Sparse Optimization and Its Applications in Bioinformatics

Shuzhong Zhang

Department of Industrial & Systems Engineering
University of Minnesota

zhangs@umn.edu

January 27, 2016
Part II
Data, Structured Data, and Information Revealing
Guess the missing number

14, 2, ?, 6, 3, 1, 21, 7, 9
Data, Structured Data, and Information Revealing

Guess the missing number

14, 2, ?, 6, 3, 1, 21, 7, 9

Guess again

1 2 3
2 6 9
7 14 21
### Movie Rating

<table>
<thead>
<tr>
<th></th>
<th>Alice</th>
<th>Bob</th>
<th>Chris</th>
<th>David</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X-Men</strong></td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Yellow Rock</strong></td>
<td>2</td>
<td>3</td>
<td>?</td>
<td>3</td>
</tr>
<tr>
<td><strong>Zoolander</strong></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>?</td>
</tr>
</tbody>
</table>
The Netflix Challenge

- The user space consists of 480,189 users.
- The rating space is {1, 2, 3, 4, 5}.
- The training data contains 100,480,507 data points.
- Test on 1,408,789 entries.

Shuzhong Zhang (ISyE@UMN) IE 5080: Sparse Optimization and Bioinformatics January 27, 2016 5 / 42
The Netflix Challenge

▶ The user space consists of 480,189 users.
▶ The rating space is $\{1, 2, 3, 4, 5\}$.
▶ The training data contains 100,480,507 data points.
▶ Test on 1,408,789 entries.
Completion and prediction problems are everywhere

- Consumer choices.
- Recommendation systems.
- Statistical learning.
- Artificial Intelligence.
- Genotypes imputing.
- Haplotype phase inferring.
The Multiplicative and Additive Principles

An Example: SAT

- Duration: 3 hours 45 minutes;
- Scoring System

<table>
<thead>
<tr>
<th>Section</th>
<th>Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math:</td>
<td>[0, 800]</td>
<td>516</td>
</tr>
<tr>
<td>Critical Reading:</td>
<td>[0, 800]</td>
<td>501</td>
</tr>
<tr>
<td>Writing:</td>
<td>[0, 800]</td>
<td>492</td>
</tr>
</tbody>
</table>
The Multiplicative and Additive Principles

An Example: SAT

- Duration: 3 hours 45 minutes;
- Scoring System

<table>
<thead>
<tr>
<th>Section</th>
<th>Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math:</td>
<td>[0, 800]</td>
<td>516</td>
</tr>
<tr>
<td>Critical Reading:</td>
<td>[0, 800]</td>
<td>501</td>
</tr>
<tr>
<td>Writing:</td>
<td>[0, 800]</td>
<td>492</td>
</tr>
</tbody>
</table>

Basic Rules:

- Skill is multiplicative;
- Composition is additive.
Movie rating revisit
Let there be $m$ movies, and $n$ viewers.
Movie rating revisit

Let there be \( m \) movies, and \( n \) viewers.

If there is only one factor underlying the rating,
Movie rating revisit
Let there be $m$ movies, and $n$ viewers.
If there is only one factor underlying the rating,
and suppose the *loading* of this factor for movie $i$ is $a_i$, $i = 1, 2, ..., m$,
and the *loading* of this factor for viewer $j$ is $b_j$, $j = 1, 2, ..., n$. 
Movie rating revisit

Let there be $m$ movies, and $n$ viewers.

If there is only one factor underlying the rating,
and suppose the loading of this factor for movie $i$ is $a_i$, $i = 1, 2, ..., m,$
and the loading of this factor for viewer $j$ is $b_j$, $j = 1, 2, ..., n$.

Then, by the multiplicative rule, the overall rating from viewer $j$ on movie $i$ should be

$$a_i \times b_j.$$
Movie rating revisit

Let there be $m$ movies, and $n$ viewers.

If there is only one factor underlying the rating,
and suppose the loading of this factor for movie $i$ is $a_i$, $i = 1, 2, \ldots, m$,
and the loading of this factor for viewer $j$ is $b_j$, $j = 1, 2, \ldots, n$.

Then, by the multiplicative rule, the overall rating from viewer $j$ on movie $i$ should be

$$a_i \times b_j.$$ 

But there might be $r$ different factors. By the additive rule, the final rating of viewer $j$ on movie $i$ should be

$$\sum_{k=1}^{r} a_i^k \times b_j^k.$$
In matrix form, the rating matrix is

\[ X = AB^\top \in \mathbb{R}^{m \times n} \]

where

\[ A = [a^1, a^2, \cdots, a^r] \]

and

\[ B = [b^1, b^2, \cdots, b^r]. \]
In matrix form, the rating matrix is

\[ X = AB^\top \in \mathbb{R}^{m \times n} \]

where

\[ A = [a^1, a^2, \ldots, a^r] \]

and

\[ B = [b^1, b^2, \ldots, b^r]. \]

We do not necessarily know in advance how many factors are there, but we know that it is a low number.
In matrix form, the rating matrix is

\[ X = AB^\top \in \mathbb{R}^{m \times n} \]

where

\[ A = [a_1, a_2, \ldots, a^r] \]

and

\[ B = [b_1, b_2, \ldots, b^r]. \]

We do not necessarily know in advance how many factors are there, but we know that it is a low number.

A possible mathematical formulation for the Netflix problem:

\[
\begin{align*}
\text{min} & \quad \text{rank}(X) \\
\text{s.t.} & \quad X_{ij} = r_{ij}, \text{ for the known entries } (i, j) \in I.
\end{align*}
\]
A 1-page introduction to the matrix spectral theory

Square matrix $X \in \mathbb{R}^{n \times n}$ has $n$ eigenvalues. *(Why?)*

Symmetric matrix $X = X^\top$ has $n$ real eigenvalues/eigenvectors. *(Why?)*

Symmetric matrix $X$ has a *spectral decomposition*: $X = Q \text{diag}(\lambda_1, \cdots, \lambda_n) Q^\top$ where $Q$ is an orthonormal matrix.

For any $X \in \mathbb{R}^{m \times n}$, the symmetric matrix $\begin{bmatrix} 0_{m \times n} & X \\ X^\top & 0_{n \times m} \end{bmatrix}$ has $m + n$ real eigenvalues. If $\lambda$ is an eigenvalue, then so is $-\lambda$: $Xp = \lambda q$, $X^\top q = \lambda u \implies X(-p) = -\lambda q$, $X^\top(q) = -\lambda(-p)$,

$$\begin{bmatrix} 0_{m \times n} & X \\ X^\top & 0_{n \times m} \end{bmatrix} = \begin{bmatrix} P & -P \\ Q & Q \end{bmatrix} \cdot \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \cdot \begin{bmatrix} P & -P \\ Q & Q \end{bmatrix}^\top,$$

& $2PP^\top = I$, $2QQ^\top = I$. Letting $U = \sqrt{2} P$ and $V = \sqrt{2} Q$: 
A 1-page introduction to the matrix spectral theory

Square matrix $X \in \mathbb{R}^{n \times n}$ has $n$ eigenvalues. (Why?)

Symmetric matrix $X = X^\top$ has $n$ real eigenvalues/eigenvectors. (Why?)

Symmetric matrix $X$ has a spectral decomposition:
$X = Q\text{diag}(\lambda_1, \cdots, \lambda_n)Q^\top$ where $Q$ is an orthonormal matrix.

For any $X \in \mathbb{R}^{m \times n}$, the symmetric matrix
$\begin{bmatrix} 0_{m \times n} & X \\ X^\top & 0_{n \times m} \end{bmatrix}$
has $m + n$ real eigenvalues. If $\lambda$ is an eigenvalue, then so is $-\lambda$:

$Xp = \lambda q, \ X^\top q = \lambda u \implies X(-p) = -\lambda q, \ X^\top(q) = -\lambda(-p),$

$\begin{bmatrix} 0_{m \times n} & X \\ X^\top & 0_{n \times m} \end{bmatrix} = \begin{bmatrix} P & -P \\ Q & Q \end{bmatrix} \cdot \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \cdot \begin{bmatrix} P & -P \\ Q & Q \end{bmatrix}^\top,$

$2PP^\top = I, \ 2QQ^\top = I.$ Letting $U = \sqrt{2}P$ and $V = \sqrt{2}Q$: $X = U\Sigma V^\top.$
The matrix norms
A matrix can be measured by norms too.

The most direct way to think of a matrix norm for \( X \in \mathbb{R}^{m \times n} \) is to view it as a vector. The associated Euclidean norm is called the *Frobenius norm*, denoted by 

\[
\|X\|_F := \left( \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j}^2 \right)^{1/2}.
\]

It is, however, more interesting to consider the matrix norm *induced* by its *singular values*, which are denoted by \( \sigma_1 > 0, \ldots, \sigma_r > 0 \), satisfying

\[
X = \sum_{\ell=1}^{r} \sigma_\ell u_\ell v_\ell^\top = U\Sigma V^\top
\]

where \( U = [u_1, \cdots, u_r] \in \mathbb{R}^{m \times r} \) and \( V = [v_1, \cdots, v_r] \in \mathbb{R}^{n \times r} \) are orthonormal, and \( \sigma_2^2, \ldots, \sigma_r^2 \) are the eigenvalues of \( X^\top X \), and \( \Sigma = \text{Diag}(\sigma_1, \cdots, \sigma_r) \).
Matrix rank and the nuclear norm

It is important to note that, although the singular value decomposition of $X$, such as $X = UΣV^T$, is not *unique* in terms of $U, V$, the set of singular values $\{σ_1, \cdots, σ_r\}$ is uniquely determined by $X$.

Clearly, $r = \text{rank}(X)$.

The so-called *matrix Schatten* norm is essentially the $L_p$-norm defined on the vector of singular values $σ := (σ_1, \cdots, σ_r)$:

$$\|X\|_{S,p} := \|σ\|_p.$$  

Three special cases:

- $\|X\|_{S,2}$ is just the Frobenius norm: $\|X\|_{S,2} = \|X\|_F$;
- $\|X\|_{S,∞}$ is also known as the *spectral norm*;
- $\|X\|_{S,1}$ is known as the *nuclear norm*, and is denoted as $\|X\|_\ast (=: \|X\|_{S,1})$. 

Just like the $L_1$-norm is a convex approximation of the $L_0$-measure, the matrix nuclear norm is a convex approximation of the rank function.

Suppose we have an under-determined system $\mathcal{L}(X) = b$, but we know that the true solution $X$ is a low-rank matrix, then we can apply the same sparse optimization idea as discussed before, to formulate an optimization model:

$$\min \quad ||X||_*$$
$$\text{s.t.} \quad \mathcal{L}(X) = b.$$
The Netflix completion problem can be viewed as *sparse* optimization in the spectrum space. Its convex relaxation problem is

$$\min \quad \|X\|_*$$
$$\text{s.t.} \quad X_{ij} = r_{ij}, \text{ for the known entries } (i, j) \in I,$$

where $\|X\|_*$ is the nuclear norm of $X$.

One may also formulate the problem as

$$\min \quad \|X\|_* + \mu \sum_{(i,j) \in I} (X_{ij} - r_{ij})^2$$
A possible CVX code

>> cvx_begin
>>     variable X(m,n)
>>     minimize norm_nuc(X)
>>     subject to
>>     L * X == b
>> cvx_end
A possible CVX code

>> cvx_begin
>> variable X(m,n)
>> minimize norm_nuc(X)
>> subject to
>> L * X == b
>> cvx_end

Testing this on the example we used before, we get

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Y</td>
<td>2</td>
<td>3</td>
<td>2.53</td>
<td>3</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.38</td>
</tr>
</tbody>
</table>
Sparse Decomposition

The low-rank hypothesis presumes that the data has a ‘simple’ structure if viewed correctly.

The data is complete, and we want to filter useful information from the data, i.e. decompose the data into

\[ D = B + F \]

- Regular part (say low-rank): background;
- Irregular part: foreground.
Background and Foreground Decomposition

- Original: \( \min \text{ rank}(B) + \mu \|F\|_0 + \nu \|N\|_F^2, \) s.t. \( D = B + F + N. \)
Background and Foreground Decomposition

- Original: \( \min \ \text{rank} \ (B) + \mu \|F\|_0 + \nu \|N\|_F^2, \text{ s.t. } D = B + F + N. \)
- Convexified: \( \min \ \|B\|_* + \mu \|F\|_1 + \nu \|N\|_F^2, \text{ s.t. } D = B + F + N. \)
Background and Foreground Decomposition

- **Original:** \( \min \text{ rank } (B) + \mu \| F \|_0 + \nu \| N \|_F^2, \) s.t. \( D = B + F + N. \)
- **Convexified:** \( \min \| B \|_* + \mu \| F \|_1 + \nu \| N \|_F^2, \) s.t. \( D = B + F + N. \)
Let us return to the gene expression analysis.

Lung cancer types:

- small cell lung cancer (SCLC);
- non-small cell lung cancer (NSCLC);
  - squamous cell carcinoma (SQ);
  - adenocarcinoma (ADCA);
  - large cell carcinoma (LC).

The identification of the ADCA subtypes is currently a major challenge.
In a recent work, we (Ma, Johnson, Ashby, Xiong, Cramer, Moore, Z., Huang, 2014) applied the ideas discussed so far to the lung cancer diagnosis problem, based on the gene expression data.

- Create a large matrix: the rows represent the genes and the columns represent the patients;
- Apply the background/foreground decomposition to extract the foreground information;
- Apply the co-clustering method on the foreground (sparse) expression data;
- The so-produced co-clusters provide the desired results.
An Illustration
An Illustration

can be decomposed as
An Illustration

can be decomposed as

\[ \begin{array}{c}
\text{Image 1} \\
\text{Image 2} \\
\text{Image 3}
\end{array} = \begin{array}{c}
\text{Image 4} \\
\text{Image 5} \\
\text{Image 6}
\end{array} + \begin{array}{c}
\text{Image 7}
\end{array} \]
Result of the SparCoC
Kaplan-Meier plot of a group of lung ADCA patients, survival curves

Oct312013_commonVector_ACC_stage1_M_noise80_79common_K1_10_K2_2_10_r

Group 1
Group 2
p = 0.0125
Genome-Wide Association Study

**SNP:** Single-Nucleotide Polymorphism.

**Haplotype:** a set of single-nucleotide polymorphisms (SNPs) on one chromosome that tend to always occur together.

**GWAS:** Genome-Wide Association Study is an examination of many common genetic variants in different individuals to see if any variant is associated with a trait.
The genotype data matrix with three reference or study panels with missing data at color-highlighted untyped SNPs, generated by three different cohorts using different genotyping chips.

<table>
<thead>
<tr>
<th></th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
<th>SNP11</th>
<th>SNP12</th>
<th>SNP13</th>
<th>SNP14</th>
<th>SNP15</th>
<th>SNP16</th>
<th>SNP17</th>
<th>SNP18</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Panel 1 sample n1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>2</td>
<td>?</td>
<td>1</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
The haplotype data matrix with color-highlighted missing data entries.
SparReC

Jiang, Ma, Hardin, Bitts, Qiao, Johnson, Z., Huang (2016), SparRec: An effective matrix completion framework of missing data imputation for GWAS.

SparReC stands for Sparse Recovery via low-rank matrix completion. Two implementations: (1) low-rank completion; (2) low number of co-clusters.
### Numerical Results

<table>
<thead>
<tr>
<th>Method</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beagle</td>
<td>0.1071</td>
<td>0.1464</td>
<td>0.1285</td>
<td>0.1797</td>
<td>0.1996</td>
<td>0.2016</td>
<td>0.2032</td>
<td>0.2066</td>
</tr>
<tr>
<td>fastPHASE</td>
<td>0.1152</td>
<td>0.1202</td>
<td>0.1216</td>
<td>0.1590</td>
<td>0.1665</td>
<td>0.1810</td>
<td>0.1766</td>
<td>0.1900</td>
</tr>
<tr>
<td>MatCompl</td>
<td>0.0192</td>
<td>0.0198</td>
<td>0.0246</td>
<td>0.0326</td>
<td>0.0441</td>
<td>0.0548</td>
<td>0.0722</td>
<td>0.1364</td>
</tr>
<tr>
<td>CoClust</td>
<td>0.0174</td>
<td>0.0159</td>
<td>0.0191</td>
<td>0.0299</td>
<td>0.0349</td>
<td>0.0477</td>
<td>0.0701</td>
<td>0.1531</td>
</tr>
</tbody>
</table>

The percentage represents the portion of the missing data.

The results are based on the error rate for completing the missing genotypes for a set of test data called Chr22.
One, two, and **THREE**

Matrix is 2-D. How about 3-D and beyond?
One, two, and THREE

Matrix is 2-D. How about 3-D and beyond?

Well, multi-dimensional matrix is known as tensor
Matrix is 2-D. How about 3-D and beyond?

Well, multi-dimensional matrix is known as *tensor*

Many data sets are in the form of tensors ...
The tensor rank saga

There are multiple ways to define a tensor rank.
The tensor rank saga

There are multiple ways to define a tensor rank.

The CP (Candecomp/Parafac) rank of \( \mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is the smallest integer \( r \) such that

\[
\mathcal{T} = \sum_{j=1}^{r} x^j \otimes y^j \otimes z^j,
\]

to be denoted as \( \text{rank}_{CP}(\mathcal{T}) = r \); that is,

\[
\mathcal{T}_{i_1i_2i_3} = \sum_{j=1}^{r} x^j_{i_1} y^j_{i_2} z^j_{i_3}, \quad \forall 1 \leq i_1 \leq n_1, 1 \leq i_2 \leq n_2, 1 \leq i_3 \leq n_3.
\]
A computational nightmare

Consider the following $9 \times 9 \times 9$ tensor (Kruskal, 1989):

\[
\begin{array}{ccc}
(1,1,1): & 1 & (4,2,1): & 1 & (7,3,1): & 1 \\
(1,4,2): & 1 & (4,5,2): & 1 & (7,6,2): & 1 \\
(1,7,3): & 1 & (4,8,3): & 1 & (7,9,3): & 1 \\
(2,1,4): & 1 & (5,2,4): & 1 & (8,3,4): & 1 \\
(2,4,5): & 1 & (5,5,5): & 1 & (8,6,5): & 1 \\
(2,7,6): & 1 & (5,8,6): & 1 & (8,9,6): & 1 \\
(3,1,7): & 1 & (6,2,7): & 1 & (9,3,7): & 1 \\
(3,4,8): & 1 & (6,5,8): & 1 & (9,6,8): & 1 \\
(3,7,9): & 1 & (6,8,9): & 1 & (9,9,9): & 1 \\
\end{array}
\]

and all other components are zero.

The CP-rank of the tensor is only known to be between 18 and 23.
The CP rank dilemma

The CP rank tells us about the fundamental number of factors behind a tensor data set. It is VERY difficult even to approximately compute the CP rank of a small tensor. However, if you DO manage to compute CP decomposition, then it is essentially unique! We need actually to use the CP rank as a function to optimize over all possible values! Something like

$$\min \text{rank}_{\text{CP}}(X) \quad \text{s.t.} \quad L(X) = b.$$
The CP rank dilemma

- The CP rank tells us about the fundamental number of factors behind a tensor data set.
The CP rank dilemma

- The CP rank tells us about the fundamental number of factors behind a tensor data set.
- It is VERY difficult even to approximately compute the CP rank of a small tensor.
The CP rank dilemma

- The CP rank tells us about the fundamental number of factors behind a tensor data set.
- It is VERY difficult even to approximately compute the CP rank of a small tensor.
- However, if you DO manage to compute CP decomposition, then it is essentially unique!
The CP rank dilemma

- The CP rank tells us about the fundamental number of factors behind a tensor data set.
- It is VERY difficult even to approximately compute the CP rank of a small tensor.
- However, if you DO manage to compute CP decomposition, then it is essentially unique!
- We need actually to use the CP rank as a function to optimize over all possible values! Something like

\[
\min \text{ rank}_{CP}(\mathbf{X}) \\
\text{s.t. } \mathbb{L}(\mathbf{X}) = b.
\]
Tensor: Grand Challenges and Opportunities

Abundance in bio-informatics:

▶ Large scale in nature.
▶ Extremely difficult to deal with, at least computationally.
▶ Contains rich amount of information.
▶ Challenges in visualization.
Tensor: Grand Challenges and Opportunities

- Abundance in bio-informatics:
Tensor: Grand Challenges and Opportunities

- Abundance in bio-informatics:

- Large scale in nature.
Tensor: Grand Challenges and Opportunities

- Abundance in bio-informatics:
  - Large scale in nature.
  - Extremely difficult to deal with, at least computationally.
Tensor: Grand Challenges and Opportunities

► Abundance in bio-informatics:

► Large scale in nature.
► Extremely difficult to deal with, at least computationally.
► Contains rich amount of information.
Tensor: Grand Challenges and Opportunities

- Abundance in bio-informatics:
  - Large scale in nature.
  - Extremely difficult to deal with, at least computationally.
  - Contains rich amount of information.
  - Challenges in visualization.
New Approximation Method

Jiang, Ma, and Z. (2014) proposed a new matricization approach to approximate the CP-rank of a tensor $\mathcal{T}$: $\text{rank}(M(\mathcal{T}))$.

The new rank is easy computable. If the CP-rank is 1, then the new computable rank is also 1. The new computable rank is sandwiched between the CP-rank and a constant ($\alpha \geq 1$) multiplying the CP-rank:

$$\text{rank}(M(\mathcal{T})) \leq \text{rank}_{\text{CP}}(\mathcal{T}) \leq \alpha \cdot \text{rank}(M(\mathcal{T})).$$
New Approximation Method

Jiang, Ma, and Z. (2014) proposed a new matricization approach to approximate the CP-rank of a tensor $\mathcal{T}$: $\text{rank}(M(\mathcal{T}))$.

The new rank is easy computable.
New Approximation Method

Jiang, Ma, and Z. (2014) proposed a new matricization approach to approximate the CP-rank of a tensor $\mathcal{T}$: $\text{rank}(M(\mathcal{T}))$.

The new rank is easy computable.

If the CP-rank is 1, then the new computable rank is also 1.
New Approximation Method

Jiang, Ma, and Z. (2014) proposed a new matricization approach to approximate the CP-rank of a tensor $\mathcal{T}$: $\text{rank}(M(\mathcal{T}))$.

The new rank is easy computable.

If the CP-rank is 1, then the new computable rank is also 1.

The new computable rank is sandwiched between the CP-rank and a constant ($\alpha \geq 1$) multiplying the CP-rank:

\[
\text{rank}(M(\mathcal{T})) \leq \text{rank}_{CP}(\mathcal{T}) \leq \alpha \cdot \text{rank}(M(\mathcal{T})).
\]
New Approximation Method

Jiang, Ma, and Z. (2014) proposed a new matricization approach to approximate the CP-rank of a tensor $\mathcal{T}$: $\text{rank}(M(\mathcal{T}))$.

The new rank is easy computable.

If the CP-rank is 1, then the new computable rank is also 1.

The new computable rank is sandwiched between the CP-rank and a constant ($\alpha \geq 1$) multiplying the CP-rank:

$$\text{rank}(M(\mathcal{T})) \leq \text{rank}_{CP}(\mathcal{T}) \leq \alpha \cdot \text{rank}(M(\mathcal{T})).$$

$$\begin{align*}
\min & \quad \|M(\mathcal{X})\|_* \\
\text{s.t.} & \quad \mathbb{L}(\mathcal{X}) = b
\end{align*}$$

tensor completion

$$\begin{align*}
\min & \quad \|M(\mathcal{Y})\|_* + \lambda \|\mathcal{Z}\|_1 \\
\text{s.t.} & \quad \mathcal{Y} + \mathcal{Z} = \mathcal{F}
\end{align*}$$

foreground/background decomposition
Colored Video Data Recovery

Three frames of the original video pictures

10% data missing

Low rank tensor recovery
50% data missing

Low rank tensor recovery
70% data missing

Low rank tensor recovery
Background and Foreground Decomposition
Other tensor ranks

The co-cluster rank \((r_1, r_2, r_3)\) of \(F \in \mathbb{R}^{n_1 \times n_2 \times n_3}\):

\[
(T)_{i,j,k} = \sum_{s=1}^{r_1} \sum_{t=1}^{r_2} \sum_{u=1}^{r_3} \lambda_{stu} X_{is} Y_{jt} Z_{ku}
\]

where \(X, Y, Z\) are assignment matrices.
Other tensor ranks

The co-cluster rank \((r_1, r_2, r_3)\) of \(\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\):

\[
(\mathcal{T})_{ijk} = \sum_{s=1}^{r_1} \sum_{t=1}^{r_2} \sum_{u=1}^{r_3} \lambda_{stu} X_{is} Y_{jt} Z_{ku}
\]

where \(X, Y, Z\) are assignment matrices.

A matrix with low co-cluster structure:

![Matrix Image]
Tensor Projection and Completion

Denote $L$ to be a certain low-complexity tensor set.

Low complexity tensor approximation

$$\begin{align*} &\text{minimize} \quad \|T - X\| \\
&\text{subject to} \quad X \in L. \end{align*}$$
Tensor Projection and Completion

Denote $\mathbf{L}$ to be a certain low-complexity tensor set.

Low complexity tensor approximation

\[
\begin{align*}
\text{minimize} & \quad \|\mathbf{T} - \mathbf{X}\| \\
\text{subject to} & \quad \mathbf{X} \in \mathbf{L}.
\end{align*}
\]

Denote $\mathbf{L}_1, \mathbf{L}_2, \cdots, \mathbf{L}_m$ to be a sequence of tensor sets with certain low-complexity structures.

Low complexity tensor decomposition

\[
\begin{align*}
\text{minimize} & \quad f_1(\mathbf{X}_1) + \cdots + f_m(\mathbf{X}_m) \\
\text{subject to} & \quad \mathbf{X}_1 + \cdots + \mathbf{X}_m = \mathbf{T} \\
& \quad \mathbf{X}_i \in \mathbf{L}_i, \ i = 1, \ldots, m.
\end{align*}
\]
More reading materials

In addition to those to be found on the Moodle site, you may also read:

- **The Matrix Theory**

- **Disease Diagnosis**
  The Merck Manuals.

- **Tensors**
Review, thoughts, and questions

- Review the basics of matrix theory.
- Review the idea of background and foreground decomposition.
- Low-rank matrix vs. sparse vector.
- What is low-rank matrix completion?
- What is Netflix Prize, and how to model it in mathematical terms?
- What is a tensor? Why is tensor difficult to deal with?
- Review the role of optimization in bioinformatics.