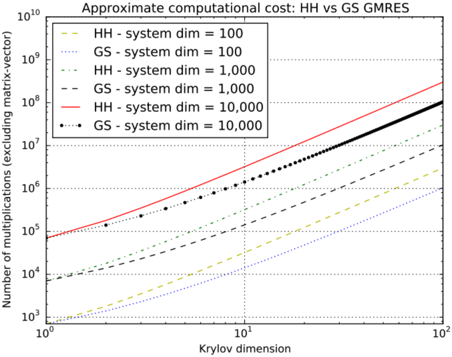
# Make GMRES more stable

- Gram-Schmidt may suffer from numerical instabilities!
- Alternatives exist:

  modified Gram Schmidt
  double Gram-Schmidt
  Given's rotations
  Householder

# Computational comparison

- N. I. Dravins, Numerical Implementations of the GMRES (2015):



# GMRES with Lanczos recurrence

- We assume $\mathbf{A} \in \mathbb{H}_n$
- Arnoldi process simplifies dramatically:
  In every iteration

$$\mathbf{w}_p = (\mathbf{I} - \mathbf{v}_{p-1}\mathbf{v}_{p-1}^\top - \mathbf{v}_{p-2}\mathbf{v}_{p-2}^\top)\mathbf{A}\mathbf{v}_{p-1}$$

- The Lanczos basis yields

$$\mathbf{V}_k^\top \mathbf{A}\mathbf{V}_k = \mathbf{J}_k$$

  where $\mathbf{J}_k$ is tri-diagonal

# GMRES with Chebyshev recurrence

- Sometimes orthogonalization is infeasible all together
- Assume $\Lambda(\mathbf{A}) \in [c \pm \delta_x, \pm\delta_y]$, and $\rho := \max\{\delta_x, \delta_y\}$
- Chebyshev recurrence:

  $\mathbf{v}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|$
  $\mathbf{v}_2 = (2\rho)^{-1}(\mathbf{A} - c\mathbf{I})\mathbf{v}_1$
  for $p = 3$ to $k$:
  $\quad \mathbf{v}_p = \frac{1}{\rho}\left((\mathbf{A} - c\mathbf{I})\mathbf{v}_{p-1} - \frac{\delta_x^2 - \delta_y^2}{4\rho}\mathbf{v}_{p-2}\right)$
  $\quad \mathbf{v}_{p-2} = \mathbf{v}_{p-2} / \|\mathbf{v}_{p-2}\|$
  $\mathbf{v}_{k-1} = \mathbf{v}_{k-1} / \|\mathbf{v}_{k-1}\|$
  $\mathbf{v}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$
- Good idea when dealing with large, sparse linear systems, where the coefficient matrix may be ill-conditioned
- Requires some estimate of the spectrum

# GMRES with sketching

- Remember, we basically want to solve

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{r}_0 - \mathbf{A}\mathbf{V}_k\mathbf{y}\|$$

  at every iteration

- As a least-squares problem it is a natural candidate for sketching

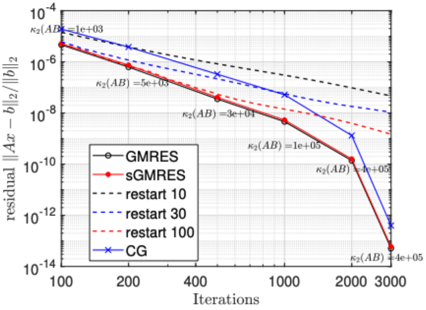
$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{S}(\mathbf{r}_0 - \mathbf{A}\mathbf{V}_k\mathbf{y})\|$$

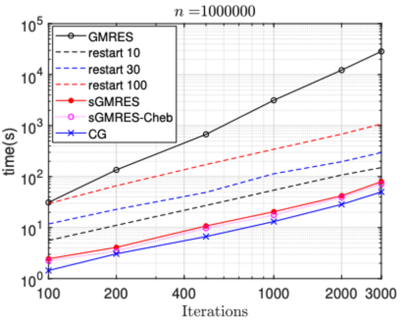
  where $\mathbf{S} \in \mathbb{R}^{d\times n}$ is the sketching matrix
  - Gaussian
  - SRTT
  - SSE

    ⋮

# Comparison sGMRES vs GMRES[1]

- Comparison of performance of MATLAB GMRES (with and without restarting) against the sGMRES algorithm (with 2-partial orthogonalization or the Chebyshev basis)
- Sparse linear system $\mathbf{Ax} = \mathbf{f}$, $\mathbf{A}$ is 2D Laplacian with $n = 10^6$.
- Left: Relative residual and condition number $\kappa_2(AB)$ of the reduced matrix associated with the $k$-truncated Arnoldi basis.
- Right: Total runtime including basis generation.




[1]Nakatsukasa and Tropp, arXiv:2111.00113

Course recap

# What we have seen ...

Randomized algorithms come in various flavors

- Random initial guess
- Application of random vectros/matrices
- Stochastic means to access quantities with high probability

# What we have seen ...

Integration

- Riemann sums (Left, Right, Upper, Lower, Middle)
- Trapezoidal rule
- Simpson
- Variational Monte Carlo
- Markov Chain Monte Carlo (MCMC)
  $\Rightarrow$ Multiple techniques to speed this up!

# What we have seen ...

Trace estimation

- Girard-Hutchinson
- Hutch$^{++}$
- XTrace
- XNysTrace
- Quantum Trace Estimation

# What we have seen ...

Sketching

- Gaussian Embedding
- Subsampled Randomized Trigonometric Transforms (SRTT)
- Sparse Sign Embeddings (SSE)
  - SSE with Cauchy random variables
  - SSE with exponential random variables
- Adaptive sampling

Application to least-squares problems

- Sketch-and-solve
- Iterative sketching

# What we have seen ...

Guest Lecture on sketching high-dimensional probability distributions

- Introduction to inherent high-dimensional problems
- Introduction to tensors
- Continuous sketching

# What we have seen ...

Matrix approximation (revisited)

- Low-rank approximation via: QR, CP-QR and ID/skeletonization
- SVD and rSVD
- Matrix Monte Carlo
- Matrix sketching

# What we have seen ...

Multi-linear algebra

- Detailed introduction to tensor spaces and tensor algebra
- Tensor diagrams
- Tensor product approximations
  - Canonical Polyadic (CP) Decomposition
  - Tucker Decomposition
  - Tensor Train (TT) Decomposition
  - Hierarchical Tucker (HT) Decomposition
- Generalizations of SVD
  - (T)HOSVD
  - S(T)HOSDV
  - r-STHOSVD
  - sketched-STHOSVD
  - sub-sketch-STHOSVD
- Sketching TT

# What we have seen ...

Eigenvalue problems

- Eigenvalue revealing decompositions
  - eigenvalue decomposition
  - unitary eigenvalue decomposition
  - Schur decomposition
- Two-step numerical procedure
  - Householder for upper Hessenbergform
  - Power iteration
  - Inverse iteration
  - “Pure” QR algorithm (without shift)
  - “Practical” QR algorithm (with shift)
- Krylov methods for eigenvalue problems
  - Rayleigh-Ritz (RR)/ Arnoldi method
  - sketched-RR

# What we have seen ...

Krylov methods for linear systems

- Steepest descent
- Conjugate Gradient (CG)
- Optimality of CG
- Generalized Minimal Residual Method (GMRES)
    - Arnoldi (Gram-Schmidt, Double Gram-Schmidt, Householder, ... )
    - Lanczos
    - Chebyshev recurrence
- sketched GMRES

# Conclusion

- Randomized techniques are a powerful tool for certain large problems in numerical linear algebra
- Randomized techniques are no silver bullet
- Randomized techniques often require careful implementation to be fully leveraged

Thank you for your attention