Unconstrained Optimization Models for Computing Several Extreme Eigenpairs of Real Symmetric Matrices

Yu-Hong Dai

LSEC, ICMSEC

Academy of Mathematics and Systems Science

Chinese Academy of Sciences

Joint work with Bo Jiang and Chun-Feng Cui

Email: dyh@lsec.cc.ac.cn

Peking University, December 22, 2013

Outline

1. Unconstrained optimization and eigenvalue computing
2. Applications of several extreme eigenpairs
3. Variational principles for computing extreme eigenpairs
   - Block unconstrained quartic model
   - Block unconstrained β-order model
   - General unconstrained model
4. Algorithm and numerical illustration
   - Alternative BB stepsize with adaptive nonmonotone line search
   - Numerical results
5. Discussions and future work

Outline

1. Unconstrained optimization and eigenvalue computing
2. Applications of several extreme eigenpairs
3. Variational principles for computing extreme eigenpairs
   - Block unconstrained quartic model
   - Block unconstrained β-order model
   - General unconstrained model
4. Algorithm and numerical illustration
   - Alternative BB stepsize with adaptive nonmonotone line search
   - Numerical results
5. Discussions and future work

Unconstrained optimization and eigenvalue computing

Quadratic Optimization

$$q(x) = g^T x + \frac{1}{2} x^T A x, \quad x \in \mathbb{R}^n$$

Eigenvalue Problem

$$Ax = \lambda x, \quad x \in \mathbb{R}^n \backslash \{0\}$$

A relation between gradient method and power method

Consider the gradient method for quadratic optimization

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k g_k \\ g_k &= g + A x_k \end{aligned}$$

It follows that $g_{k+1} = (I - \alpha_k A) g_k$. If $\alpha_k \equiv \alpha$, we have that

$$\frac{g_{k+1}}{\|g_{k+1}\|} = \frac{(I - \alpha A)^k g_1}{\|(I - \alpha A)^k g_1\|}$$

The value $g_k^T A g_k / \|g_k\|^2$ will return some eigenvalue of A under suitable assumptions. Therefore the gradient method with constant stepsizes can be regarded as a shifted power method. On the other hand, the (ordinary) power method can be treated as the gradient iteration with infinite stepsizes.

Finite termination property of the gradient method

For the gradient method, we generally have

$$\begin{aligned} g_{k+1} &= g_k - \alpha_k A g_k \\ &= (I - \alpha_k A) g_k \\ &= \left[\prod_{j=1}^{k} (1 - \alpha_j A) \right] g_1 \end{aligned}$$

Assuming that

$$\lambda(A) = \{\lambda_1, \lambda_2, ..., \lambda_n\}$$

we have by the Caylay-Hamilton theorem that $g_{n+1} = 0$ if

$$\left\{ \alpha_k : \; k = 1, ..., n \right\} = \left\{ \lambda_k^{-1} : \; k = 1, ..., n \right\}$$

This result was due to Yan-Lian Lai (1983).

The Barzilai-Borwein method

- Two-point stepsize gradient method [Barzilai & Borwein, 1988]
  Ask $\alpha_k I$ or $\alpha_k^{-1} I$ to have certain quasi-Newton property and solve

$$\min_{\alpha_k} \|s_{k-1} - \alpha_k y_{k-1}\|_2 \quad \text{or} \quad \min_{\alpha_k} \|\alpha_k^{-1} s_{k-1} - y_{k-1}\|_2,$$

  where $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$.

- The large and short BB stepsizes are respectively defined as

$$\alpha_k^{\text{LBB}} = \frac{\|s_{k-1}\|_2^2}{s_{k-1}^{\text{T}} y_{k-1}} \quad \text{and} \quad \alpha_k^{\text{SBB}} = \frac{s_{k-1}^{\text{T}} y_{k-1}}{\|y_{k-1}\|_2^2}.$$

- Remark that for quadratic optimization, the stepsize α_k^{LBB} reduces to

$$\alpha_k = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T A g_{k-1}},$$

  which is exactly the inverse of Reighley quotient of A with respect to $-g_{k-1}$.

Superlinear results for BB-like gradient methods

- [Barzilai & Borwein, 1988]
  $n = 2$, R-superlinear
  $\left(\alpha_{k_{i_1}}^{-1} \to \lambda_1, \quad \alpha_{k_{i_2}}^{-1} \to \lambda_2\right)$
- [Dai & Fletcher, 2005]
  $n = 3$, R-superlinear
- [Dai & Fletcher, 2005]
  Cyclic SD method $(\alpha_{mk+i} = \alpha_{mk+1}^{SD},\ 1 \le i \le m)$,
  $m \ge \dfrac{n}{2} + 1$, R-superlinear
  $(\alpha_{k_i} \to \lambda_i^{-1}$ for $i = 1, 2, \cdots, n)$

Unconstrained optimization model for the smallest eigenpair

- General unconstrained optimization [Auchmuty, 1989]

$$\min_{x\in\mathbb{R}^n} \; E(x) = \Phi\left(\frac{1}{2}\|x\|^2\right) + \Psi\left(\frac{1}{2}x^{\mathrm{T}}Ax\right)$$

- Unconstrained quartic model [Auchmuty, 1991; Mongeau & Torki, 2004]

$$\min_{x\in\mathbb{R}^n} \; E_4(x) = \frac{1}{4}\|x\|^4 + \frac{1}{2}x^{\mathrm{T}}Ax \qquad (1.1)$$

Noticing that $g_k = Ax_k + \|x_k\|^3 x_k$, we may consider some special gradient method (see [Gao, Dai & Tong, 2012])

Outline

1. Unconstrained optimization and eigenvalue computing
2. Applications of several extreme eigenpairs
3. Variational principles for computing extreme eigenpairs
   - Block unconstrained quartic model
   - Block unconstrained β-order model
   - General unconstrained model
4. Algorithm and numerical illustration
   - Alternative BB stepsize with adaptive nonmonotone line search
   - Numerical results
5. Discussions and future work

Eigenvalue decomposition of real symmetric matrices

$A \in \mathbb{R}^{n\times n}$ is real symmetric matrix

- Eigenvalue decomposition

$$A = Q\Lambda Q^{\mathrm{T}}$$

- The r-truncated decomposition (r largest/smallest eigenpairs)

$$AQ_{(r)} = Q_{(r)}\Lambda_{(r)}$$

  - $M_{(r)}$ stands for the first r columns of M
  - $Q_{(r)} \in \mathbb{R}^{n\times r}$ with orthonormal columns; $r \ll n$
  - $\Lambda_{(r)}$ is diagonal with largest/smallest r eigenvalues

Many applications

▲ A is large and sparse

▲ Compute a big portion of specturm

Application 1: Principal component analysis (PCA)

- Data analysis in many fields
  - pattern recognition (computer science)
  - chemical component analysis
- Given: $A \in \mathbb{R}^{I \times J}$ with I observations and J variables

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1J} \\ a_{21} & a_{22} & \cdots & a_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1} & a_{I2} & \cdots & a_{IJ} \end{pmatrix}$$

- Goal: extract r principal components

$$X_{[r]} \in \mathbb{R}^{I \times r}$$

Application 1: Principal component analysis (Cont'd)

- Principal component score matrix

$$X_{[r]} = \arg \min_{\text{rank}(X_1) \le r} \left\{ \sum_{ij} (a_{ij} - x_{ij})^2 = \|A - X_1\|_F^2 \right\}$$

- Low-rank matrix recovery

$$X_{[r]} = Q_{[r]} \Delta_{[r]} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ir} \end{pmatrix}$$

  - x_{ij} is the score of sample i on the principal j
  - $\Delta_{[r]}$ and $Q_{[r]}$ are the r largest singularpairs of A

Normally, X is the covariance matrix of real data, so it is symmetric.

▶ Compute r largest eigenpairs or singularpairs

Low-rank matrix recovery with missing values

Netflix: Given $A \in \mathbb{R}^{n \times n}$ whose values are known on the set $\mathcal{K}$

- Recovery the rank r matrix A

$$\min_{\text{rank}(X) \le r} \left\{ \sum_{(i,j)\in\mathcal{K}} (a_{ij} - x_{ij})^2 = \|A - X\|_{\mathcal{K}}^2 \right\}$$

- Nuclear norm regularization

$$\min_{X} \quad \|A - X\|_{\mathcal{K}}^2 + \lambda \|X\|_*$$

$$\Longleftrightarrow \quad X = U\,\text{diag}((\sigma_1 - 2\lambda)_+, \ldots, (\sigma_n - 2\lambda)_+)V^{\mathrm{T}},$$

where U and V is from the SVD $A_0 = U\text{diag}(\sigma_1, \ldots, \sigma_n)V^{\mathrm{T}}$

▶ Compute singular values greater than 2λ

Application 2: Electronic structure of material

- Density functional theory + local density approximation ⇒ The Kohn-Sham equation [Kohn & Sham, 1965]

$$\left(-\frac{\nabla^2}{2} + V_N(r) + V_H(r) + V_{xc}[n(r)]\right) \psi_i(r) = E_i \psi_i(r)$$

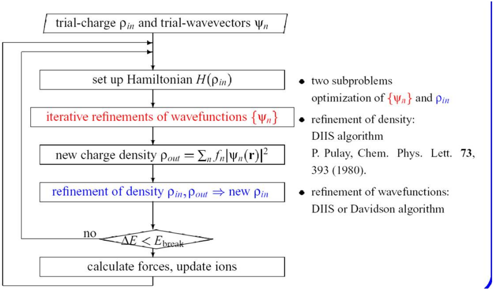
where

  - $\psi_i(r)$ and E_i are the i-th electron wave function and energy level
  - $n(r) = \sum_{i=1}^{\text{occup}} |\psi_i(r)|^2$ is the electron density distribution
  - $V_N(r)$ is the ionic pseudopotential
  - $V_H(r) = \int \frac{n(r)}{|r-\hat{r}|} d\hat{r}$ is the Hartree potential
  - $V_{xc}(r) = \frac{\delta E_{xc}(n)}{\delta n(r)}$ is the exchange-correlation potential

Application 2: Electronic structure of material (Cont'd)

$$\left(-\frac{\nabla^2}{2} + V_N(r) + V_H(r) + V_{xc}[n(r)]\right)\psi_i(r) = E_i\psi_i(r)$$

Figure: Solving the Kohn-Sham equation by iterating to self-consistency



- Compute the occupied eigenpairs every iteration

Application 3: Three dimensional photonic crystals

- Maxwell equation + discreting with FCC lattice vector $\Rightarrow$

$$Ax = \lambda Bx,$$

where $A \in \mathbb{C}^{3n\times 3n}$ is Hermitian positive semi-definite, B is positive and diagonal.

- Difficulties
  - n of the eigenvalues are zeros
  - to find k ($k = 10$) smallest positive eigenpairs

- Some existing methods
  - explicit matrix representation of the double-curl operator [Hwang, 2012]
  - project out of the null space [Hwang, 2013]

Outline

1. Unconstrained optimization and eigenvalue computing

2. Applications of several extreme eigenpairs

3. Variational principles for computing extreme eigenpairs
   - Block unconstrained quartic model
   - Block unconstrained β-order model
   - General unconstrained model

4. Algorithm and numerical illustration
   - Alternative BB stepsize with adaptive nonmonotone line search
   - Numerical results

5. Discussions and future work

Some existing methods

- Numerical algebraic methods
  - Lanczos algorithm [Lanczos, 1951]
  - Davidson's method [Davidson, 1975]
  - LOBPCG [Knyazev, 2001]
- Optimization methods
  - the Rayleigh quotient minimization [Longsine & McCormick, 1980]

$$\min_{X \in \mathbb{R}^{n\times r}} \ \mathrm{tr}\left(X^{\mathrm{T}} A X (X^{\mathrm{T}} X)^{-1}\right)$$

  - the trace minimization [Sameh & Wisniewski, 1982]

$$\min_{X \in \mathbb{R}^{n\times r}} \ \mathrm{tr}(X^{\mathrm{T}} A X) \ \text{ s.t. } \ X^{\mathrm{T}} X = I_r$$

▲ A feasible framework on the Stiefel manifold [Jiang & Dai, 2012]

$$Y(\tau, X) = \underbrace{XR(\tau)}_{\text{value space}} + \underbrace{WN(\tau)}_{\text{null space}}$$

  - what's more?

Several new block unconstrained models

1. Block unconstrained quartic model

$$\min_{X\in\mathbb{R}^{n\times r}} P(X) = \frac{1}{4}\mathrm{tr}\left(X^{\mathrm{T}}XX^{\mathrm{T}}X\right) + \frac{1}{2}\mathrm{tr}\left(X^{\mathrm{T}}AX\right) \tag{3.1}$$

2. Block unconstrained β-order model

$$\min_{X\in\mathbb{R}^{n\times r}} \widehat{P}(X;\mu,\beta,\theta) = \frac{\theta}{\beta}\|X^{\mathrm{T}}X\|_F^{\frac{\beta}{2}} + \frac{1}{2}\mathrm{tr}\left(X^{\mathrm{T}}(A-\mu I_n)X\right) \tag{3.2}$$

3. The general model

$$\min_{X\in\mathbb{R}^{n\times r}} G(X) = \Phi\left(\frac{1}{2}\|X^{\mathrm{T}}X\|_F\right) + \Psi\left(\frac{1}{2}\mathrm{tr}(X^{\mathrm{T}}AX)\right) \tag{3.3}$$

▼ They seem to be ordinary, however $\cdots$

Advantage of proposed models

- Main work

$$X^{\mathrm{T}}X, \quad X(X^{\mathrm{T}}X), \quad AX$$

whose cost is $3nr^2 + 2Nr$, where N is number of nonzero elements in A

- No orth(X) $\Longrightarrow$ parallelize

- An independent model by Wen, Yang, Liu & Zhang (2012):

$$\min_{X \in \mathbb{R}^{n \times r}} \frac{1}{2}\mathrm{tr}(X^T A X) + \frac{\mu}{4}\|X^T X - I\|_F^2$$

Stationary points of model (3.1)

$$\min_{X \in \mathbb{R}^{n\times r}} P(X) = \frac{1}{4}\mathrm{tr}\left(X^{\mathrm{T}}XX^{\mathrm{T}}X\right) + \frac{1}{2}\mathrm{tr}\left(X^{\mathrm{T}}AX\right)$$

▼ The stationary points are related to the eigenpairs of A.

Lemma 3.1

Any stationary point of (3.1) is of the thin SVD form

$$X = Q_{p,s}\left(-\Lambda_p\right)^{1/2} V_p^{\mathrm{T}},$$

where p is the rank of X, $Q_{p,s}$ consists of the $j_1, \cdots, j_p$ columns of Q with

$$1 \le j_1 \le \cdots \le j_p \le s := \arg\max_{\lambda_i < 0} i,$$

$\Lambda_p = \mathrm{diag}(\lambda_{j_1}, \cdots, \lambda_{j_p})$, *and $V_p \in \mathbb{R}^{r\times p}$ is any matrix orthonormal columns.*

Proof: The stationary point satisfies

$$\left.\begin{array}{rcl} \nabla P(X) &=& XX^{\mathrm{T}}X + AX = 0 \\ X &=& U_1\Sigma_1 V_1^{\mathrm{T}} \end{array}\right\} \Rightarrow AU_1 = U_1(-\Sigma_1^2)$$

Global minimizer of model (3.1)

$$\min_{X\in\mathbb{R}^{n\times r}} P(X) = \frac{1}{4}\mathrm{tr}\left(X^{\mathrm{T}}XX^{\mathrm{T}}X\right) + \frac{1}{2}\mathrm{tr}\left(X^{\mathrm{T}}AX\right)$$

▼ The global minimizer is related to the smallest r eigenpairs of A.

Theorem 3.2

Problem (3.1) has a rank-r stationary point if and only if $\lambda_r < 0$. *Furthermore, the global minimizer* X^* *of (3.1) is of the thin SVD form*

$$X^* = Q_{(r)}\left(\mu I_r - \Lambda_r\right)^{1/2} V_r^{\mathrm{T}} \tag{3.4}$$

and the global minimum is $P^* = -\frac{1}{4}\sum_{i=1}^{r}\lambda_i^2$.

Proof:

$$P(X) = -\frac{1}{4}\sum_{i=1}^{p}\lambda_{j_i}^2 \ \geq\ -\frac{1}{4}\sum_{i=1}^{r}\lambda_i^2 = P(X^*)$$

No undesired local minimizers

▼ Either saddle point or global minimizer $\Longrightarrow$ numerical a big merit

Theorem 3.3

If $\lambda_r < 0$*, then*

(i) *any nonzero stationary point of problem* (3.1) *is either a saddle point or a global minimizer defined in* (3.4)*.*

(ii) *Further, if* $\lambda_r < 0 \leq \lambda_{[r+1]}$*, where* $\lambda_{[r+1]}$ *is the smallest eigenvalue strictly greater than* λ_r*, all the rank-r stationary points are global minimizers.*

Model 2: Block unconstrained β-order model

$$\min_{X\in\mathbb{R}^{n\times r}} \widehat{P}(X;\mu,\beta,\theta) = \frac{\theta}{\beta}\|X^{\mathrm{T}}X\|_F^{\frac{\beta}{2}} + \frac{1}{2}\mathrm{tr}\left(X^{\mathrm{T}}(A-\mu I_n)X\right), \quad \beta>2,\ \theta>0$$

▼ All the three properties for the quartic model hold

Theorem 3.4

Problem (3.2) has a rank-r stationary point if and only if $\mu > \lambda_r$. Furthermore, there hold the following properties

(i) *the stationary point X has the form* $X = Q_{p,s}\left[c_p^{2-\frac{\beta}{2}}\theta^{-1}(\mu I_p - \Lambda_p)\right]^{1/2} V_p^{\mathrm{T}}$.

(ii) *if $\mu > \lambda_r$, the global minimizer X^* of (3.2) is of the thin SVD form*

$$X^* = Q_{(r)}\left[c^{2-\frac{\beta}{2}}\theta^{-1}(\mu I_r - \Lambda_r)\right]^{1/2} V_r^{\mathrm{T}},$$

and the global minimum is $\widehat{P}^*_{\mu,\beta,\theta} = -\frac{\theta^{-\frac{2}{\beta-2}}(\beta-2)}{2\beta}\left(\sum_{i=1}^r(\mu-\lambda_i)^2\right)^{\frac{\beta}{2(\beta-2)}}$.

(iii) *if $\mu > \lambda_r$, any nonzero stationary point of problem (3.2) is either a saddle point or a global minimizer.*

Model 3: General unconstrained model

$$\min_{X \in \mathbb{R}^{n\times r}} \; G(X) = \Phi\left(\frac{1}{2}\|X^{\mathrm{T}}X\|_F\right) + \Psi\left(\frac{1}{2}\mathrm{tr}(X^{\mathrm{T}}AX)\right)$$

▼ The stationary points are related to the eigenpairs of A.

Theorem 3.5

Under some assumptions, any nonzero stationary point of (3.3) *can be expressed by*

$$X = Q_p\Sigma_1 V_p^{\mathrm{T}}.$$

Moreover, there holds

$$\Lambda_p = -\Psi'\left(\frac{1}{2}tr(\Lambda_p\Sigma_1^2)\right)^{-1}\Phi'\left(\frac{1}{2}\|\Sigma_1^2\|_F\right)\|\Sigma_1^2\|_F^{-1}\Sigma_1^2.$$

.

The global minimizer is related to the specific formulation.

Outline

1. Unconstrained optimization and eigenvalue computing
2. Applications of several extreme eigenpairs
3. Variational principles for computing extreme eigenpairs
   - Block unconstrained quartic model
   - Block unconstrained β-order model
   - General unconstrained model
4. Algorithm and numerical illustration
   - Alternative BB stepsize with adaptive nonmonotone line search
   - Numerical results
5. Discussions and future work

BB vs CG

[Fletcher, 2005], "On the Barzilai-Borwein method":

$$\triangle u = -f, \qquad u \in [0,1]^3$$

$$f = x(x-1)y(y-1)z(z-1)w(x,y,z)$$

$$w = \exp\Big(-\frac{1}{2}\sigma^2\big((x-\alpha)^2 + (y-\beta)^2 + (z-\gamma)^2\big)\Big)$$

$$A\,u = b, \qquad n = 10^6$$

$$\left(\Leftrightarrow \min\ \frac{1}{2}u^T A u - b^T u\right)$$

$$u_1 = 0, \qquad \|g_k\|_2 \le 10^{-6}\|g_1\|_2$$

BB vs CG

Numerical Results

$(\sigma, \alpha, \beta, \gamma)$		BB	CG
$(20, 0.5, 0.5, 0.5)$	double	543(859)	162(178)
	single	462(964)	254(387)
$(50, 0.4, 0.7, 0.5)$	double	640(1009)	285(306)
	single	310(645)	290(443)

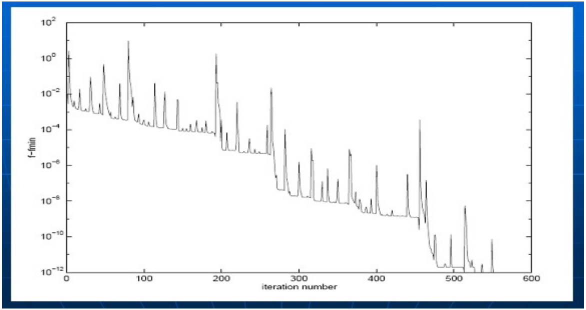
But SD: 2000, $\frac{\|g_{2000}\|}{\|g_1\|} = 0.18$!

Scholar google BB:

704 times (by May 16, 2013)

Nonmonotone performance of BB

A Typical Nonmonotone Performance of BB



For any dimensional strictly convex quadratics

- [Raydan,1993]: global convergence
- [Dai & Liao, (2002)]: R-linear convergence

Implication: *The BB stepsize can be asymptotically accepted by the nonmonotone line search in the context of unconstrained optimization*

ABB stepsize

- Let $S_{k-1} = X_k - X_{k-1}$, $Y_{k-1} = \nabla P(X_k) - \nabla P(X_{k-1})$. The large and short BB stepsizes are respectively defined as

$$\tau_k^{\mathrm{LBB}} = \frac{\mathrm{tr}(S_{k-1}^{\mathrm{T}} S_{k-1})}{|\mathrm{tr}(S_{k-1}^{\mathrm{T}} Y_{k-1})|} \quad \text{and} \quad \tau_k^{\mathrm{SBB}} = \frac{|tr(S_{k-1}^{\mathrm{T}} Y_{k-1})|}{\mathrm{tr}(Y_{k-1}^{\mathrm{T}} Y_{k-1})}.$$

- We used the alternative BB (ABB) stepsize [Dai & Fletcher, 2005]

$$\tau_k^{\mathrm{ABB}} = \begin{cases} \tau_k^{\mathrm{SBB}}, & \text{for odd } k; \\ \tau_k^{\mathrm{LBB}}, & \text{for even } k. \end{cases} \tag{4.1}$$

Adaptive nonmonotone line search strategy

- Armijo line search + adaptive nonmonotone strategy [Dai & Zhang, 2001]

$$P(X_k - \gamma^{i_k}\tau_k^{(1)}\nabla P(X_k)) \le P_r - \delta\gamma^{i_k}\tau_k^{(1)}\|\nabla P(X_k)\|_F^2,$$

where P_r is reference value.

Algorithm 1: Adaptive nonmonotone line search strategy

if $P_{k+1} < P_{best}$ **then**
  $P_{best} = P_{k+1},\ P_c = P_{k+1},\ l = 0$
  else
    $P_c = \max\{P_c, P_{k+1}\}, l = l + 1$
    if $l = L$, **then**
      $P_r = P_c,\ P_c = P_{k+1},\ l = 0$

ABB algorithm

Algorithm 2: Adaptive ABB Method

Step 0 Give a starting point and initialize the parameters.

Step 1 If $\|\nabla P_\mu(X_k)\|_F \leq tol$, return approximated eigenparis via RR procedure and stop.

Step 2 Find the least nonnegative integer i_k satisfying

$$P(X_k - \gamma^{i_k}\tau_k^{(1)}\nabla P_\mu(X_k)) \leq P_r - \delta\gamma^{i_k}\tau_k^{(1)}\|\nabla_\mu P(X_k)\|_F^2$$

and set $\tau_k = \gamma^{i_k}\tau_k^{(1)}$.

Step 3 $X_{k+1} = X_k - \tau_k\nabla P_\mu(X_k), P_{k+1} = P(X_{k+1})$, and update P_r by Algorithm 1.

Step 4 Calculate τ_k^0 by ABB (4.1) and set $\tau_k^{(1)} = \max\{\tau_{\min}, \min\{\tau_k^{(0)}, \tau_{\max}\}\}$.

Step 5 $k := k + 1$. Go to Step 1.

Global convergence

Lemma 4.1

$\{X_k, k > 0\}$ is the sequence generated by above Algorithm 2 when tol $= 0$. Then, either $\|\nabla P(X_k)\|_F = 0$ for some finite k, or

$$\lim_{k\to\infty} \|\nabla P(X_k)\|_F = 0.$$

Denote $Y(X) = \mathrm{orth}(X), R(X) = AY(X) - Y(X)\left(Y(X)^{\mathrm{T}}AY(X)\right)$.

♣ $Y(X)$ spans the eigenspace of $A \iff R(X) = 0$.

Theorem 4.2

For any rank-r matrix X, we have

$$\|R(X)\|_F \le \sigma_1(X)^{-1}\|\nabla P(X)\|_F.$$

- $\|\nabla P(X)\|_F \le tol \implies R(X) \approx 0$.

Numerical experiments: EigUncABB

- Test matrix: 3D negative Laplacian on a rectangular finite-difference grid
- Guard vectors [Liu, 2012]: set $\bar{r} = r + 5$
- The parameters

$$\text{tol} = 10^{-3},\ \gamma = 0.5,\ \delta = 0.001,\ \tau_{\min} = 10^{-20},\ \tau_{\max} = 10^{20},\ L = 4$$

$$\mu = \begin{cases} 1.01 \times \lambda_r(X_0^{\mathrm{T}} A X_0), & \text{if } \lambda_r(X_0^{\mathrm{T}} A X_0) > 0 \\ 0.99 \times \lambda_r(X_0^{\mathrm{T}} A X_0), & \text{otherwise} \end{cases}$$

Comparison of EIGS, LOBPCG and EigUncABB

Table: Comparison of EIGS, LOBPCG and EigUncABB, $n = 16000, \bar{r} = r + 5$

	EIGS				LOBPCG				EigUncABB			
r	err	nAx	resi	time	err	iter	resi	time	err	nfe	resi	time
20	4.37e-15	1220	2.31e-14	5.7	5.51e-07	106	7.79e-04	9.4	5.92e-13	242	1.75e-06	4.7
50	4.45e-15	1433	2.47e-14	12.5	1.32e-06	96	8.76e-04	18.2	3.58e-09	233	7.20e-05	9.8
100	5.75e-15	1757	2.53e-14	25.9	8.67e-07	112	8.31e-04	37.1	1.60e-12	316	7.42e-07	27.9
150	8.22e-15	2144	2.72e-14	45.3	2.20e-06	155	9.73e-04	50.9	5.06e-07	184	1.31e-04	26.3
200	1.40e-14	2543	2.61e-14	70.2	1.01e-06	231	6.41e-04	122.4	4.41e-08	342	2.45e-05	69.8
250	1.18e-14	2700	3.18e-14	91.3	7.82e-07	255	6.67e-04	101.1	3.16e-09	249	7.91e-06	66.3
300	1.47e-14	3015	3.54e-14	122.7	2.10e-06	305	8.56e-04	211.9	5.79e-09	350	2.01e-05	125.5
350	1.98e-14	3105	3.19e-14	142.8	1.39e-06	355	7.47e-04	253.2	3.57e-10	312	1.18e-05	135.1
400	1.54e-14	3480	3.20e-14	184.9	1.08e-06	405	6.32e-04	326.0	1.43e-10	345	1.09e-05	184.9
450	1.37e-14	3662	3.16e-14	217.1	1.03e-06	455	6.47e-04	312.0	4.84e-09	367	6.26e-05	228.6
500	1.83e-14	4008	3.65e-14	266.7	1.03e-06	505	5.42e-04	397.0	2.63e-06	383	1.48e-04	288.5

$\longrightarrow$ best $\qquad \longrightarrow$ worst

- competitive with LOBPCG

  compared with EIGS, sometimes find a lower accuracy solution in less time

Comparison of different β

Table: Comparison of different β's in model (3.2) by using EigUncABB

	$\beta = 3$				$\beta = 4$				$\beta = 5$			
r	err	nfe	resi	time	err	nfe	resi	time	err	nfe	resi	time
20	1.41e-08	208	5.49e-05	3.7	5.92e-13	242	1.75e-06	4.3	7.55e-10	261	2.92e-05	4.7
50	5.69e-09	241	4.26e-05	9.9	3.58e-09	233	7.20e-05	9.6	2.77e-08	272	3.49e-05	11.2
100	6.96e-09	270	2.27e-05	23.4	1.60e-12	316	7.42e-07	27.2	1.07e-07	304	9.45e-05	26.7
150	1.55e-08	228	2.04e-05	32.6	5.06e-07	184	1.31e-04	26.8	4.58e-07	240	1.18e-04	34.3
200	6.02e-07	295	5.49e-05	61.0	4.41e-08	342	2.45e-05	70.2	5.58e-07	462	6.12e-05	94.4
250	4.99e-07	232	1.03e-04	61.0	3.16e-09	249	7.91e-06	64.9	8.04e-09	314	1.19e-05	81.5
300	1.89e-08	307	4.43e-05	106.5	5.79e-09	350	2.01e-05	125.4	1.70e-09	397	3.17e-05	142.6
350	2.40e-07	281	7.16e-05	119.8	3.57e-10	312	1.18e-05	136.8	8.03e-10	394	3.49e-06	174.0
400	3.26e-10	297	1.38e-05	160.4	1.43e-10	345	1.09e-05	180.1	6.85e-09	643	7.06e-05	333.1
450	5.77e-10	287	3.04e-06	180.2	4.84e-09	367	6.26e-05	235.5	4.81e-10	328	8.25e-06	203.0
500	1.57e-10	442	3.41e-06	327.0	2.63e-06	383	1.48e-04	283.6	9.19e-09	495	5.48e-05	366.5

- the 5-order (quintic) model is worst
  the 3-order (cubic) and 4-order (quartic) models is similar

Discussions and future work

1. Our unconstrained models can easily be parallelized. How to design faster algorithms taking advantage of parallelization
2. Faster gradient algorithms using more approximated eigenvalues

Thank you!