A Block Wiedemann Rank Algorithm

William J. Turner

Department of Mathematics & Computer Science
Wabash College
Crawfordsville, IN 47933

ISSAC 2006 : Genova, Italy
July 2006

Outline

Minimal Generating Matrix Polynomial

The minimal generating matrix polynomial F for a linearly generated matrix sequence $\{B_i\}_{i=0}^{\infty} \in \left(\mathbb{F}^{\beta_l \times \beta_r}\right)^{\mathbb{Z}_{\geq 0}}$:

- Popov canonical form
- Columns form basis over $\mathbb{F}[\lambda]$ of the module of generating vector polynomials

Minimal Generating Matrix Polynomial

The minimal generating matrix polynomial F for a linearly generated matrix sequence $\{B_i\}_{i=0}^{\infty} \in (\mathbb{F}^{\beta_l \times \beta_r})^{\mathbb{Z}_{\geq 0}}$:

- Popov canonical form
- Columns form basis over $\mathbb{F}[\lambda]$ of the module of generating vector polynomials

I.e., if $C = \sum_{i=0}^{d} C_i \lambda^i \in \mathbb{F}^{\beta_r}[\lambda]$ is a column of F, then

$$\sum_{i=0}^{d} B_{i+j} C_i = 0^{\beta_l}, \qquad \forall j \geq 0.$$

Important Definitions

Let

- $H(\nu_l, \nu_r) = \begin{bmatrix} B_0 & B_1 & \cdots & B_{\nu_r-1} \\ B_1 & B_2 & \cdots & B_{\nu_r} \\ \vdots & \vdots & \ddots & \vdots \\ B_{\nu_l-1} & B_{\nu_l} & \cdots & B_{\nu_l+\nu_r-2} \end{bmatrix}$

- $\gamma_r = \deg(F)$

- γ_l be the smallest positive integer such that the block Hankel matrix $H(\gamma_l, \gamma_r + 1)$ has maximal rank

Modular Equivalence

Let

- $C = \sum_{i=0}^{d} C_i \lambda^i \in \mathbb{F}^{\beta_r}[\lambda]$
- $\nu_l \geq \gamma_l$ and $\nu_r \geq d$
- $\hat{C} = \mathrm{rev}_d(C)$

Modular Equivalence

Let

- $C = \sum_{i=0}^{d} C_i \lambda^i \in \mathbb{F}^{\beta_r}[\lambda]$
- $\nu_l \geq \gamma_l$ and $\nu_r \geq d$
- $\hat{C} = \mathrm{rev}_d(C)$

C generates $\{B_i\}_{i=0}^{\infty}$ if and only

$$\left(\sum_{i=0}^{\nu_l+\nu_r-1} B_i \lambda^i \right) \hat{C} \equiv C^{(\mathrm{res})} \pmod{\lambda^{\nu_l+\nu_r}}$$

for a vector polynomial $C^{(\mathrm{res})}$ of degree at most $d-1$.

Hermite-Padé Approximants

Let

- $\mathbf{P} = \begin{bmatrix} C^{(\text{res})} \\ \text{rev}_d(C) \end{bmatrix} \in \mathbb{F}^{\beta_l+\beta_r}[\lambda]$
- G_i be the ith column of

$$G = \begin{bmatrix} I & -(\sum_{i=0}^{\nu_l+\nu_r-1} B_i \lambda^i) \end{bmatrix} \in \mathbb{F}^{\beta_l \times (\beta_l+\beta_r)}[\lambda]$$

- $\mathbf{n} = (n_1, \ldots, n_m)$ where

$$n_i = \begin{cases} \nu_r - 1 & \text{if } 1 \le i \le \beta_l \\ \nu_r & \text{if } \beta_r + 1 \le i \le \beta_l + \beta_r \end{cases}$$

Hermite-Padé Approximants

P is vector Hermite-Padé approximant of type $(\mathbf{n}, \nu_l + \nu_r)$ of $\{G_i\}$ [Beckermann & Labahn, 1992, Def. 2.1].

Hermite-Padé Approximants

P is vector Hermite-Padé approximant of type $(\mathbf{n}, \nu_l + \nu_r)$ of $\{G_i\}$ [Beckermann & Labahn, 1992, Def. 2.1].

Equivalently, **P** is power Hermite-Padé approximant of type $(\mathbf{n}, (\nu_l + \nu_r)\beta_l, \beta_l)$ of $\{f_i\}$ where

$$f_i = (1, \lambda, \lambda^2, \ldots, \lambda^{s-1}) \cdot G_i(\lambda^s)$$

for $1 \leq i \leq m$ [Beckermann & Labahn, 1994, Ex. 2.5].

Hermite-Padé Approximants

Conversely, let

- $\mathbf{P} = (P_1, \ldots, P_{\beta_l+\beta_r})$ solve the power Hermite-Padé approximant problem
- $\mathrm{dct}(\mathbf{P}) = \min_{1 \le i \le \beta_l+\beta_r} \{n_i - \deg(P_i) + 1\}$
- $d = \nu_r + 1 - \mathrm{dct}(\mathbf{P})$
- $C^{(\mathrm{res})} = \begin{bmatrix} P_1 & \ldots & P_{\beta_l} \end{bmatrix}^{\mathsf{T}} \in \mathbb{F}^{\beta_l}[\lambda]$
- $\hat{C} = \begin{bmatrix} P_{\beta_l+1} & \ldots & P_{\beta_l+\beta_r} \end{bmatrix}^{\mathsf{T}} \in \mathbb{F}^{\beta_r}[\lambda]$

Hermite-Padé Approximants

Conversely, let

- $\mathbf{P} = (P_1, \ldots, P_{\beta_l+\beta_r})$ solve the power Hermite-Padé approximant problem
- $\mathrm{dct}(\mathbf{P}) = \min_{1 \le i \le \beta_l+\beta_r} \{n_i - \deg(P_i) + 1\}$
- $d = \nu_r + 1 - \mathrm{dct}(\mathbf{P})$
- $C^{(\mathrm{res})} = \begin{bmatrix} P_1 & \ldots & P_{\beta_l} \end{bmatrix}^{\mathsf{T}} \in \mathbb{F}^{\beta_l}[\lambda]$
- $\hat{C} = \begin{bmatrix} P_{\beta_l+1} & \ldots & P_{\beta_l+\beta_r} \end{bmatrix}^{\mathsf{T}} \in \mathbb{F}^{\beta_r}[\lambda]$

Then $C = \mathrm{rev}_d(\hat{C})$ generates $\{B_i\}_{i=0}^{\infty}$.

Computing Minimal Generating Matrix Polynomial

- Use Beckerman-Labahn Fast Power Hermite-Padé Solver (FPHPS) to compute basis for **P**
- Exactly β_r of the polynomial tuples will have positive defect
- These β_r polynomial tuples give a basis over $\mathbb{F}[\lambda]$ of the module of generating vector polynomials
- Convert matrix with these columns to Popov form : F

Kaltofen-Saunders Rank Algorithm

Precondition A so with high probability

- minimal polynomial $f^{\tilde{A}} = \lambda f(\lambda)$
- $\mathrm{rank}(A) = \deg(f)$
- $f(0) \neq 0$

Kaltofen-Saunders Rank Algorithm

Precondition A so with high probability

- minimal polynomial $f^{\tilde{A}} = \lambda f(\lambda)$
- $\text{rank}(A) = \deg(f)$
- $f(0) \neq 0$

The algorithm returns

$$\text{rank}(A) = \deg(f_u^{\tilde{A},v}) - 1 = \deg(f_u^{\tilde{A},v}) - \text{codeg}(f_u^{\tilde{A},v}).$$

Converting to Block Version

Kaltofen-Saunders: relies on $f_u^{A,v} = f^A = s_n(\lambda I - A)$ with probability at least $1 - 2\nu/|S| \geq 1 - 2n/|S|$

Converting to Block Version

Kaltofen-Saunders: relies on $f_u^{A,v} = f^A = s_n(\lambda I - A)$ with probability at least $1 - 2\nu/|S| \geq 1 - 2n/|S|$

Block version:

$$s_{\beta_r - i}(F_X^{A,Y}) = s_{n-i}(\lambda I - A), \qquad 0 \leq i \leq \beta_r - 1,$$

with probability at least $1 - 2\nu/|S| \geq 1 - 2n/|S|$

Converting to Block Version

Kaltofen-Saunders: relies on $f_u^{A,v} = f^A = s_n(\lambda I - A)$ with probability at least $1 - 2\nu/|S| \geq 1 - 2n/|S|$

Block version:

$$s_{\beta_r - i}(F_X^{A,Y}) = s_{n-i}(\lambda I - A), \qquad 0 \leq i \leq \beta_r - 1,$$

with probability at least $1 - 2\nu/|S| \geq 1 - 2n/|S|$

Proof uses Schwartz-Zippel lemma and bounds

$$\text{rank}(H_X^{A,Y}(\nu_l, \nu_r + 1)) = \det(\deg(F_X^{A,Y}))$$

Block Rank

Theorem

Let $\mathbb{F}$ be a field, S be a finite subset of $\mathbb{F}$, $A \in \mathbb{F}^{n \times n}$, and $1 \leq \beta_r \leq \beta_l \leq n$. Let $X \in S^{n \times \beta_l}$ and $Y \in S^{n \times \beta_r}$ be matrices whose entries are chosen uniformly and independently from S, and let $D = \mathrm{diag}(d_1, \ldots, d_n)$ where $d_1, \ldots, d_n$ are chosen uniformly and independently from S. Let A have a nonzero $r \times r$ principal minor. Then, the difference $\deg(\det(F_X^{AD,Y})) - \mathrm{codeg}(\det(F_X^{AD,Y}))$ equals the rank of A, and we can compute $F_X^{AD,Y}$ from the first $\lceil n/\beta_l \rceil + \lceil n/\beta_r \rceil$ matrices in the block Wiedemann sequence $\{X^T (AD)^i Y\}_{i=0}^{\infty}$ with probability at least $1 - n(n+3)/(2|S|)$.

Block Rank Algorithm

1. $B_1, B_2 \leftarrow$ butterfly network preconditioners with parameters chosen uniformly and independently from S
2. $D \leftarrow \text{diag}(d_1, \ldots, d_n)$, $d_1, \ldots, d_n$ chosen uniformly and independently from S
3. $\tilde{A} \leftarrow B_1^T A B_2 D$
   {Implement via black box model}
4. Choose $X \in S^{n \times \beta_l}$ and $Y \in S^{n \times \beta_r}$ uniformly and independently
5. $\nu_l \leftarrow \lceil n/\beta_l \rceil$ and $\nu_r \leftarrow \lceil n/\beta_r \rceil$
6. Compute $F_X^{\tilde{A},Y}$ from $\{X^T \tilde{A}^i Y\}_{i=0}^{\nu_l+\nu_r-1}$
   {Possibly returning failure}
7. $r \leftarrow \deg(\det(F_X^{\tilde{A},Y})) - \text{codeg}(\det(F_X^{\tilde{A},Y}))$

Block vs. Kaltofen-Saunders

Block algorithm:

- at most $2\frac{n\lceil \log_2(n)\rceil}{2} + n + \beta_l n + \beta_r n = n\left(\beta_l + \beta_r + 1 + \lceil \log_2(n)\rceil\right)$ random elements from S
- Returns the correct rank with probability at least $\left(1 - \frac{2r\lceil \log_2(n)\rceil}{|S|}\right)\left(1 - \frac{n(n+3)}{2|S|}\right) \geq 1 - \frac{n(n+3+4\lceil \log_2(n)\rceil)}{2|S|}$

Block vs. Kaltofen-Saunders

Block algorithm:

- at most $2\frac{n\lceil\log_2(n)\rceil}{2} + n + \beta_l n + \beta_r n = n\left(\beta_l + \beta_r + 1 + \lceil\log_2(n)\rceil\right)$ random elements from S
- Returns the correct rank with probability at least $\left(1 - \frac{2r\lceil\log_2(n)\rceil}{|S|}\right)\left(1 - \frac{n(n+3)}{2|S|}\right) \geq 1 - \frac{n(n+3+4\lceil\log_2(n)\rceil)}{2|S|}$

Kaltofen-Saunders algorithm:

- No more than $2\frac{n\lceil\log_2(n)\rceil}{2} + 3n = n\left(3 + \lceil\log_2(n)\rceil\right)$ random elements from S
- Returns the correct rank with probability at least $1 - \frac{4\deg(f^{\tilde{A}})+r(r+1)+4r\lceil\log_2(n)\rceil)}{2|S|} \geq 1 - \frac{n(n+3+4\lceil\log_2(n)\rceil)}{2|S|}$

Other Advantages

- Parallel algorithm [Coppersmith, 1994; Kaltofen, 1995; Villard, 2000]
- Captures more than just the largest invariant factor $\lambda I - A$
- Allows use of different blocking factors (*i.e.*, $\beta_r < \beta_l$)

Other Advantages

- Parallel algorithm [Coppersmith, 1994; Kaltofen, 1995; Villard, 2000]
- Captures more than just the largest invariant factor $\lambda I - A$
- Allows use of different blocking factors (*i.e.*, $\beta_r < \beta_l$)
  - This can reduce the number of black box operations that various computations require [Kaltofen, 1995].

Bound on Rank

Theorem

Let $\mathbb{F}$ be a field, $A \in \mathbb{F}^{n \times n}$ have rank r, and $1 \leq \beta_r \leq \beta_l \leq n$. Let $X \in \mathbb{F}^{n \times \beta_l}$, $Y \in \mathbb{F}^{n \times \beta_r}$, and $D = \mathrm{diag}(d_1, \ldots, d_n) \in \mathbb{F}^{n \times n}$. Then, the rank of A is bounded from below by

$$\begin{aligned} r &\geq \deg(\det(\lambda I - AD)) - \mathrm{codeg}(\det(\lambda I - AD)) \\ &\geq \deg(\det(F_X^{AD,Y})) - \mathrm{codeg}(\det(F_X^{AD,Y})). \end{aligned}$$

Future Work

- New preconditioners
  - Take advantage of capturing more than one invariant factor
- Detailed comparison with Eberly's block Lanczos method [Eberly, 2004]

References

B. Beckermann & G. Labahn (1992). A Uniform Approach for Hermite Padé and Simultaneous Padé Approximants and Their Matrix-Type Generalizations. *Numer. Algorithms*, 3:45–54.

— (1994). A Uniform Approach for the Fast Computation of Matrix-Type Padé Approximants. *SIAM J. Matrix Anal. Appl.*, 15(3):804–823.

D. Coppersmith (1994). Solving Homogeneous Linear Equations Over GF(2) via Block Wiedemann Algorithm. *Math. Comp.*, 62:333–350.

W. Eberly (2004). Reliable Krylov-Based Algorithms for Matrix Null Space and Rank (Extended Abstract). In J. Gutierrez, ed., *Proc. ISSAC 2004*, pp. 127–134. ACM Press.

E. Kaltofen (1995). Analysis of Coppersmith's Block Wiedemann Algorithm for the Parallel Solution of Sparse Linear Systems. *Math. Comp.*, 64(210):777–806.

G. Villard (2000). Processor Efficient Parallel Solution of Linear Systems of Equations. *J. Algorithms*, 35(1):122–126.