A New Convex Relaxation for Tensor Completion

Bernardino Romera-Paredes
Massimiliano Pontil

29/11/2013

CSML Lunch Seminar

Outline

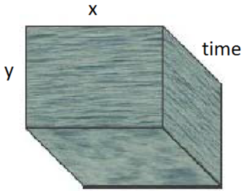
- Problem and motivation
- Convex methodology for this problem
- Rethinking the convex approach
- Alternative method
- Experimental results
- Conclusion

Problem

General problem: Learning a tensor from a set of linear measurements.

Examples:

- Tensor completion



- Video denoising/completion
- 3D scanning denoising/completion
- Context-aware recomendation
- Entities-relationships learning

- Multilinear multitask learning

Problem

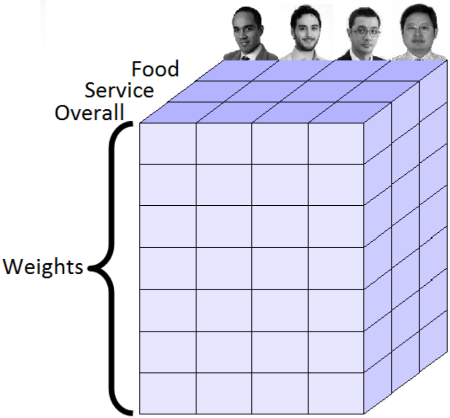
General problem: Learning a tensor from a set of linear measurements.

Examples:

- Tensor completion
- Multilinear multitask learning (Romera-Paredes et al, 2013)

Multitask learning (MTL) scenario in which tasks can be referenced by multiple indices

E.g: (, Food)



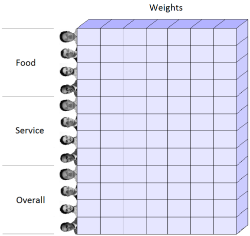
Problem modelling

We model the previous scenarios by formulating the following optimization problem:

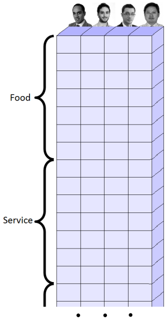
$$\underset{\mathcal{W}}{\text{argmin}}\, F(\mathcal{W}) + \frac{\gamma}{N}\sum_{n=1}^{N} \text{rank}\left(W_{(n)}\right)$$

$W_{(n)}$ is the n-th matricization of the tensor. E.g:

$W_{(1)}$



$W_{(3)}$



Convex approach

The trace norm is a widely used convex surrogate for the rank. Therefore, we can consider the following convex relaxation:

$$\underset{\mathcal{W}}{\operatorname{argmin}} F(\mathcal{W}) + \frac{\gamma}{N} \sum_{n=1}^{N} \left\| W_{(n)} \right\|_{\mathrm{Tr}}$$

Regularizer previously employed for Tensor Completion (Liu et al, 2009), (Gandy et al, 2011), (Signoretto et al, 2012)

Alternating Direction Method of Multipliers (ADMM)

- The regularizer $\frac{1}{N}\sum_{n=1}^{N}\left\|W_{(n)}\right\|_{\mathrm{Tr}}$ is composite.
- Alternating Direction Method of Multipliers is able to decouple the problem
  - This is done by introducing auxiliary tensors $\mathcal{B}^n$, $\forall n \in \mathbb{N}_N$ accounting for each term in the sum, adding the constraints $\mathcal{B}^n = \mathcal{W}$, $\forall n \in \mathbb{N}_N$.
  - Optimizing the resultant Lagrangian w.r.t. each of those tensors only involves computing the proximal operator of the trace norm:

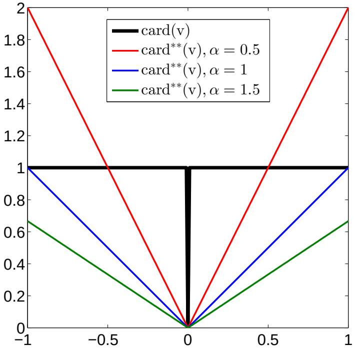
$$\mathrm{prox}_{\|\cdot\|_{\mathrm{Tr}}}(V) = \underset{X\in\mathbb{R}^{d\times d}}{\mathrm{argmin}}\frac{1}{2}\|X - V\|_{\mathrm{Fro}}^2 + \|X\|_{\mathrm{Tr}}$$

Rethinking the convex approach

Convex envelope of a function f on a set S is the largest convex function f^{**} majorized by f for all points in S

E.g: cardinality of a vector:

- $f(v) = card(v)$
- $S = \{v : ||v||_\infty \leq \alpha\}$
- $f^{**}(v) = ||v||_1/\alpha$



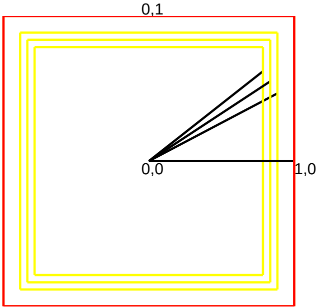
In practise α is unknown and tuned by cross validation.
Trade off: the smaller S, the tighter the convex envelope.

Rethinking the convex approach

- (Fazel 2001) $\|W\|_{\mathrm{Tr}} / \alpha$ is the convex envelope of $\mathrm{rank(W)}$ in the set
$$\left\{ W \ : \ \|W\|_{\mathrm{Sp}} \leq \alpha \right\}$$
- By using the regularizer $\sum_{n=1}^{N} \left\| W_{(n)} \right\|_{\mathrm{Tr}}$ we implicitly assume the same α for the different matricizations.
- However:
$$\left\| W_{(1)} \right\|_{\mathrm{Sp}} \neq \left\| W_{(2)} \right\|_{\mathrm{Sp}} \neq \ldots \neq \left\| W_{(N)} \right\|_{\mathrm{Sp}}$$

Rethinking the convex approach

- $\|W_{(1)}\|_{\text{Sp}} \neq \|W_{(2)}\|_{\text{Sp}} \neq \ldots \neq \|W_{(N)}\|_{\text{Sp}}$
- Let us consider $\mathcal{W} \in \mathbb{R}^{2\times 2\times 2\times 2}$. Then:

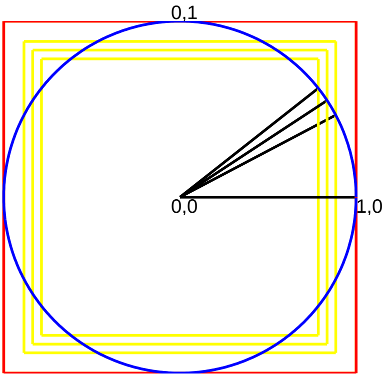


- ■ Vectors of singular values of each matricization
- ■ Smallest ℓ_∞ ball containing each of the vectors
- ■ Smallest ℓ_∞ ball containing all vectors

Rethinking the convex approach

- We are interested in convex functions over matrices invariant to matricizations of a tensor.
- The Frobenius norm is very appealing:
  - $\left\| W_{(1)} \right\|_{\mathrm{Fro}} = \left\| W_{(2)} \right\|_{\mathrm{Fro}} = \ldots = \left\| W_{(N)} \right\|_{\mathrm{Fro}}$
  - It is also a spectral function
- Therefore, we consider the set $S = \{W \ : \ \|W\|_{\mathrm{Fro}} \leq \alpha\}$
- In that set, calculating the convex envelope of the rank can be reduced to calculate the convex envelope of $\mathrm{card}\,(\mathrm{v})$ on the set $\{v \ : \ \|v\|_2 \leq \alpha\}$, where v is the vector of singular values of W.

Convex envelope of the cardinality of a vector in the ℓ_2 ball



- Vectors of singular values of each matricization
- Smallest ℓ_∞ ball containing each of the vectors
- Smallest ℓ_∞ ball containing all vectors
- Smallest ℓ_2 ball containing all vectors

Convex envelope of the cardinality of a vector in the ℓ_2 ball

- **Problem**: Getting the convex envelope of $f_\alpha(x) = \operatorname{card}(x)$, when $x \in \mathcal{B}_2$,
  - where $\mathcal{B}_2 = \left\{x \in \mathbb{R}^d : \|x\|_2 \leq \alpha\right\}$,
- The conjugate of f_α, $\forall s \in \mathbb{R}^d$, is

$$f_\alpha^*(s) = \sup_{x \in \mathcal{B}_2} x^\top s - \operatorname{card}(x) =$$
$$\max_{r \in \{0,\ldots,d\}} \{\alpha \|s_{1:r}\|_2 - r\},$$

- The biconjugate of f, $\forall v \in \mathcal{B}_2$, can be expressed as

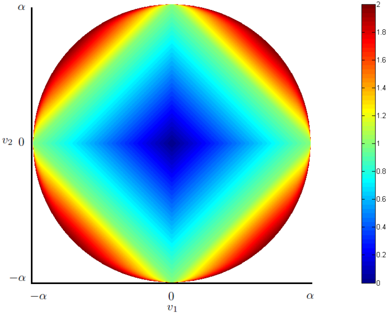
$$f_\alpha^{**}(v) = \sup_{s \in \mathbb{R}^d} s^\top v - f_\alpha^*(s) =$$
$$\sup_{s \in \mathbb{R}^d} s^\top v - \max_{r \in \{0,\ldots,d\}} \{\alpha \|s_{1:r}\|_2 - r\},$$

Convex envelope of the cardinality of a vector in the ℓ_2 ball

- $f_\alpha^{**}(v) = \sup_{s \in \mathbb{R}^d} s^\top v - \max_{r \in \{0,\ldots,d\}} \{\alpha \|s_{1:r}\|_2 - r\}, \forall v \in \mathcal{B}_2.$

- Lemma:
  If $\|v\|_2 = \alpha \rightarrow$
  $f(v) = f^{**}(v)$

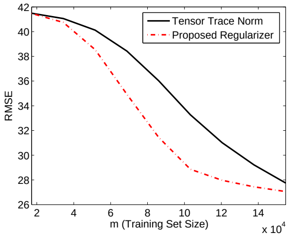
- That provides a key insight in showing that $\frac{1}{N}\sum_{n=1}^{N} \|W_{(n)}\|_{\mathrm{Tr}}$ is not tight in the spectral ball



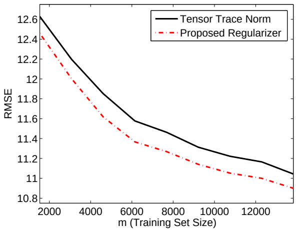
- The resultant function is difficult to compute explicitly.
- However, it is feasible to compute its proximal operator.
  - That is all we need to solve the problem via ADMM!

Experiments on tensor completion

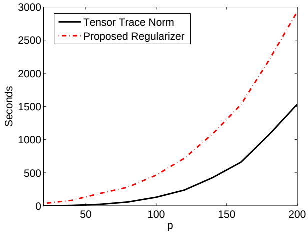
Video compression
($160 \times 112 \times 32 \times 3$ tensor)



Exam score prediction
($139 \times 11 \times 3 \times 3 \times 2$ tensor)



Time comparison:



Conclusions

- The average of the trace norm is a convex relaxation for the average of the ranks of the matricizations, but it is not tight!
- We provide a new regularizer based on the derivation of the convex envelope of the rank in the Frobenius ball.
- The proposed regularizer is tighter than the average of trace norms at all points of the Frobenius ball and at some points in the Spectral ball.
- According to empirical evidence, the proposed method is superior to the trace norm, being computationally comparable.