

Note: These are notes, not final report. The contents will definitely change over time. Partly still in Dutch, initially was only internally used as notes.

0.1 Introduction

Waarom überhaupt geïnteresseerd in Krylov methodes?

- schetsen waarom men geïnteresseerd zou moeten zijn in iteratieve methodes
- vergelijking direct/iteratief, "rule of thumb" dat het voor 2D ongeveer zelfde is, 3D zou het sneller moeten worden. Zie hiervoor bijv. boek wat mischa had gevonden

3 verschillende strategieën voor de Krylov methodes

- er zijn eigenlijk 3 keuzes wat men kan doen met het verkregen residu.
 1. zorgen dat het volgende residu zo klein mogelijk wordt met gegeven set aan vectoren (minres-strategie)
 2. zorgen dat het volgende residu orthogonaal is met een ruimte die steeds groter wordt (CG-strategie)
 3. Methodes waarbij het residu steeds ligt in een ruimte die kleiner wordt (IDR-strategie)
- voor/nadelen van elke methode. C
- [1]

Schets over convergentie idee 1

- Schets dat orthogonale methodes gegarandeerd convergeren in N iteraties indien alles wordt meegenomen (vgm alleen bij CG kan dit als A Hermitisch is).
- 1. Volgende residu is orthogonaal met $K(A, r_0)$
- 2. Voor elke nieuwe iteratie wordt dimensie van K een groter.
- 3. De dimensie van K is maximaal N .
- Therefore na N iteraties spant K dus de gehele ruimte op.
- Dan is er geen residu te vinden waarvoor het volgende orthogonaal is met K . I.e. het inproduct van r_{next} met alles in K is dus 0.
- Dit impliceert dat r_{next} geen componenten meer heeft in de gehele ruimte, en dus 0 moet zijn.

Schets over convergentie idee 2

- Schets dat GMRES (zonder restarts) convergeert in N iteraties
- 1. GMRES zoekt een lineaire combinatie van alle vectoren in K waarvoor volgende residu minimaal is.
- 2. Voor elke nieuwe iteratie wordt dimensie van K een groter.
- 3. De dimensie van K is maximaal N .
- Therefore na N iteraties spant K dus de gehele ruimte op.
- Als K de gehele ruimte opspant, dan is elk willekeurig residu uit te drukken in een lineaire combinatie van de vectoren in K .
- Therefore als $iter=N$ is het residu geleverd door GMRES 0.

0.2 Local Minimum Residual LMR

Simplest minimum residual method one could imagine

Dit komt uit het boek van optimal iteration methods for large linear systems of equations Sleijpen, van der Vorst

[7]. Linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (1)$$

Residual k

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \quad (2)$$

Goal is to approximate inverse of \mathbf{A} with a polynomial;

$$\mathbf{A}^{-1}\mathbf{b} \approx P(\mathbf{A})\mathbf{b}. \quad (3)$$

Krylov subspace of matrix \mathbf{A} and vector \mathbf{v} is the space $\mathcal{K}_k(\mathbf{A}, \mathbf{v})$;

$$\mathcal{K}_k(\mathbf{A}, \mathbf{v}) = \text{span} \{ \mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \dots, \mathbf{A}^{k-1}\mathbf{v} \}. \quad (4)$$

Here it can be seen that the polynomial that approximates the inverse $P(\mathbf{A})\mathbf{b}$ can be expressed as a linear combination of vectors in the Krylov subspace. However the question remains: Which coefficients should be used? This is where the methods differ.

One of the simplest Krylov subspace methods is the Local Minimum Residual (LMR) method. Which makes it nice for explaining. Starting with some initial guess \mathbf{x}_0 and corresponding residual \mathbf{r}_0 , this method seeks an update to \mathbf{x}_0 of the form:

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathbf{r}_0. \quad (5)$$

Where α is a coefficient that has to be determined at each iteration. This method chooses the value of α such that the next residual is minimized. Note that the next residual, \mathbf{r}_{i+1} , is given by

$$\mathbf{r}_{i+1} = \mathbf{b} - \mathbf{A}(\mathbf{x}_i + \alpha_i \mathbf{r}_i). \quad (6)$$

Which can be written as

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{A}\mathbf{r}_i, \quad (7)$$

or alternatively

$$\mathbf{r}_{i+1} = (\mathbf{I} - \alpha_i \mathbf{A})\mathbf{r}_i, \quad (8)$$

To obtain the value for α_i for which the next residual $\|\mathbf{r}_{i+1}\|_2$ is minimal a least squares problem is solved for α_i ;

$$\mathbf{0} = \mathbf{r}_i - \alpha_i \mathbf{A}\mathbf{r}_i. \quad (9)$$

This results in a recurrence relation

$$\alpha_i = \frac{\langle \mathbf{A}\mathbf{r}_i, \mathbf{r}_i \rangle}{\langle \mathbf{A}\mathbf{r}_i, \mathbf{A}\mathbf{r}_i \rangle}, \quad (10)$$

from which it can be deduced that the vector \mathbf{r}_i lies in the Krylov subspace $\mathcal{K}_{i+1}(\mathbf{A}, \mathbf{r}_0)$. Since $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{r}_i$ and \mathbf{x}_1 is given by equation 5, the vector \mathbf{x}_{i+1} exists in the affine space $\mathbf{x}_0 + \mathcal{K}_{i+1}(\mathbf{A}, \mathbf{r}_0)$.

- laten zien dat dit convergeert?
- men kan voorbeeld construeren waarvoor het niet convergeert (rotatiematrix 90 graden)
- voor algemene rotatiematrix kan men laten zien wat de update is dan
- BiCGStab gebruikt ook dit soort dingen op het einde, die heeft zelfde probleem
- als men bijv. 2 vectoren meeneemt is dit probleem weg. Dat is ook dan wat men in de polynomial stap doet bij bijv. idrstab en BiCGStab.

Resulting algorithm:

Algorithm 1 Basic LMR algorithm

```

1: function LMR( $\mathbf{A}, \mathbf{b}, \mathbf{x}, tol$ )
2:    $\mathbf{r} \leftarrow \mathbf{b} - \mathbf{A}\mathbf{x}$ 
3:   while  $\|\mathbf{r}\|_2 > tol$  do
4:      $\alpha_i \leftarrow \frac{\langle \mathbf{A}\mathbf{r}, \mathbf{r} \rangle}{\langle \mathbf{A}\mathbf{r}, \mathbf{A}\mathbf{r} \rangle}$ 
5:      $\mathbf{x} \leftarrow \mathbf{x} + \alpha \mathbf{r}$ 
6:      $\mathbf{r} \leftarrow \mathbf{r} - \alpha \mathbf{A}\mathbf{r}$ 
7:   end while
8:   return  $\mathbf{x}$ 
9: end function

```

0.3 Conjugate gradient (CG)

CG en familie komt uit het boek van Saad

[5]

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (11)$$

Here \mathbf{A} is symmetric, positive definite. The concept to be exploited in CG relies on a generalized notion of conjugate (orthogonality);

$$\mathbf{u}^T \mathbf{A}\mathbf{v} = 0. \quad (12)$$

Then an inner product can be defined

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{A}} := \langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{A}^T \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{A}\mathbf{v} \rangle = \mathbf{u}^T \mathbf{A}\mathbf{v} \quad (13)$$

0.4 BiCG

- uitbereiding om te zorgen dat CG ook werkt voor asymmetrisch

0.5 CGS

- problem is dat BiCG veel onnodig werk doet doordat het ook transpose probleem oplost
- Er schijnen ook scenarios te zijn waarin de transpose niet beschikbaar is, maar dat is volgens mij niet relevant voor de usecase hier.

0.6 BiCGStab

- problem is dat de BiCG methode niet echt smooth convergeert, maar het residu ook allerlei pieken kan hebben. Dit wilt men fixen door een GMRES stap op het einde toe te voegen, deze forceert dat het residu afneemt.
- CGS, maar dan met een GMRES(1) stap erbij.
- Deze GMRES(1) stap is eigenlijk precies LMR algorithm
- problem is wel als eigenwaardes sterk imaginair zijn, dan is inproduct in teller voor alpha ongeveer 0.
- Dan volgt dat je met de nieuwe vector in de krylov basis het residu nauwelijks kunt verkleinen; er staat namelijk niks in de richting van het residu.

0.7 BiCGStab(L)

[6]

- bicgstab moet de problemen met een 1e orde polynoom als "smoothing fix" oplossen door een hogere orde te pakken
- zo is de hoop dat het wel werkt met sterk imaginair
- is het mogelijk een matix te construeren waarvoor N stappen niet werken?

0.8 IDR(s)

[4]

- hier dan uitleggen wat het idee is achter het idr theorema, en waarom men dit kan gebruiken om een iterative solver te maken.
- Deze solver maakt gebruik van residu duwen in een ruimte die steeds kleiner word.

0.9 IDR(s)Stab(L)

[2][3]

- Verhaal van men begint met BiCG, CGS, BiCGStab, BiCGStab(L), IDR, idrstab
- idrstab combineert eigenlijk IDR(s) en BiCGStab(L). IDR is een ruimte die steeds kleiner word, BiCG dat r orthogonaal is met ruimte die steeds groter word. Zo word de oplossing van 2 kanten ingesloten in zekere zin.
- die van kensuke bleek allerlei voordelen te hebben zoals goede schatting van residu en zou sneller moeten zijn doordat het minder dots nodig heeft.
- uitleggen wat de "residual gap" is.

0.10 BiCGStab(L) implementation in Eigen

- Original idea, why should this implementation exist?
- Residual gaps & reliable updating strategies
- flying restarts
- argmin step
- breakdown (rotation matrix and example of the gmres case where no residual reduction until final iteration)
- BiCG convergence for trivial matrices
- preconditioner
- complex matrices
- Convex polynomial
- HouseholderQR
- Analytical, original version
- BiCGStab(L), BiCGStab comparison/(mathematical/numerical?) equivalence for L=1
- r0-rshadow orthogonality

0.11 Other definitions

0.11.1 Arnoldi iteration

- truc om een orthogonale krylov basis te construeren, Lanczos is vergelijkbaar, maar construeert een bi-orthogonale basis.

0.11.2 Inner product

A inner product is an operation for which:

$$\langle x, y \rangle = \overline{\langle y, x \rangle} \quad (14)$$

$$\langle ax, y \rangle = a \langle x, y \rangle \quad (15)$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \quad (16)$$

$$\langle x, x \rangle > 0 \quad (17)$$

$$\langle x, y \rangle = 0 \Leftrightarrow x = 0. \quad (18)$$

Bibliography

- [1] H. der Vorst. Iterative krylov methods for large linear systems. *Scientific Computing*, 2003.
- [2] M. v. G. G.L.Sleijpen. Exploiting bicgstab(ℓ) strategies to induce dimension reduction. *SIAM Journal on Scientific Computing*, 32:1035–1062, 2010.
- [3] E. I. Kensuke Aihara, Kuniyoshi Abe. A variant of idrstab with reliable update strategies for solving sparse linear systems. *Journal of Computational and Applied Mathematics*, 259:244–258, Mar 2014.
- [4] M. v. G. P. Sonneveld. Idr(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations. *SIAM Journal on Scientific Computing*, 31:1035–1062, 2008.
- [5] Y. Saad. *Iterative Methods for Sparse Linear Systems: Second Edition*. SIAM.
- [6] G. L. Sleijpen and D. R. Fokkema. Bicgstab (l) for linear equations involving unsymmetric matrices with complex spectrum. *Electronic Transactions on Numerical Analysis*, 1(11):2000, 1993.
- [7] G. L. G. Sleijpen. Optimal iteration methods for large linear systems of equations. In W. Hiller and E. Krausse, editors, *Workshop on Aspects of Parallelization regarding Finite Elements and Ocean Modelling*, volume 52 of *Berichte aus dem Fachbereich Physik*, Bremerhaven, 1994. Alfred Wegener Inst.

Todo list

Note: These are notes, not final report. The contents will definitely change over time. Partly still in Dutch, intially was only interally used as notes.	1
Simplest minimum residual method one could imagine	2
Dit komt uit het boek van optimal iteration methods for large linear systems of equations Sleijpen, van der Vorst	2
CG en familie komt uit het boek van Saad	3