Global Optimality in Structured Matrix Factorization

René Vidal
Center for Imaging Science
Institute for Computational Medicine
High-Dimensional Data

• In many areas, we deal with high-dimensional data
  – Signal processing
  – Speech processing
  – Computer vision
  – Medical imaging
  – Medical robotics
  – Bioinformatics
Low Rank Modeling

• Models involving factorization are ubiquitous
  – PCA
  – Nonnegative Matrix Factorization
  – Dictionary Learning
  – Matrix Completion
  – Robust PCA

Face clustering and classification

Affine structure from motion
Convex Formulations of Matrix Factorization

- **Nuclear Norm Matrix Approximation**
  \[
  \min_X \frac{1}{2} \| Y - X \|_F^2 + \lambda \| X \|_*
  \]

- **Robust Principal Component Analysis**
  \[
  \min_X \| Y - X \|_1 + \lambda \| X \|_*
  \]

\[
\| X \|_* = \sum \sigma_i(X)
\]
Non-Convex Formul. of Matrix Factorization

• Principal Component Analysis

\[
\min_{U,V} \|Y - UV^\top\|_F^2 \quad \text{s.t.} \quad U^\top U = I
\]

• Nonnegative Matrix Factorization

\[
\min_{U,V} \|Y - UV^\top\|_F^2 \quad \text{s.t.} \quad U \geq 0, V \geq 0
\]

• Sparse Dictionary Learning

\[
\min_{U,V} \|Y - UV^\top\|_F^2 \quad \text{s.t.} \quad \|U_i\|_2 \leq 1, \|V_i\|_0 \leq r
\]
Typical Low-Rank Formulations

- **Convex formulations**
  \[
  \min_X \ell(Y, X) + \lambda \Theta(X)
  \]
  - Robust PCA
  - Matrix completion

- **Large problem size**
- **Unstructured factors**

- **Convex**

- **Factorized formulations**
  \[
  \min_{U, V} \ell(Y, UV^\top) + \lambda \Theta(U, V)
  \]
  - Nonnegative matrix factorization
  - Dictionary learning

- **Non-Convex**
- **Small problem size**
- **Structured factors**
Why Do We Need Structured Factors?

- Given a low-rank video $Y \in \mathbb{R}^{p \times t}$
  $$\min_X \|Y - X\|_1 + \lambda\|X\|_*$$

  (a) Original frames  (b) Low-rank $\hat{L}$  (c) Sparse $\hat{S}$

$$\min_{U,V} \ell(Y, UV^\top) + \lambda \Theta(U, V)$$

- **U**: spatial basis
  - Low total-variation
  - Non-negative

- **V**: temporal basis
  - Sparse on particular basis set
  - Non-negative

Why Do We Need Structured Factors?

- Nonnegative matrix factorization

\[
\min_{U,V} \| Y - U V^\top \|_F^2 \quad \text{s.t.} \quad U \geq 0, V \geq 0
\]

- Sparse dictionary learning

\[
\min_{U,V} \| Y - U V^\top \|_F^2 \quad \text{s.t.} \quad \| U_i \|_2 \leq 1, \| V_i \|_0 \leq r
\]

- **Challenges to state-of-the-art methods**
  - Need to pick size of U and V a priori
  - Alternate between U and V, without guarantees of convergence to a global minimum
Why do We Care About Convexity?

- A local minimizer of a convex problem is a global minimizer.

Why is Non Convexity a Problem?
Contributions

\[
\min_{U,V} \ell(Y, UV^\top) + \lambda \Theta(U, V)
\]

- **Assumptions:**
  - \(\ell(Y, X)\): convex and once differentiable in \(X\)
  - \(\Theta\): sum of positively homogeneous functions of degree 2

\[
f(\alpha X^1, \ldots, \alpha X^K) = \alpha^p f(X^1, \ldots, X^K) \quad \forall \alpha \geq 0
\]

- **Theorem 1:** A local minimizer \((U, V)\) such that for some \(i\)

\[
U_i = V_i = 0
\]

is a global minimizer

- **Theorem 2:** If the size of the factors is large enough, local descent can reach a global minimizer from any initialization
Contributions

\[
\min_{U,V} \ell(Y, UV^\top) + \lambda \Theta(U, V)
\]

• **Assumptions:**
  - \(\ell(Y, X)\): convex and once differentiable in \(X\)
  - \(\Theta\): sum of positively homogeneous functions of degree 2
    \[
    f(\alpha X^1, \ldots, \alpha X^K) = \alpha^p f(X^1, \ldots, X^K) \quad \forall \alpha \geq 0
    \]

• **Theorem 2:**

```latex
Figure 4.1: Left: Example critical points of a non-convex function (shown in red). (a) Saddle plateau (b,d) Global minima (c,e,g) Local maxima (f,h) Local minima (i - right panel) Saddle point. Right: Guaranteed properties of our framework. From any initialization a non-increasing path exists to a global minimum. From points on a flat plateau a simple method exists to find the edge of the plateau (green points). Plateaus (a,c) for which there is no local descent direction, there is a simple method to find the edge of the plateau from which there will be a descent direction (green points). Taken together, these results will imply a theoretical meta-algorithm that is guaranteed to find a global minimum of the non-convex factorization problem if from any point one can either find a local descent direction or verify the non-existence of a local descent direction. The primary challenge from a theoretical perspective (which is not solved by our results and is potentially NP-hard for certain problems within our framework) is thus how to find a local descent direction (which is guaranteed to exist) from a non-globally-optimal critical point.
```

Two concepts will be key to establishing our analysis framework: 1) the dimensionality of the factorized elements is not assumed to be fixed, but instead fit to the data through regularization (for example, in matrix factorization the number of columns in \(U\) and \(V\) is allowed to change) 2) we require the mapping, \(\map\), and the regularization on the factors, \(\Theta\), to be positively homogeneous (defined below).

Note that points in the interior of these plateaus could be considered both local maxima and local minima as there exists a neighborhood around these points such that the point is both maximal and minimal on that neighborhood.
Tackling Non-Convexity: Nuclear Norm Case

- **Convex problem**
  \[
  \min_X \ell(Y, X) + \lambda \|X\|_*
  \]

- **Factorized problem**
  \[
  \min_{U,V} \ell(Y, UV^T) + \lambda \Theta(U, V)
  \]

- **Variational form of the nuclear norm**
  \[
  \|X\|_* = \min_{U,V} \sum_{i=1}^{r} |U_i|_2 |V_i|_2 \quad \text{s.t.} \quad UV^T = X
  \]

- **Theorem:** Assume loss \( \ell \) is convex and once differentiable in \( X \). A local minimizer of the factorized problem such that for some \( i \) \( U_i = V_i = 0 \) is a global minimizer of both problems.

- **Intuition:** regularizer \( \Theta \) “comes from a convex function”
Tackling Non-Convexity: Nuclear Norm Case

• Convex problem
\[
\min_X \ell(Y, X) + \lambda \|X\|_*
\]

• Factorized problem
\[
\min_{U, V} \ell(Y, UV^\top) + \lambda \Theta(U, V)
\]

• **Theorem:** Assume loss \(\ell\) is convex and once differentiable in \(X\). A local minimizer of the factorized problem such that for some \(i\) \(U_i = V_i = 0\) is a global minimizer of both problems.
Tackling Non-Convexity: Tensor Norm Case

• A natural generalization is the **projective tensor norm** \([1,2]\)

\[
\|X\|_{u,v} = \min_{U,V} \sum_{i=1}^{r} \|U_i\|_u \|V_i\|_v \quad \text{s.t.} \quad UV^\top = X
\]

• **Theorem 1 [3,4]:** A local minimizer of the factorized problem

\[
\min_{U,V} \ell(Y, UV^\top) + \lambda \sum_{i=1}^{r} \|U_i\|_u \|V_i\|_v
\]

such that for some \(i\) \(U_i = V_i = 0\), is a global minimizer of both the factorized problem and of the convex problem

\[
\min_X \ell(Y, X) + \lambda \|X\|_{u,v}
\]

---

Theorem 2: If the number of columns is large enough, local descent can reach a global minimizer from any initialization.

Meta-Algorithm:
- If not at a local minima, perform local descent to reach a local minima.
- If optimality condition is satisfied, then local minima is global.
- If condition fails, choose descent direction \((u,v)\), and set

\[
\begin{align*}
r &\leftarrow r + 1 \\
U &\leftarrow \begin{bmatrix} U & u \end{bmatrix} \\
V &\leftarrow \begin{bmatrix} V & v \end{bmatrix}
\end{align*}
\]

Critical Points of Non-Convex Function

Guarantees of Our Framework
Optimization

\[
\min_{U,V} \ell(Y, UV^\top) + \lambda \sum_{i=1}^{r} \|U_i\|_u \|V_i\|_v
\]

- Convex in U given V and vice versa
- Alternating proximal gradient descent
  - Calculate gradient of smooth term
  - Compute proximal operator
  - Acceleration via extrapolation
- Advantages
  - Easy to implement
  - Highly parallelizable
  - Guaranteed convergence to Nash equilibrium (may not be local min)
Example: Nonnegative Matrix Factorization

- Original formulation

\[
\min_{U,V} \|Y - UV^\top\|^2_F \quad \text{s.t.} \quad U \geq 0, V \geq 0
\]

- New factorized formulation

\[
\min_{U,V} \|Y - UV^\top\|^2_F + \lambda \sum_i |U_i|_2 |V_i|_2 \quad \text{s.t.} \quad U, V \geq 0
\]

- Note: regularization limits the number of columns in \((U,V)\)
Example: Sparse Dictionary Learning

• Original formulation

\[
\min_{U,V} \|Y - UV^\top\|_F^2 \quad \text{s.t.} \quad \|U_i\|_2 \leq 1, \|V_i\|_0 \leq r
\]

• New factorized formulation

\[
\min_{U,V} \|Y - UV^\top\|_F^2 + \lambda \sum_i |U_i|_2(\|V_i\|_2 + \gamma |V_i|_1)
\]
Non Example: Robust PCA

• Original formulation [1]

$$\min_{X,E} \|E\|_1 + \lambda \|X\|_* \quad \text{s.t.} \quad Y = X + E$$

• Equivalent formulation

$$\min_X \|Y - X\|_1 + \lambda \|X\|_*$$

• New factorized formulation

$$\min_{U,V} \|Y - UV^\top\|_1 + \lambda \sum_i |U_i|_2 |V_i|_2$$

• Not an example because loss is not differentiable

Neural Calcium Image Segmentation

- Find neuronal shapes and spike trains in calcium imaging

\[
\min_{U,V} \| Y - \Phi(UV^T) \|_F^2 + \lambda \sum_{i=1}^{r} \| U_i \|_u \| V_i \|_v
\]

Neuron Shape

Data

True Signal

Spike Times

Why Do We Need Structure?

\[
p = \text{number of pixels}
\]

\[
t = \text{number of video frames}
\]

\[
U_1 \quad U_2
\]

\[
V_1 \quad V_2
\]

\[
Y
\]

\[
\Phi(UV^T)
\]

Why Do We Need Structure?
In Vivo Results (Small Area)

\[
\min_{U,V} \|Y - \Phi(UV^\top)\|_F^2 + \lambda \sum_{i=1}^{r} \|U_i\|_u \|V_i\|_v
\]

\[
\| \cdot \|_u = \| \cdot \|_2 + \| \cdot \|_1 + \| \cdot \|_{TV}
\]

\[
\| \cdot \|_v = \| \cdot \|_2 + \| \cdot \|_1
\]

60 microns

Raw Data
Sparse + Low Rank + Total Variation
In Vivo Results

- **PCA**
  - Sensitive to noise
  - Hard to interpret

- **Proposed method**
  - Found 46/48 manually identified active regions
  - Features are easy to interpret
  - Minimal post-processing for segmentation

![Mean Fluorescence](image1.png) ![Feature obtained by PCA](image2.png)

![Example Image Frames](image3.png) ![Features by Our Method](image4.png)
Neural Calcium Image Segmentation

Manual

Sparse

Sparse + Low-Rank

Sparse + Low-Rank + TV

Same Neuron
Hyperspectral Compressed Recovery

- $Y \in \mathbb{R}^{p \times t}$: hyperspectral image of a certain area at multiple ($t>100$) wavelengths of light

- Different regions in space correspond to different materials
  - $\text{rank}(Y) = \text{number of materials}$

- $U$: spatial features
  - Low total-variation
  - Non-negative

- $V$: spectral features
  - Non-negative

\[
\min_{U,V} \ell(Y, UV^\top) + \lambda \Theta(U, V)
\]

Hyperspectral Compressed Recovery

- Prior method: NucTV (Golbabaee et al., 2012)

\[
\min_X \|X\|_* + \lambda \sum_{i=1}^{t} \|X_i\|_{TV} \quad \text{s.t.} \quad \|Y - \Phi(X)\|_F^2 \leq \epsilon
\]

- 180 Wavelengths
- 256 x 256 Images
- Computation per Iteration
  - SVT of whole image volume
  - 180 TV Proximal Operators
  - Projection onto Constraint Set
Our method

\[
\min_{U,V} \|Y - \Phi(UV^T)\|_F^2 + \lambda \sum_{i=1}^{r} \|U_i\|_u \|V_i\|_v
\]

(U,V) have 15 columns
Problem size reduced by 91.6%
Computation per Iteration
- Calculate gradient
- 15 TV Proximal Operators
Random Initializations
Hyperspectral Compressed Recovery

\[
\frac{\|X_{true} - UV^\top\|_F}{\|X_{true}\|_F}
\]

\[
\|X_{true} - UV^\top\|_F
\]

\[
\|X_{true}\|_F
\]

![Graph showing reconstruction error for different SNR levels and subsampling ratios.](image)
Conclusions

• Structured Low Rank Matrix Factorization
  – Structure on the factors captured by the Projective Tensor Norm
  – Efficient optimization for Large Scale Problems

• Local minima of the non-convex factorized form are global minima of both the convex and non-convex forms

• Advantages in Applications
  – Neural calcium image segmentation
  – Compressed recovery of hyperspectral images
Acknowledgements

• PhD Students
  – Ben Haeffele, JHU

• Collaborators
  – Eric Young, JHU

• Grants
  – NIH DC00115
  – NIH DC00032
  – NSF 1218709
  – NSF 1447822

[1] Haeffele, Young, Vidal. Structured Low-Rank Matrix Factorization: Optimality, Algorithm, and Applications to Image Processing, ICML ’14
More Information,

Vision Lab @ Johns Hopkins University
http://www.vision.jhu.edu

Center for Imaging Science @ Johns Hopkins University
http://www.cis.jhu.edu

Thank You!