Inverse Power Method for Non-linear Eigenproblems
Matthias Hein and Thomas Bühler

Anubhav Dwivedi

Department of Aerospace Engineering & Mechanics

7th March, 2017
Motivation

Non-Linear Eigenproblems

Inverse Power Method

Applications
  1-Spectral Clustering
  Sparse PCA
Motivation

- Generalized eigenvalue problems important in machine learning and statistics
- Variational formulation of eigenproblems leads to optimization of ratio of quadratic objectives
- Many constraint optimization problems with non-quadratic objectives and constraints understood as nonlinear eigenproblems
Non-Linear Eigenproblems

- Standard eigenproblem for symmetric $A \in \mathbb{R}^{n \times n}$ is of form
  \[ Af - \lambda f = 0 \]

- For symmetric matrix, A eigenvectors of A can be characterized as critical points of:
  \[ F_{\text{standard}}(f) = \frac{\langle f, Af \rangle}{\|f\|^2} \quad (1) \]

- Use Courant-Fischer Min-Max principle to compute eigenvectors
  - Central idea is to generalize this to functional of the form
  \[ F(f) = \frac{R(f)}{S(f)} \]
Non-Linear Eigenproblems

- Standard eigenproblem for symmetric $A \in \mathbb{R}^{n \times n}$ is of form

  $$Af - \lambda f = 0$$

- For symmetric matrix, A eigenvectors of A can be characterized as critical points of:

  $$F_{standard}(f) = \frac{\langle f, Af \rangle}{\|f\|^2} \quad (1)$$

- Use Courant-Fischer Min-Max principle to compute eigenvectors

- Central idea is to generalize this to functional of the form

  $$F(f) = \frac{R(f)}{S(f)}$$
For the functional,

\[ F(f) = \frac{R(f)}{S(f)} \]  \hspace{1cm} (2)

Assume,

- \( R : \mathbb{R}^n \rightarrow \mathbb{R}_+ \), \( S : \mathbb{R}^n \rightarrow \mathbb{R}_+ \) are convex and lipschitz continuous
- \( R \) and \( S \) are even and positively \( p \)-homogeneous with \( p \geq 1 \), i.e.
  \[ R(\gamma x) = \gamma^p R(x), \quad S(\gamma x) = \gamma^p S(x) \]
- \( S(f) = 0 \iff f = 0 \)
Non-linear Eigenproblems

For the functional,

\[ F(f) = \frac{R(f)}{S(f)} \]  \hspace{1cm} (2)

Assume,

- \( R : \mathbb{R}^n \rightarrow \mathbb{R}_+ \), \( S : \mathbb{R}^n \rightarrow \mathbb{R}_+ \) are convex and lipschitz continuous
- \( R \) and \( S \) are even and positively \( p \)-homogeneous with \( p \geq 1 \), i.e. \( R(\gamma x) = \gamma^p R(x) \), \( S(\gamma x) = \gamma^p S(x) \)
- \( S(f) = 0 \iff f = 0 \)
Non-linear Eigenproblems

For the functional,

\[ F(f) = \frac{R(f)}{S(f)} \quad (2) \]

Assume,

- \( R : \mathbb{R}^n \to \mathbb{R}_+ \), \( S : \mathbb{R}^n \to \mathbb{R}_+ \) are convex and lipschitz continuous
- \( R \) and \( S \) are even and positively \( p \)-homogeneous with \( p \geq 1 \), i.e.
  \[ R(\gamma x) = \gamma^p R(x), \quad S(\gamma x) = \gamma^p S(x) \]
- \( S(f) = 0 \iff f = 0 \)
Intuition

- For $R$ and $S$ to be differentiable, critical points $f^*$ satisfy
  \[ \nabla F(f^*) = 0 \iff \nabla R(f^*) - \frac{R(f^*)}{S(f^*)} \cdot \nabla S(f^*) = 0 \]

- Let, $r = \nabla R, s = \nabla S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be operators and $\lambda^* = \frac{R(f^*)}{S(f^*)}$

- Every critical point $f^*$ of $F$ satisfies the non-linear eigenproblem
  \[ r(f^*) - \lambda^* s(f^*) = 0 \] (3)

- Can be extended to non-smooth functions
Intuition

- For $R$ and $S$ to be differentiable, critical points $f^*$ satisfy

\[
\nabla F(f^*) = 0 \iff \nabla R(f^*) - \frac{R(f^*)}{S(f^*)} \cdot \nabla S(f^*) = 0
\]

- Let, $r = \nabla R$, $s = \nabla S : \mathbb{R}^n \to \mathbb{R}^n$ be operators and $\lambda^* = \frac{R(f^*)}{S(f^*)}$

- Every critical point $f^*$ of $F$ satisfies the non-linear eigenproblem

\[
r(f^*) - \lambda^* s(f^*) = 0 \tag{3}
\]

- Can be extended to non-smooth functions
For $R$ and $S$ to be differentiable, critical points $f^*$ satisfy

$$\nabla F(f^*) = 0 \iff \nabla R(f^*) - \frac{R(f^*)}{S(f^*)} \cdot \nabla S(f^*) = 0$$

Let, $r = \nabla R, s = \nabla S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be operators and $\lambda^* = \frac{R(f^*)}{S(f^*)}$

Every critical point $f^*$ of $F$ satisfies the non-linear eigenproblem

$$r(f^*) - \lambda^* s(f^*) = 0$$

Can be extended to non-smooth functions
Intuition

- For $R$ and $S$ to be differentiable, critical points $f^*$ satisfy

$$\nabla F(f^*) = 0 \iff \nabla R(f^*) - \frac{R(f^*)}{S(f^*)} \cdot \nabla S(f^*) = 0$$

- Let, $r = \nabla R, s = \nabla S : \mathbb{R}^n \to \mathbb{R}^n$ be operators and $\lambda^* = \frac{R(f^*)}{S(f^*)}$

- Every critical point $f^*$ of $F$ satisfies the non-linear eigenproblem

$$r(f^*) - \lambda^* s(f^*) = 0 \quad (3)$$

- Can be extended to non-smooth functions
Non-Differantiable Case

Definition 1
The generalized gradient at \( f \) denoted by \( \partial F(f) \) is given by

\[
\partial F(f) = \{ \xi \in \mathbb{R}^n | F^0(f, v) \geq \langle \xi, v \rangle \}
\]

where, \( F^0(f, v) = \lim_{g \to f, t \to 0} \sup_{v} F(g + tv) - F(g) \)

Definition 2
A point \( f \in \mathbb{R}^n \) is a critical point of \( F \), if \( 0 \in \partial F \)
Theorem 3

For the $R$ and $S$ satisfying the previously stated conditions. A necessary condition for $f^*$ to be a critical point is

$$0 \in \partial R(f^*) - \lambda^* \partial S(f^*), \quad \text{where} \quad \lambda^* = \frac{R(f^*)}{S(f^*)}$$

If $S$ is continuously differentiable at $f^*$, then this is also sufficient.
Algorithm: Inverse Power Method
IPM for Non-linear Eigenproblem

- Motivation
  - For linear problem the iterative scheme,
    \[ Af^{k+1} = f^k \]
    converges to the smallest eigenvector of \( A \)
  - Can be written as an optimization problem
    \[ f^{k+1} = \arg\min_u \frac{1}{2} \langle u, Au \rangle - \langle u, f^k \rangle \]

- The direct generalization solves
  \[ 0 \in r(f^{k+1}) - s(f^k) \]
  or,
  \[ f^{k+1} = \arg\min_u R(u) - \langle u, s(f^k) \rangle \]

Here, \( r(f) \in \partial R(f) \) and \( s(f) \in \partial S(f) \)
IPM for Non-linear Eigenproblem

- **Motivation**
  - For linear problem the iterative scheme,
    \[ Af^{k+1} = f^k \]
    converges to the smallest eigenvector of \( A \)
  - Can be written as an optimization problem
    \[ f^{k+1} = \arg \min_u \frac{1}{2} \langle u, Au \rangle - \langle u, f^k \rangle \]
- The direct generalization solves
  \[ 0 \in r(f^{k+1}) - s(f^k) \]
  or,
  \[ f^{k+1} = \arg \min_u R(u) - \langle u, s(f^k) \rangle \]

Here, \( r(f) \in \partial R(f) \) and \( s(f) \in \partial S(f) \)
Algorithm for degree-1 R and S

**Input:** $f^0 = \text{random with } \|f^0\| = 1, \lambda^0 = F(f^0)$

1. **repeat**
2. $f^{k+1} = \arg\min_{\|u\| \leq 1} R(u) - \lambda^k \langle u, s(f^k) \rangle$ where $s(f^k) \in \partial S(f^k)$
3. $\lambda^{k+1} = R(f^{k+1})/S(f^{k+1})$
4. **until** $|\lambda^{k+1} - \lambda^k| < \epsilon$

**Output:** eigenvalue $\lambda^{k+1}$ and eigenvector $f^{k+1}$

- Notice the requirement for unit ball constraint
Algorithm for degree-p R and S

Input: $f^0 = \text{random with } \|f^0\| = 1, \lambda^0 = F(f^0)$

1: repeat
2: $g^{k+1} = \arg\min_u R(u) - \langle u, s(f^k) \rangle$ \text{ where } $s(f^k) \in \partial S(f^k)$
3: $f^{k+1} = g^{k+1} / S(g^{k+1})^{1/p}$
4: $\lambda^{k+1} = R(f^{k+1}) / S(f^{k+1})$
5: until $|\lambda^{k+1} - \lambda^k| / \lambda^k < \epsilon$

Output: eigenvalue $\lambda^{k+1}$ and eigenvector $f^{k+1}$
Convergence

**Lemma 4**

The sequence $f^k$ produced by Alg. 1 and 2 satisfy $F(f^k) > F(f^{k+1})$ for all $k \geq 0$ or the sequences terminate.

**Theorem 5**

The sequence $f^k$ produced by Algorithms 1 and 2 converge to an eigenvector $f^*$ with eigenvalue $\lambda^* \in [0, F(f^0)]$ in the sense that it solves the non-linear eigenproblem in Theorem 3. If $S$ is continuously differentiable at $f^*$, then $F$ has a critical point at $f^*$. 
APPLICATION-1: SPECTRAL CLUSTERING
Spectral Clustering for Graph Partitioning

- **Problem**: Find optimal balanced cut of an undirected graph

  - Several ways to quantify the objective of achieving balanced cuts:
    - **Ratio Cut**, 
      \[
      R\text{Cut}(C, \overline{C}) = \frac{\text{cut}(C, \overline{C})}{|C|} + \frac{\text{cut}(C, \overline{C})}{|\overline{C}|}
      \]
    - **Ratio Cheeger cut**, 
      \[
      R\text{Cut}(C, \overline{C}) = \frac{\text{cut}(C, \overline{C})}{\min\{|C|, |\overline{C}|\}}
      \]

  - **Spectral Clustering**: solve a relaxed version of the problem Eg. Relaxation of RCut,
    \[
    v^{(2)} = \arg\min_{f \in \mathbb{R}^V} \left\{ \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} \mid \langle f, 1 \rangle = 0 \right\}
    \]
    Here, \(\Delta_2 = D - W\) represents the unnormalized graph Laplacian.
Spectral Clustering for Graph Partitioning

- **Problem:** Find optimal balanced cut of an undirected graph
- Several ways to quantify the objective of achieving balanced cuts:
  - Ratio Cut,
    \[
    RCut(C, \overline{C}) = \frac{\text{cut}(C, \overline{C})}{|C|} + \frac{\text{cut}(C, \overline{C})}{|\overline{C}|}
    \]
  - Ratio Cheeger cut,
    \[
    RCut(C, \overline{C}) = \frac{\text{cut}(C, \overline{C})}{\min\{|C|, |\overline{C}|\}}
    \]
- **Spectral Clustering:** solve a relaxed version of the problem Eg. Relaxation of RCut,
  \[
  \nu^{(2)} = \arg \min_{f \in \mathbb{R}^V} \left\{ \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} \mid \langle f, 1 \rangle = 0 \right\}
  \]
  Here, \( \Delta_2 = D - W \) represents the unnormalized graph Laplacian
Spectral Clustering for Graph Partitioning

- **Problem:** Find optimal balanced cut of an undirected graph
- Several ways to quantify the objective of achieving balanced cuts:
  - Ratio Cut,
    \[
    RCut(C, \overline{C}) = \frac{cut(C, \overline{C})}{|C|} + \frac{cut(C, \overline{C})}{|\overline{C}|}
    \]
  - Ratio Cheeger cut,
    \[
    RCut(C, \overline{C}) = \frac{cut(C, \overline{C})}{\min\{|C|, |\overline{C}|\}}
    \]
- **Spectral Clustering:** solve a relaxed version of the problem Eg. Relaxation of RCut,
  \[
  \nu^{(2)} = \arg \min_{f \in \mathbb{R}^V} \left\{ \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} \mid \langle f, 1 \rangle = 0 \right\}
  \]
  Here, \( \Delta_2 = D - W \) represents the **unnormalized graph Laplacian**
Using $\nu^{(2)}$ to partition $C, \overline{C}$ Thresholding:

$$C = \arg \min_{C_t = \{i \in V | \nu^{(2)}(i) > t\}} RCC(C_t, \overline{C}_t)$$

**Important Result:** How good is the partition?

$$\frac{h_{RCC}}{\max_{i \in V} d_i} \leq \frac{h^*_{RCC}}{\max_{i \in V} d_i} \leq 2 \left( \frac{h_{RCC}}{\max_{i \in V} d_i} \right)^{\frac{1}{2}}$$

Here, $h_{RCC} = \inf_C \{RCC(C, \overline{C})\}$ and $h^*_{RCC}$ obtained by optimal thresholding the second eigenvector.
The graph $p$-Laplacian

- Analogous to standard graph Laplacian, define $\Delta_p$ for $p \geq 1$

\[ \langle f, \Delta_p f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p \]

- The unnormalized Graph $p$–Laplacian is given by

\[ (\Delta_p f)_i = \sum_{j \in V} w_{ij} \phi_p(f_i - f_j) \]

Here, $\phi_p(x) = |x|^{p-1} \text{sign}(x)$

- **Motivation:** The relation between $h_{RCC}$ and $h_{RCC}^*$ for Graph $p$–Laplacian

\[ \frac{h_{RCC}}{\max_{i \in V} d_i} \leq \frac{h_{RCC}^*}{\max_{i \in V} d_i} \leq p \left( \frac{h_{RCC}}{\max_{i \in V} d_i} \right)^{\frac{1}{p}} \quad \forall p > 1 \]

- For $p \to 1$, the upper bound is tight
The graph $p$-Laplacian

- Analogous to standard graph Laplacian, define $\Delta_p$ for $p \geq 1$
  \[
  \langle f, \Delta_p f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p
  \]

- The unnormalized Graph $p$–Laplacian is given by
  \[
  (\Delta_p f)_i = \sum_{j \in V} w_{ij} \phi_p(f_i - f_j)
  \]

Here, $\phi_p(x) = |x|^{p-1} \text{sign}(x)$

- **Motivation:** The relation between $h_{RCC}$ and $h^*_{RCC}$ for Graph $p$–Laplacian
  \[
  \frac{h_{RCC}}{\max_{i \in V} d_i} \leq \frac{h^*_{RCC}}{\max_{i \in V} d_i} \leq p \left( \frac{h_{RCC}}{\max_{i \in V} d_i} \right)^{\frac{1}{p}} \quad \forall p > 1
  \]

- For $p \to 1$, the upper bound is tight
1-Laplacian and Eigenvectors

Consider the functional for 1–Laplacian, $\Delta_1$

$$F_1(f) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} |f_i - f_j| \|f\|_1 = \frac{\langle f, \Delta_1 f \rangle}{\|f\|_1}$$

Here,

$$(\Delta_1 f)_i = \left\{ \sum_{j=1}^{n} w_{ij} u_{ij} |u_{ij} = -u_{ij}, u_{ij} \in \text{sign}(f_i - f_j) \right\}$$

The associated non-linear eigenproblem is $0 \in \Delta_1 f - \lambda \text{sign}(f)$

Theorem 6

Any non-constant eigenvector $f^*$ of the 1–Laplacian has median zero. Moreover, let $\lambda_2$ be the second eigenvalue of the 1–Laplacian, then if $G$ is connected it holds $\lambda_2 = h_{RCC}$
1-Laplacian and Eigenvectors

- Consider the functional for 1−Laplacian, $\Delta_1$

$$F_1(f) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} |f_i - f_j| \frac{\|f\|_1}{\|f\|_1} = \frac{\langle f, \Delta_1 f \rangle}{\|f\|_1}$$

Here,

$$(\Delta_1 f)_i = \left\{ \sum_{j=1}^{n} w_{ij} u_{ij} |u_{ij} = -u_{ij}, u_{ij} \in sign(f_i - f_j) \right\}$$

- The associated non-linear eigenproblem is $0 \in \Delta_1 f - \lambda \text{sign}(f)$

**Theorem 6**

*Any non-constant eigenvector $f^*$ of the 1−Laplacian has median zero. Moreover, let $\lambda_2$ be the second eigenvalue of the 1−Laplacian, then if $G$ is connected it holds $\lambda_2 = h_{RCC}$*
IPM for second Eigenvector of $1$-Laplacian

- Global minimizer is first eigenvector, which is **constant**
- Mutual orthogonality doesn’t hold true in non-linear case
- Modified IPM for computing the non-constant eigenvector of $1$–Laplacian
- No guarantee for convergence to the second eigenvector.
IPM for second Eigenvector of 1-Laplacian

- Global minimizer is first eigenvector, which is **constant**
- Mutual orthogonality doesn’t hold true in non-linear case
- Modified IPM for computing the non-constant eigenvector of 1–Laplacian
- No guarantee for convergence to the second eigenvector.
IPM for second Eigenvector of 1-Laplacian

- Global minimizer is first eigenvector, which is constant
- Mutual orthogonality doesn’t hold true in non-linear case
- Modified IPM for computing the non-constant eigenvector of 1–Laplacian
- No guarantee for convergence to the second eigenvector.
Algorithm for Non-Constant 1-eigenvector

**Input:** Weight matrix $W$

1. **Initialization:** non-constant $f^0$ with median($f^0 = 0$) and $\|f^0\|_1 = 1$, accuracy $\epsilon$

2. **repeat**

3. $g^{k+1} = \arg \min \left\{ \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} |f_i - f_j| - \lambda_k \langle f, v^k \rangle \bigg\} \quad \|f\|_2^2 \leq 1$

4. $f^{k+1} = g^{k+1} - \text{median}(g^{k+1})$

5. $v^{k+1}_i = \begin{cases} \text{sign}(f^{k+1}_i), & \text{if } f^{k+1}_i \neq 0 \\ -\frac{|f^{k+1}_i| - |f^{k+1}_i^-|}{|f^{k+1}_0|}, & \text{if } f^{k+1}_i = 0 \end{cases}$

6. $\lambda^{k+1} = F^{k+1}_1$

7. **until** $\frac{|\lambda^{k+1} - \lambda^k|}{\lambda^k} < \epsilon$

Here, $|f^{k+1}_+|, |f^{k+1}_-|, |f^{k+1}_0| \equiv \text{Cardinality of positive, negative and zero elements}$
Quality Guarantee for 1—Spectral Clustering

- No guarantee for obtaining optimal ratio Cheeger cuts
- But always at least as good as the one found by standard spectral clustering
- The inner problem is convex but is non-smooth and can be solved with standard methods like subgradient methods
Experiment

Figure 1: Second Eigenvector of the $1-$ Laplacian and $2-$ Laplacian respectively

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. RCC</td>
<td>0.0195 (± 0.0015)</td>
<td>0.0195 (± 0.0015)</td>
<td>0.0196 (± 0.0016)</td>
<td>0.0247 (± 0.0016)</td>
</tr>
<tr>
<td>Avg. error</td>
<td>0.0462 (± 0.0161)</td>
<td>0.0491 (± 0.0181)</td>
<td>0.0578 (± 0.0285)</td>
<td>0.1685 (± 0.0200)</td>
</tr>
</tbody>
</table>
APPLICATION-2: SPARSE PRINCIPAL COMPONENT ANALYSIS
Principal Component Analysis (PCA)

- A standard technique for dimensionality reduction and data analysis.
- **Idea:** find the $k-$ dimensional subspace with maximal variance in data.
- Given a data matrix, $X \in \mathbb{R}^{n \times p}$ (each column has zero mean) and $k = 1$

$$f^* = \arg \max_{f \in \mathbb{R}^p} \frac{\langle f, X^T X f \rangle}{\|f\|_2^2} \quad (4)$$

- **Solution:** $f^*$, the largest eigenvector of the covariance matrix $\Sigma = X^T X \in \mathbb{R}^{p \times p}$
- **Issue:** Interpretation of $f^*$ difficult as all entries of $f^*$ are non-zero
Principal Component Analysis (PCA)

- A standard technique for dimensionality reduction and data analysis
- **Idea:** find the $k$– dimensional subspace with maximal variance in data.
- Given a data matrix, $X \in \mathbb{R}^{n \times p}$ (each column has zero mean) and $k = 1$

$$f^* = \arg \max_{f \in \mathbb{R}^p} \frac{\langle f, X^T X f \rangle}{\| f \|^2_2} \quad (4)$$

- **Solution:** $f^*$, the largest eigenvector of the covariance matrix $\Sigma = X^T X \in \mathbb{R}^{p \times p}$
- **Issue:** Interpretation of $f^*$ difficult as all entries of $f^*$ are non-zero
Principal Component Analysis (PCA)

- A standard technique for dimensionality reduction and data analysis
- **Idea:** find the $k$–dimensional subspace with maximal variance in data.
- Given a data matrix, $X \in \mathbb{R}^{n \times p}$ (each column has zero mean) and $k = 1$

$$ f^* = \arg \max_{f \in \mathbb{R}^p} \frac{\langle f, X^T X f \rangle}{\|f\|_2^2} \quad (4) $$

- **Solution:** $f^*$, the largest eigenvector of the covariance matrix $\Sigma = X^T X \in \mathbb{R}^{p \times p}$

- **Issue:** Interpretation of $f^*$ difficult as all entries of $f^*$ are non-zero
Principal Component Analysis (PCA)

- A standard technique for dimensionality reduction and data analysis
- **Idea**: find the $k$– dimensional subspace with maximal variance in data.
- Given a data matrix, $X \in \mathbb{R}^{n \times p}$ (each column has zero mean) and $k = 1$

$$f^* = \arg\max_{f \in \mathbb{R}^p} \frac{\langle f, X^T X f \rangle}{\| f \|_2^2}$$  \hspace{1cm} (4)

- **Solution**: $f^*$, the largest eigenvector of the covariance matrix $\Sigma = X^T X \in \mathbb{R}^{p \times p}$
- **Issue**: Interpretation of $f^*$ difficult as all entries of $f^*$ are non-zero
Sparse PCA

- Gaining sparsity by placing constraint on cardinality (no. of non-zero component)
  - **Issue:** Problem becomes NP-hard
- Solve a relaxed problem
- Problem in Eq. 4 is equivalent to

\[
f^* = \arg\min_{f \in \mathbb{R}^p} \frac{\|f\|^2}{\langle f, \Sigma f \rangle} = \arg\min_{f \in \mathbb{R}^p} \frac{\|f\|_2}{\|Xf\|_2}
\]

- To enforce sparsity, use a convex combination of \(L_1\) norm and \(L_2\) norm in the enumerator

\[
F(f) = \frac{(1 - \alpha)\|f\|_2 + \alpha \|f\|_1}{\|Xf\|_2}
\]

Here, \(\alpha \in [0, 1]\)

- Here, \(R(f) = (1 - \alpha)\|f\|_2 + \alpha \|f\|_1\) and \(S(f) = \|Xf\|_2 = \langle f, \Sigma f \rangle\) are homogeneous of degree-1
Sparse PCA

- Gaining sparsity by placing constraint on cardinality (no. of non-zero component)
  - **Issue:** Problem becomes NP-hard
- Solve a relaxed problem
- Problem in Eq. 4 is equivalent to

\[
\begin{align*}
    f^* &= \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|^2}{2} = \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|^2}{2} \\
    &= \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|^2}{\|Xf\|^2}
\end{align*}
\]

- To enforce sparsity, use a convex combination of $L_1$ norm and $L_2$ norm in the enumerator

\[
F(f) = \frac{(1 - \alpha)\|f\|_2 + \alpha\|f\|_1}{\|Xf\|_2}
\]

(5)

Here, $\alpha \in [0, 1]$

- Here, $R(f) = (1 - \alpha)\|f\|_2 + \alpha\|f\|_1$ and $S(f) = \|Xf\|_2 = \langle f, \Sigma f \rangle$ are homogeneous of degree-1
Sparse PCA

- Gaining sparsity by placing constraint on cardinality (no. of non-zero component)
  - **Issue**: Problem becomes NP-hard
- Solve a relaxed problem
- Problem in Eq. 4 is equivalent to

$$f^* = \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|_2^2}{\langle f, \Sigma f \rangle} = \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|_2}{\|Xf\|_2}$$

- To enforce sparsity, use a convex combination of $L_1$ norm and $L_2$ norm in the enumerator

$$F(f) = \frac{(1 - \alpha)\|f\|_2 + \alpha\|f\|_1}{\|Xf\|_2} \quad (5)$$

Here, $\alpha \in [0, 1]$
- Here, $R(f) = (1 - \alpha)\|f\|_2 + \alpha\|f\|_1$ and $S(f) = \|Xf\|_2 = \langle f, \Sigma f \rangle$ are homogeneous of degree-1
Gaining sparsity by placing constraint on cardinality (no. of non-zero component)

**Issue:** Problem becomes NP-hard

Solve a relaxed problem

Problem in Eq. 4 is equivalent to

$$f^* = \arg \min_{f \in \mathbb{R}^p} \frac{\|f\|^2}{2} = \arg \min_{f \in \mathbb{R}^p} \|f\|_2$$

$$\langle f, \Sigma f \rangle = \arg \min_{f \in \mathbb{R}^p} \|Xf\|_2$$

To enforce sparsity, use a convex combination of $L_1$ norm and $L_2$ norm in the enumerator

$$F(f) = \frac{(1 - \alpha)\|f\|_2 + \alpha\|f\|_1}{\|Xf\|_2} \quad (5)$$

Here, $\alpha \in [0, 1]$

Here, $R(f) = (1 - \alpha)\|f\|_2 + \alpha\|f\|_1$ and $S(f) = \|Xf\|_2 = \langle f, \Sigma f \rangle$ are homogeneous of degree-1
Experiments

- Application of IPM for sparse PCA on gene expression datasets [1]
- Comparison with $L_1$ based single-unit power algorithm [2] and EM-based algorithm [3]

According to the authors, the data matches with other works considered almost exactly.
References


Thank You . . .
Backup Slides . . .
Inner Optimization Problem

- The inner problem can be rewritten as,

\[
\begin{align*}
f^{k+1} &= \arg \min_{\|f\|_2 \leq 1} R(f) - \lambda^k \langle u, s(f^k) \rangle \\
&= \arg \min_{\|f\|_2 \leq 1} (1 - \alpha)\|f\|_2 + \alpha\|f\|_1 - \lambda^k \langle f, \mu^k \rangle
\end{align*}
\]

Here, \( \mu^k = \frac{\Sigma f^k}{\sqrt{\langle f^k, \Sigma f^k \rangle}} \)

- This problem is convex and admits an analytic solution,

\[
f_i^{k+1} = \frac{1}{s} \text{sign}(\mu_i^k)(\lambda^k |\mu_i^k| - \alpha)_+ \quad \text{Where, } s = \sqrt{\sum_{i=1}^n (\lambda^k |\mu_i^k - \alpha|)^2}_+
\]

Here, \( x_+ = \max\{0, x\} \)
Inner Optimization Problem

The inner problem can be rewritten as,

\[
f^{k+1} = \arg \min_{\|f\|_2 \leq 1} R(f) - \lambda^k \langle u, s(f^k) \rangle = \arg \min_{\|f\|_2 \leq 1} (1 - \alpha)\|f\|_2 + \alpha\|f\|_1 - \lambda^k \langle f, \mu^k \rangle
\]

Here, \( \mu^k = \frac{\sum f^k}{\sqrt{\langle f^k, \Sigma f^k \rangle}} \)

This problem is convex and admits an analytic solution,

\[
f_i^{k+1} = \frac{1}{s} \text{sign}(\mu_i^k)(\lambda^k |\mu_i^k| - \alpha)_+ \quad \text{Where, } s = \sqrt{\sum_{i=1}^n (\lambda^k |\mu_i^k - \alpha|)}_+
\]

Here, \( x_+ = \max\{0, x\} \)