Solving Constrained Rayleigh Quotient Optimization Problem by Projected QEP Method

Yunshen Zhou
Advisor: Prof. Zhaojun Bai

University of California, Davis

yszhou@math.ucdavis.edu

June 15, 2017
Overview

1. Introduction

2. Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP

3. Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt

4. Convergence Analysis

5. Application
Outline

1 Introduction

2 Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP

3 Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt

4 Convergence Analysis

5 Application
The Constrained Rayleigh Quotient Optimization Problem

- Consider the problem CRQopt

\[
\begin{align*}
\min \ g(v) &= v^T A v, \\
\text{s.t.} \ v^T v &= 1, \\
C^T v &= t.
\end{align*}
\]

- \( A \in \mathbb{R}^{n \times n}, \ A = A^T, \ C \in \mathbb{R}^{n \times m} \) is full column rank, \( t \in \mathbb{R}^m \) and \( m \ll n \).

- We are interested in the case where \( A \) is large and sparse and \( t \neq 0 \).
Applications

• Ridge regression.
• Trust region subproblems.
• Constrained least square problems.
• Transductive learning.
• Graph cut.
Methods:

Secular equations [Gander, Golub and von Matt, 1989].

Inner-outer iteration of Lanczos method [Golub, Zhang and Zha, 2000].

Projected power method [Xu, Li and Schuurmans, 2009].


Comments:

Only works for small matrices.

Only works for the case $t = 0$.

The convergence is slow, the shifting parameter is hard to determine.

The eigenvalue computation involved in each iteration of line search results in a huge amount of calculation.
Outline

1. Introduction

2. Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP

3. Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt

4. Convergence Analysis

5. Application
Vector Decomposition

Let

\[ v = n_0 + Pv. \]

where \( n_0 = (C^T)^\dagger t \) is the minimum norm solution of \( C^T v = t \) and \( u = Pv \), where \( P = I - CC^\dagger \).

We are interested in the case where \( \|n_0\| < 1 \). There are multiple feasible vectors.
Problem Transformation

By the vector decomposition, $u$ is the solution of the following CQ optimization problem (CQopt):

$$
\begin{align*}
\min \ f(u) &= u^T PAPu + 2u^T b_0, \\
\text{s.t. } &\|u\| = \gamma, \\
&u \in \mathcal{N}(C^T),
\end{align*}
$$

where $b_0 = PA n_0$ and $\gamma = \sqrt{1 - \|n_0\|^2}$. 
The Lagrange Equations

By Lagrange multiplier theory, the Lagrange equations for CQopt is

\[
\begin{align*}
(PAP - \lambda I)u &= -b_0, \\
\|u\| &= \gamma, \\
\text{where } u &\in \mathcal{N}(C^T).
\end{align*}
\]
The Lagrange Optimization

• $\lambda_1 < \lambda_2 \iff f(u_1) < f(u_2)$.

• Solving CQopt is equivalent to solving the following Lagrange optimization problem (LGopt):

$$\begin{cases}
(PAP - \lambda I)u = -b_0, \\
\|u\| = \gamma, \\
\text{when } u \in \mathcal{N}(C^T), \\
\lambda = \min.
\end{cases}$$

• When $\lambda = \min$, the second order necessary condition of Lagrange equations is satisfied.

• For standard trust-region subproblems, $PAP - \lambda I$ is positive semidefinite.
The Equivalent QEP

Back to the Lagrange equations, assuming that $\lambda \notin \lambda(PAP)$, define

$$z = (PAP - \lambda I)^{-2}b_0,$$

then we have

$$\begin{cases}
(PAP - \lambda I)^2 z = b_0, \\
\gamma^2 = b_0^T z.
\end{cases}$$

So

$$\begin{cases}
(PAP - \lambda I)^2 z = \frac{1}{\gamma^2} b_0 b_0^T z, \\
z \in \mathcal{N}(C^T).
\end{cases}$$
Equivalency of LGopt and QEPmin

LGopt:

\[
\begin{align*}
(PAP - \lambda I)u &= -b_0, \\
\|u\| &= \gamma, \\
\lambda &= \min.
\end{align*}
\]

QEPmin:

\[
\begin{align*}
(PAP - \lambda I)^2z &= \frac{1}{\gamma^2}b_0b_0^Tz, \\
z \in N(C^T), \\
\lambda &= \min, \text{ real}.
\end{align*}
\]

Theorem 1

- If \((\lambda_*, z)\) is the solution of QEPmin, there exists \(u\) such that \((\lambda_*, u)\) is the solution of LGopt.
- Conversely, let \((\lambda_*, u)\) be the solution of LGopt, there exists \(z\) such that \((\lambda_*, z)\) is the solution of QEPmin.

Therefore, LGopt is equivalent to QEPmin.
Distribution of the Eigenvalues

**Proposition 1**

The smallest real eigenvalue of the QEP

\[
(PAP - \lambda I)^2z = \frac{1}{\gamma^2}b_0b_0^Tz, \\
\gamma \in N(C^T),
\]

is the leftmost eigenvalue.
Distribution of the Eigenvalues

Figure 1: Plot of the eigenvalues of QEP. Problem size: \( n = 100, \ m = 10 \).
Outline

1. Introduction

2. Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP

3. Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt

4. Convergence Analysis

5. Application
Lanczos Algorithm

- Given a real symmetric matrix $M$, the Lanczos process generates an orthogonal matrix $Q_k$ whose column space is the Krylov subspace

$$K_k(M, r_0) = \mathcal{R}(r_0, Mr_0, \cdots, M^{k-1}r_0).$$

- The first $k$ steps of Lanczos algorithm take the matrix form

$$MQ_k = Q_kT_k + \beta_{k+1}q_{k+1}e_k^T,$$

where $T_k$ is a diagonal symmetric matrix with diagonal entries $\alpha_1, \cdots, \alpha_k$ and subdiagonal entries $\beta_2, \cdots, \beta_k$. 
Lanczos Algorithm

Require: $M, r_0$
Ensure: $\alpha, \beta, Q = [q_1, q_2, \ldots, q_k], k$

1: $\beta_1 \leftarrow \|r_0\|$
2: \textbf{if} $\beta_1 = 0$ \textbf{then}
3: \hspace{1em} \textbf{stop}
4: \textbf{end if}
5: $q_1 \leftarrow r_0/\beta_1$, $q_0 \leftarrow 0$

6: \textbf{for} $k \leftarrow 1, 2, \ldots$ \textbf{do}
7: \hspace{1em} $p \leftarrow M q_k$
8: \hspace{1em} $r \leftarrow p - \beta_k q_{k-1}$
9: \hspace{1em} $\alpha_k \leftarrow q_k^T p$
10: \hspace{1em} $r \leftarrow r - \alpha_k q_k$
11: \hspace{1em} $\beta_{k+1} \leftarrow \|r\|$
12: \hspace{1em} \textbf{if} $\beta_{k+1} = 0$ \textbf{then}
13: \hspace{2em} \textbf{stop}
14: \hspace{1em} \textbf{end if}
15: \hspace{1em} $q_{k+1} \leftarrow r/\beta_{k+1}$
16: \textbf{end for}
Matrix-Vector Multiplication \( PAPx \)

- Apply Lanczos to \( M = PAP \). If \( r_0 \in \mathcal{N}(C^T) \), then \( Mq_k = PAPq_k = PAq_k \).
- \( p = PAq_k \) can be computed by
  \[ PAq_k = (I - CC^\dagger)Aq_k = Aq_k - CC^\dagger Aq_k = Aq_k - Cy, \]
  where
  \[ y = \arg \min_{y \in \mathbb{R}^m} \|Cy - Aq_k\|. \]
Approximate QEPmin by Projection

Let $b_0 \in \mathcal{N}(C^T)$ be the starting vector of Lanczos iteration, then the projection of QEPmin is

$$\begin{align*}
(T_k - \lambda I)^2 y &= \left( \frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T - \beta_{k+1}^2 e_k e_k^T \right) y, \\
y &\in \mathbb{R}^k, \\
\lambda &= \min, \text{ real.}
\end{align*}$$
Existence of Real Eigenvalues

- **QEP**

\[
(T_k - \lambda I)^2 y = \left( \frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T - \beta_{k+1}^2 e_k e_k^T \right) y
\]  

(1)

may not have any real eigenvalues.

- **Example:**

![Plot of eigenvalues](image)

**Figure 2:** Plot of the eigenvalues of QEP (1). Problem: 
\( A = \text{diag}(1, 2, 3, 4, 5), \ C = [4, -8, 3, 1, -1]^T, \ t = 3 \) and \( k = 3 \).
Existence of Real Eigenvalues

In order to make sure that the reduced QEP has at least one real eigenvalue, we can solve the following QEP (rQEPmin):

\[
(T_k - \lambda I)^2 w = \frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T w, \\
w \in \mathbb{R}^k, \\
\lambda = \min, \text{ real.}
\]

Let \((\mu_1^{(k)}, w)\) be the eigenpair of rQEPmin, then \((\mu_1^{(k)}, Q_kw)\) is the approximation of eigenpair of QEPmin. Note that \(Q_kw \in \mathcal{N}(C^T)\).
Approximation of the Eigenvalues

Figure 3: Plot of the eigenvalues of QEP and projected QEP. Problem size: $n = 100$, $m = 10$, $k = 25$. 
Solving rQEPmin by Linearization

Let \( y = (T_k - \lambda I)w \), then rQEPmin

\[
(T_k - \lambda I)^2 w = \frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T w
\]

can be written as

\[
\begin{cases}
T_k y - \frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T w = \lambda y, \\
T_k w - y = \lambda w.
\end{cases}
\]

Let \( s = \begin{bmatrix} y \\ w \end{bmatrix} \), then

\[
\begin{bmatrix}
T_k & -\frac{\|b_0\|^2}{\gamma^2} e_1 e_1^T \\
-I & T_k
\end{bmatrix}
\begin{bmatrix} y \\ w \end{bmatrix} = \lambda \begin{bmatrix} y \\ w \end{bmatrix}.
\]
Stopping Condition

• Let $\mu_1^{(k)}$ be the smallest real eigenvalue of $rQEP_{\text{min}}$ and $w$ be the corresponding eigenvector. Let $\hat{w}^{(k)} = Q_k w$ and the residual of the QEP be

$$r_k = \left( PAP - \mu_1^{(k)} I \right)^2 \hat{w}^{(k)} - \gamma^{-2} b_0 b_0^T \hat{w}^{(k)}.$$

• Stopping condition: $\|r_k\| < \varepsilon$.

• Observation 1:

$$r_k = \left[ -2\mu_1^{(k)} \beta_{k+1} q_{k+1} + \beta_{k+1}^2 q_k + \alpha_k \beta_{k+1} q_{k+1} 
\quad + \beta_{k+1} (\alpha_{k+1} q_{k+1} + \beta_{k+2} q_{k+2}) \right] w_k + q_{k+1} \beta_k \beta_{k+1} w_{k-1}.$$
Stopping Condition

Observation 2: \[ \| r_k \| \approx |w_k| . \]

Figure 4: Plot of \( \| r_k \| \) and \( |w_k| . \) Problem size: \( n = 100, m = 10. \)
Compute the Solution of CRQopt

- Let the approximated solution of CRQopt be

\[ \hat{\nu}(k) = -\frac{\gamma^2}{\|b_0\|w_1}Q_k y + n_0, \]

where \( w_1 \) is the first component of \( w \) and \( y = (T_k - \mu_{1}^{(k)} I)w \).

- \( \hat{\nu}(k) \) is the “best” approximation of the solution of CRQopt in \( n_0 + \mathcal{K}_k(PAP, b_0) \), namely,

\[ \hat{\nu}(k) = \arg\min_{s \in n_0 + \mathcal{K}_k(PAP, b_0), \|s\| = 1} s^T A s. \]
Numerical Results: Running Time

Figure 5: The running time of different algorithms.
Outline

1 Introduction

2 Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP

3 Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt

4 Convergence Analysis

5 Application
Convergence Analysis

Let $\hat{\theta}$ be the eigenvalue of QEPmin and

$$D = \{ \lambda | PAPy = \lambda y, y \in \mathcal{N}(C^T) \},$$

suppose $\hat{\theta} \notin D$,

1. the sequence $\{g(\hat{v}^{(k)})\}$ is nonincreasing,
2. let $k_{\text{max}}$ be the smallest $k$ such that $\beta_{k_{\text{max}}+1} = 0$, then $g(\hat{v}^{(k_{\text{max}})}) = g(v)$, and
3. let $\theta_* = \min(D)$ and $\tilde{\theta} = \max(D)$, then

$$0 \leq g(\hat{v}^{(k)}) - g(v) \leq 4\|PAP - \hat{\theta}I\| \frac{\gamma}{p_{k+1}(\eta)}^2,$$

where

$$\eta = 1 + 2\frac{\theta_* - \hat{\theta}}{\tilde{\theta} - \theta_*},$$

and $p_k(x)$ is the $k$-th Chebyshev polynomial of the first kind.
Chebyshev Polynomials

Figure 6: Chebyshev polynomials of different degrees.
Numerical Results: Rate of Convergence

Figure 7: The error of objective function for different $\eta$. Problem size: $n = 500, m = 50$. 
Numerical Results: Tightness of the Bound

Figure 8: The actual error and the error bound of the objective function for different $k$. The left plot is the case where $\eta = 1.0608$ and the right plot is the case where $\eta = 1.0023$. Problem size: $n = 500, m = 50$. 
Outline

1. Introduction
2. Theory
   - Problem Decomposition
   - The Lagrange Equations
   - The Equivalent QEP
3. Algorithm
   - Lanczos Algorithm
   - Projected QEP Method
   - Compute the Solution of CRQopt
4. Convergence Analysis
5. Application
Application: Constrained Normalized Cut

Let $i_1, \cdots, i_k$ be the indexes of training data. The CRQ form of constrained normalized cut can be written as

\[
\begin{align*}
\min \quad & z^T L z, \\
\text{s.t.} \quad & z^T z = 1, \\
& C^T z = t,
\end{align*}
\]

where

\[
C = [D^{\frac{1}{2}} 1, e_{i_1}, \cdots, e_{i_k}],
\]

and

\[
t = [0, \pm \frac{1}{\sqrt{n}}, \cdots, \pm \frac{1}{\sqrt{n}}]^T.
\]
Introduction

Example of Application: Image Cut

(a) Normalized cut

(b) Constrained normalized cut

Figure 9: Numerical results of normalized cut with and without constraints. Problem size: \(n = 41616, m = 20\).
Example of Application: Image Cut

(a) Normalized cut  
(b) Constrained normalized cut 1  
(c) Constrained normalized cut 2

Figure 10: Numerical results of normalized cut with and without constraints. Problem size for constrained normalized cut 1: $n = 5184$, $m = 18$. Problem size for constrained normalized cut 2: $n = 5184$, $m = 24$. 
Future Work

- Tighter bound for the error of objective function.
- Compute $\hat{v}^{(k)}$ when $\hat{\theta} \in D$.
- Convergence analysis when matrix-vector multiplication has error.
References


Thank you!
LSQR

Consider the least square problem

\[
\min_{y \in \mathbb{R}^m} \|Cy - x\|_2.
\]

Let

\[
B = \begin{bmatrix}
0 & C \\
C^T & 0
\end{bmatrix},
\]

\[
W = \begin{bmatrix}
u_1 & 0 & \cdots & u_k & 0 \\
0 & v_1 & \cdots & 0 & v_k
\end{bmatrix}.
\]

We apply Lanczos algorithm to matrix \(B\) with starting vector \(\begin{bmatrix} u_1 \\ 0 \end{bmatrix}\).
LSQR

Then

\[
\begin{align*}
U_{k+1}(\beta_1 e_1) &= b, \\
AV_k &= U_{k+1}B_k, \\
A^T U_{k+1} &= V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T,
\end{align*}
\]

where

\[
B_k = \begin{bmatrix}
\alpha_1 & 0 & \cdots & \cdots & \cdots & 0 \\
\beta_2 & \alpha_2 & 0 & \cdots & \cdots & 0 \\
0 & \beta_3 & \alpha_3 & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \alpha_k \\
0 & \cdots & \cdots & \cdots & 0 & \beta_{k+1}
\end{bmatrix}.
\]

Then \( x_k = V_k y_k \) is the approximation of the solution of the least square problem, where \( y_k \) is the solution of

\[
\min_{y_k \in \mathbb{R}^k} \| \beta_1 e_1 - B_k y_k \|.
\]
Application: Constrained Normalized Cut

- Let $x$ be an indicator vector, $x_i = 1$ if $i \in A$ and $-1$ otherwise. The normalized cut can be simplified to

$$Ncut(A, B) = \frac{y^T(D - W)y}{y^TDy},$$

where $D$ is a diagonal matrix with the row sums of $W$ on the diagonal, $y$ is a linear transformation of $x$ with the property that $y^TD1 = 0$.

- Set $z = D^{\frac{1}{2}}y$, $L = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$ and rewrite the problem of minimizing the normalized cut:

$$\begin{cases}
\min z^TLz, \\
\text{s.t. } z^Tz = 1, \\
z^TD^{\frac{1}{2}}1 = 0.
\end{cases}$$
Application: Constrained Normalized Cut

- Linear constraints: information for cluster $z_i = \pm \frac{1}{\sqrt{n}}$ and the constraint $z^T D^{1/2} 1 = 0$.

- Weight Matrix:

$$w_{ij} = \begin{cases} 
    e^{-\frac{\|F(i) - F(j)\|^2}{\delta^2_F}} & \text{if } \|X(i) - X(j)\|_2 < r, \\
    0 & \text{otherwise.}
\end{cases}$$

where $F(i)$ is the brightness of pixel $i$ and $X(i)$ is the location of pixel $i$. 