Term Project Presentation

A Comparison of Algorithms Related to Trace Minimization to Compute a Small Number of Eigenvalues of a Real Symmetric Matrix

Giuseppe Accaputo, 27.01.2017

Outline

- Tasks
- Trace Theorem and a Proof
- Trace Minimization Method
- Trace Minimization Method as a Quasi-Newton Method
- Jacobi-Davidson / Newton with Subspace Acceleration
- Davidson-Type Trace Minimization Method
- Conclusion

Tasks

1. Detailed formulation of the problem, including proof of the trace theorem
2. Formulation of the trace minimization method as a Newton method
3. Derive TraceMin and Jacobi-Davidson algorithms and compare them

Problem Description

Compute a few of the smalles eigenvalues or eigenvectors of the large, sparse, generalized eigenvalue problem

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{B}\boldsymbol{x} \quad , \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n, \ \lambda \in \mathbb{R}$ and $\boldsymbol{A}, \boldsymbol{B}$ are $n \times n$ symmetric matrices, with $\boldsymbol{B}$ being positive-definite.

- The matrix $\boldsymbol{A} - \lambda \boldsymbol{B}$ is called a matrix *pencil*

Theorem 1 [5]

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric $n \times n$ matrices. If $\boldsymbol{B}$ is positive-definite then there is an $n \times n$ matrix $\boldsymbol{Z}$ for which

$$\boldsymbol{Z}^T \boldsymbol{B} \boldsymbol{Z} = \boldsymbol{I}_n \quad \text{and} \quad \boldsymbol{Z}^T \boldsymbol{A} \boldsymbol{Z} = \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n), \tag{2}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the pencil $(\boldsymbol{A}, \boldsymbol{B})$ from problem (1) and the columns of $\boldsymbol{Z}$ are their associated eigenvectors. Furthermore, if $\boldsymbol{A}$ is positive-definite, then all of the eigenvalues λ_i are positive.

Theorem 2: The Trace Theorem [8]

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be given as in Theorem 1 and $\mathcal{Y}^*$ be the set of all $n \times p$ matrices $\boldsymbol{Y}$ for which $\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{Y} = \boldsymbol{I}_p$. Then

$$\min_{\boldsymbol{Y}\in\mathcal{Y}^*} \mathrm{tr}(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}) = \sum_{i=1}^{p} \lambda_i \,. \tag{3}$$

In other words,

$$\min_{\boldsymbol{Y}\in\mathcal{Y}^*} \mathrm{tr}(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) \tag{4}$$

with

$$\boldsymbol{X}^T\boldsymbol{B}\boldsymbol{X} = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)\,, \tag{5}$$

where $\boldsymbol{X}$ corresponds to the first p columns of the matrix $\boldsymbol{Z}$ of Theorem 1.

Proof of the Trace Theorem (1/6)

Theorem 3 *(Poincaré Separation Theorem [4, 6])*

Let $\boldsymbol{A}$ be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and let $\boldsymbol{G}$ be a semi-unitary $n \times k$ matrix $(1 \leq k \leq n)$, so that $\boldsymbol{G}^T\boldsymbol{G} = \boldsymbol{I}_k$. Then the eigenvalues $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$ of $\boldsymbol{G}^T\boldsymbol{A}\boldsymbol{G}$ satisfy

$$\lambda_i \leq \mu_i \leq \lambda_{n-k+i} \qquad (i = 1, 2, \ldots, k)\,. \tag{6}$$

Proof of the Trace Theorem (2/6)

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ given as in Theorem 1, i.e., $\boldsymbol{Z} \in \mathbb{R}^{n\times n}$ is the matrix for which $\boldsymbol{Z}^T\boldsymbol{B}\boldsymbol{Z} = \boldsymbol{I}_n$ and $\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{Z} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the pencil $(\boldsymbol{A}, \boldsymbol{B})$.

Let $\boldsymbol{Y} \in \mathcal{Y}^*$ and set $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{G}$ for some $\boldsymbol{G} \in \mathbb{R}^{n\times p}$. $\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{Y} = \boldsymbol{I}_p \implies \boldsymbol{G}$ is unitary and

$$\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{G}^T\boldsymbol{\Lambda}\,\boldsymbol{G}\,. \tag{7}$$

Proof of the Trace Theorem (3/6)

From Theorem 3, for the eigenvalues μ_i of $\boldsymbol{G}^T \boldsymbol{\Lambda} \, \boldsymbol{G}$ it follows that $\lambda_i \leq \mu_i$ for $i = 1, \ldots, p$ and thus

$$\sum_{i=1}^{p} \lambda_i \leq \sum_{i=1}^{p} \mu_i \, . \tag{8}$$

Proof of the Trace Theorem (4/6)

$\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G}$ is symmetric $\implies$ there exists a spectral decomposition [10, Theorem 4.33] of the form

$$\boldsymbol{Q}^T (\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G}) \boldsymbol{Q} = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_p), \tag{9}$$

where $\boldsymbol{Q}$ is a unitary matrix with columns $\boldsymbol{q}_i$ being the eigenvectors of $\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G}$.

Proof of the Trace Theorem (5/6)

Consider the trace of the spectral decomposition in Eq. (9):

$$\operatorname{tr}(\boldsymbol{Q}^T (\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G})\, \boldsymbol{Q}) = \operatorname{tr}(\boldsymbol{Q}\boldsymbol{Q}^T (\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G})) = \operatorname{tr}(\boldsymbol{G}^T \boldsymbol{\Lambda} \boldsymbol{G}) = \sum_{i=1}^{p} \mu_i \,. \quad (10)$$

From Eqs. (7), (8) and (10) it follows that

$$\sum_{i=1}^{p} \lambda_i \leq \operatorname{tr}(\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y})\,. \quad (11)$$

Proof of the Trace Theorem (6/6)

By the the spectral decomposition [10, Theorem 4.33] equality holds if $\boldsymbol{Y} = \boldsymbol{Z}_p = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_p]$, where the columns $\boldsymbol{z}_i$ are the eigenvectors of the pencil $(\boldsymbol{A}, \boldsymbol{B})$.

$\boldsymbol{Z}_p$ hence diagonalizes the matrix $\boldsymbol{A}$ from problem (1) and thus leads to

$$\boldsymbol{Z}_p^T \boldsymbol{A} \boldsymbol{Z}_p = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \, .$$

□

The Trace Minimization Method [8, 7]

Trace Minimization (TRACEMIN): Use trace theorem (Theorem 2) and treat problem (1) as the quadratic minimization problem

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}(\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y}) \\ \text{subject to} \quad & \boldsymbol{Y}^T \boldsymbol{B} \boldsymbol{Y} = \boldsymbol{I}_p \,. \end{aligned} \tag{12}$$

The Trace Minimization Method [8, 7]

Idea is to compute a correction term $\boldsymbol{\Delta}_k$ that is chosen as to

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}((\boldsymbol{Y}_k - \boldsymbol{\Delta}_k)^T \boldsymbol{A} (\boldsymbol{Y}_k - \boldsymbol{\Delta}_k)) \\ \text{subject to} \quad & \boldsymbol{Y}_k^T \boldsymbol{B} \, \boldsymbol{\Delta}_k = \boldsymbol{0} \, . \end{aligned} \tag{13}$$

The Trace Minimization Method [8, 7]

Next iterate $\boldsymbol{Y}_{k+1}$ is formed by B-orthonormalizing $\boldsymbol{Y}_k - \boldsymbol{\Delta}_k$. By also enforcing $\boldsymbol{Y}_k^T \boldsymbol{B} \boldsymbol{\Delta}_k = \boldsymbol{0}$ in the minimization problem (13) it guarantees that

$$\operatorname{tr}(\boldsymbol{Y}_{k+1}^T \boldsymbol{A} \boldsymbol{Y}_{k+1}) \leq \operatorname{tr}((\boldsymbol{Y}_k - \boldsymbol{\Delta}_k)^T \boldsymbol{A} (\boldsymbol{Y}_k - \boldsymbol{\Delta}_k)) \leq \operatorname{tr}(\boldsymbol{Y}_k^T \boldsymbol{A} \boldsymbol{Y}_k) \,. \tag{14}$$

The Trace Minimization Method [8, 7]

Solution of the minimization problem (13) can be obtained by solving the saddle-point problem

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}\boldsymbol{Y}_k \\ \boldsymbol{Y}_k^T \boldsymbol{B} & 0 \end{pmatrix} \begin{pmatrix} \Delta_k \\ \boldsymbol{L}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}\boldsymbol{Y}_k \\ 0 \end{pmatrix} \tag{15}$$

where $\boldsymbol{L}_k$ represents the Lagrange multipliers

The Trace Minimization Method [8, 7]

The saddle-point problem is further reduced to the following positive-semidefinite system

$$(\boldsymbol{PAP})\boldsymbol{\Delta}_k = \boldsymbol{PAY}_k, \qquad \boldsymbol{Y}_k^T\boldsymbol{B}\boldsymbol{\Delta}_k = \boldsymbol{0} \tag{16}$$

where

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{BY}_k(\boldsymbol{Y}_k^T\boldsymbol{B}^2\boldsymbol{Y}_k)^{-1}\boldsymbol{Y}_k^T\boldsymbol{B} \tag{17}$$

is the orthogonal projector onto the space B-orthogonal to $\boldsymbol{Y}_k$, which guarantees that $\boldsymbol{Y}_k^T\boldsymbol{B}\boldsymbol{\Delta}_k = \boldsymbol{0}$.

The Trace Minimization Method [8, 7]

If the projected system in Eq. (16) is solved exactly at each iteration step, TRACEMIN is mathematically equiv. to inverse iteration.

- Inherits robust global convergence property
- Also inherits linear convergence rate
  - TRACEMIN can be accelerated by using shifting strategies

Trace Minimization Method as a Quasi-Newton Method

Newton's method: Solve

$$F(\boldsymbol{x}) = \boldsymbol{0}\,. \tag{18}$$

Newton step: Use $F(\boldsymbol{x}) = \operatorname{grad} f(\boldsymbol{x})$, then:

$$\boldsymbol{p}_k = -\operatorname{Hess}_f(\boldsymbol{x}_k)^{-1} \; \operatorname{grad} f(\boldsymbol{x}_k)\,. \tag{19}$$

Quasi-Newton step:

$$\boldsymbol{p}_k = -\boldsymbol{B}_k^{-1} \; \operatorname{grad} f(\boldsymbol{x}_k)\,, \tag{20}$$

with $\boldsymbol{B}_k$ being an approximation of the true Hessian $\operatorname{Hess}_f(\boldsymbol{x}_k)$.

Trace Minimization Method as a Quasi-Newton Method

TRACEMIN's objective function is given by

$$f : \mathbb{R}_*^{n\times p} \to \mathbb{R} : \boldsymbol{Y} \mapsto \mathrm{tr}((\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{Y})^{-1}(\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}))\,, \tag{21}$$

where $\mathbb{R}_*^{n\times p}$ denotes the set of full-rank $n \times p$ matrices.

Trace Minimization Method as a Quasi-Newton Method

A second-order expansion of f around $\boldsymbol{\Delta}_k = \mathbf{0}$ gives:

$$
\begin{aligned}
f(\boldsymbol{Y}_k + \boldsymbol{\Delta}_k) = {} & \operatorname{tr}((\boldsymbol{Y}_k^T \boldsymbol{B} \boldsymbol{Y}_k)^{-1} (\boldsymbol{Y}_k^T \boldsymbol{A} \boldsymbol{Y}_k)) \\
& + \operatorname{tr}((\boldsymbol{Y}_k^T \boldsymbol{B} \boldsymbol{Y}_k)^{-1} \boldsymbol{\Delta}_k^T 2 \boldsymbol{A} \boldsymbol{Y}_k) \\
& + \frac{1}{2} \operatorname{tr}((\boldsymbol{Y}_k^T \boldsymbol{B} \boldsymbol{Y}_k)^{-1} \boldsymbol{\Delta}_k^T 2 (\boldsymbol{A} \boldsymbol{\Delta}_k \\
& - \boldsymbol{B} \boldsymbol{\Delta}_k (\boldsymbol{Y}_k^T \boldsymbol{B} \boldsymbol{Y}_k)^{-1} \boldsymbol{Y}_k^T \boldsymbol{A} \boldsymbol{Y}_k)) + H.O.T.
\end{aligned} \tag{22}
$$

Trace Minimization Method as a Quasi-Newton Method

Introduce $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{B}\boldsymbol{Y}_k(\boldsymbol{Y}_k^T\boldsymbol{B}^2\boldsymbol{Y}_k)^{-1}\boldsymbol{Y}_k^T\boldsymbol{B}$ as the orthogonal projector onto the space B-orthogonal to $\boldsymbol{Y}_k$

Further, introduce the inner product [1, 3]

$$\langle \boldsymbol{Z}_1, \boldsymbol{Z}_2 \rangle := \operatorname{tr}((\boldsymbol{Y}_k^T\boldsymbol{B}\boldsymbol{Y}_k)^{-1}\boldsymbol{Z}_1^T\boldsymbol{Z}_2), \quad \boldsymbol{Z}_1, \boldsymbol{Z}_2 \ B\text{-orthogonal to } \boldsymbol{Y}_k \, . \tag{23}$$

Trace Minimization Method as a Quasi-Newton Method

Now rewrite second-order expansion as

$$f(\boldsymbol{Y}_k + \boldsymbol{\Delta}_k) = f(\boldsymbol{Y}_k) + \langle \boldsymbol{\Delta}_k, 2\boldsymbol{PAY} \rangle + \frac{1}{2} \langle \boldsymbol{\Delta}_k, 2\boldsymbol{P}(\boldsymbol{A\Delta}_k - \boldsymbol{B\Delta}_k(\boldsymbol{Y}_k^T \boldsymbol{BY}_k)^{-1} \boldsymbol{Y}_k^T \boldsymbol{AY}_k)\rangle + H.O.T. \tag{24}$$

Identify $2\boldsymbol{PAY}$ to be the gradient of f at $\boldsymbol{\Delta}_k = \mathbf{0}$ and the operator

$$\mathrm{Hess}_f : \boldsymbol{\Delta}_k \mapsto 2\boldsymbol{P}(\boldsymbol{A\Delta}_k - \boldsymbol{B\Delta}_k(\boldsymbol{Y}_k^T \boldsymbol{BY}_k)^{-1} \boldsymbol{Y}_k^T \boldsymbol{AY}_k) \tag{25}$$

to be the Hessian of f at $\boldsymbol{\Delta}_k = \mathbf{0}$ [1, 3]

Trace Minimization Method as a Quasi-Newton Method

Newton correction equation now becomes

$$\boldsymbol{P}(\boldsymbol{A}\boldsymbol{\Delta}_k - \boldsymbol{B}\boldsymbol{\Delta}_k(\boldsymbol{Y}_k^T\boldsymbol{B}\boldsymbol{Y}_k)^{-1}\boldsymbol{Y}_k^T\boldsymbol{A}\boldsymbol{Y}_k) = -\boldsymbol{P}\boldsymbol{A}\boldsymbol{Y}\,. \tag{26}$$

Substitute the Hessian of f with the approximate Hessian $2\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}$ and the correction equation becomes

$$(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P})\boldsymbol{\Delta}_k = -\boldsymbol{P}\boldsymbol{A}\boldsymbol{Y}, \quad \boldsymbol{Y}_k^T\boldsymbol{B}\boldsymbol{\Delta}_k = \boldsymbol{0}\,, \tag{27}$$

which is the same as Eq. (16) solved in the TRACEMIN method [3, § 4.3.2].

Trace Minimization Method as a Quasi-Newton Method

Important to mention: further calculations needed to capture TRACEMIN's global convergence theory; see [3, § 4.3.2] for further details.

Nonetheless, TRACEMIN can be described as inexact, quasi-Newton method.

- Method yields linear (instead of quadratic) convergence rate due to the usage of approx. Hessian
- Authors of TRACEMIN knew this result due to relationship between TRACEMIN and inverse iteration

Trace Minimization Method as a Quasi-Newton Method

In [1], the authors present a two-phase algorithm using a Riemannian trust-region algorithm:

1. Use basic TRACEMIN far away from solution, i.e. use approximate Hessian
2. When a switching criterion is satisfied (i.e. algorithm is close to solution), continue calculations with exact Hessian

Result: superlinear convergence.

Jacobi-Davidson [9, 2]

The Jacobi-Davidson (JD) method calculates the eigenvectors and eigenvalues of the pencil $(\boldsymbol{A}, \boldsymbol{B})$ by constructing a correction, for a given eigenvector approximation, in a subspace orthogonal to the given approximation.

Name follows from two basic principles:

1. Jacobi's idea: compute orthogonal corrections
2. Davidson approach: Computation of the correction in a given subspace different from Krylov subspaces

Jacobi-Davidson: Newton with Subspace Acceleration [9, 2]

Derive a Newton update from the generalized Rayleigh quotient

$$\rho(\boldsymbol{x}) = \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}}\,, \quad \forall \boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}\,. \tag{28}$$

The Newton equation $\operatorname{Hess}_\rho(\boldsymbol{x}_k)\boldsymbol{t}_k = -\operatorname{grad}\rho(\boldsymbol{x}_k)$ becomes

$$\left(\boldsymbol{I} - \frac{2}{\boldsymbol{x}_k^T \boldsymbol{B} \boldsymbol{x}_k} \boldsymbol{B} \boldsymbol{x}_k \boldsymbol{x}_k^T\right)(\boldsymbol{A} - \rho(\boldsymbol{x}_k)\boldsymbol{B})\left(\boldsymbol{I} - \frac{2}{\boldsymbol{x}_k^T \boldsymbol{B} \boldsymbol{x}_k} \boldsymbol{x}_k \boldsymbol{x}_k^T \boldsymbol{B}\right) = -(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{B}\boldsymbol{x}_k \rho(\boldsymbol{x}_k)). \tag{29}$$

Problem: Hessian is always singular when $\boldsymbol{x}$ is an eigenvector, because then $\operatorname{Hess}_\rho(\boldsymbol{x})\boldsymbol{x} = -F(\boldsymbol{x}) = \boldsymbol{0}$ [2]

Jacobi-Davidson: Newton with Subspace Acceleration [9, 2]

Apply the Newton method instead to

$$F(\boldsymbol{x}, \lambda) := \begin{pmatrix} (\boldsymbol{A} - \lambda \boldsymbol{B})\boldsymbol{x} \\ \boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x} - 1 \end{pmatrix}. \tag{30}$$

New Newton step:

$$\begin{pmatrix} \boldsymbol{A} - \lambda_k \boldsymbol{B} & \boldsymbol{B}\boldsymbol{x}_k \\ \boldsymbol{x}_k^T \boldsymbol{B} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{t}_k \\ \epsilon_k \end{pmatrix} = \begin{pmatrix} -\boldsymbol{r}_k \\ 0 \end{pmatrix}, \tag{31}$$

Note: Is only singular if $(\boldsymbol{x}_k, \lambda_k)$ is an eigenpair of the pencil $(\boldsymbol{A}, \boldsymbol{B})$ with λ_k being a multiple eigenvalue [2]

Jacobi-Davidson: Newton with Subspace Acceleration [9, 2]

Davidson's approach: consecutive corrections $\boldsymbol{t}_k$ are now used to build the search space.

- Solution $\boldsymbol{t}_k$ of the Jacobian correction equation is appended to $\boldsymbol{V}_k$, resulting in $\boldsymbol{V}_{k+1} = [\boldsymbol{V}_k, \boldsymbol{t}_k]$
- Speeds up convergence by increasing dimension of trial space by one

Block JD approx. l eigenvalues simultaneously. Further, the trial space dimension is increased by l.

Davidson-Type Trace Minimization Method [7]

Problems of block JD:

1. Shifting strategy forces algorithm to converge to eigenvalues closest to Ritz values (often far away from desired eigenvalues at the beginning)
2. Subspace expanding decreases Ritz values; block JD is forced to converge to smalles eigenpairs
3. Ill-conditioning when Ritz value approaches multiple eigenvalue or cluster of eigenvalues

Davidson-Type Trace Minimization Method [7]

Solution:

1. Use multiple dynamic shifting strategy
2. Use implicit deflation technique ($\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{d}_{k,i} = 0$)
3. Use dynamic stopping strategy for accuracy when solving inner system

Conclusion

- Proof of trace theorem using Poincaré separation theorem
- Trace minimization characterization as quasi-Newton more difficult than expected (requires background in differential geometry)
- Block JD and TRACEMIN are quite similar
- Although JD has better convergence rate in some cases, it still depends on a good starting subspace
  - Davidson-type TRACEMIN is not affected; is more robust

References

[1] P-A Absil, Christopher G Baker, Kyle A Gallivan, and Ahmed Sameh. Adaptive model trust region methods for generalized eigenvalue problems. In *International Conference on Computational Science*, pages 33–41. Springer, 2005.

[2] Peter Arbenz and Richard B Lehoucq. A comparison of algorithms for modal analysis in the absence of a sparse direct method. 2003.

[3] Christopher G Baker. *Riemannian manifold trust-region methods with applications to eigenproblems*. ProQuest, 2008.

[4] Richards Bellman. *Introduction to matrix analysis*. Society for Industrial and Applied Mathematics, 1970.

[5] Joel N Franklin. *Matrix theory*. Courier Corporation, 2012.

[6] Jan R Magnus, Heinz Neudecker, et al. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 1995.

[7] Ahmed Sameh and Zhanye Tong. The trace minimization method for the symmetric generalized eigenvalue problem. *Journal of Computational and Applied Mathematics*, 123(1):155–175, 2000.

[8] Ahmed H Sameh and John A Wisniewski. A trace minimization algorithm for the generalized eigenvalue problem. *SIAM Journal on Numerical Analysis*, 19(6):1243–1259, 1982.

[9] Gerard LG Sleijpen, Albert GL Booten, Diederik R Fokkema, and Henk A Van der Vorst. Jacobi-davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT Numerical Mathematics*, 36(3):595–633, 1996.

[10] G.W. Stewart. *Matrix algorithms: volume 1: basic decompositions*. Society for Industrial and Applied Mathematics, 1998.