A Modified Frank-Wolfe Algorithm for Tensor Factorization with Unimodal Signals

Junting Chen
The Chinese University of Hong Kong, Shenzhen
Guangdong, China

Urbashi Mitra
University of Southern California
CA, USA

Acknowledgement
This research has been funded in part by one or more of the following grants:
Many signals to be estimated have unimodal properties

False estimate (known as outlier) when not exploiting unimodality

Known spectrum curves (unimodal) (Bro&Sidiropoulos'98)

Estimate spectra of different chemicals from compound samples

Signal propagation (spatially unimodal)

Source localization exploiting only unimodality in an unknown environment (e.g., underwater)
Formulation: Add a unimodality constraint to improve the estimation

\[
\mathcal{P} : \quad \text{minimize } f(x) \\
\text{subject to } x \in \mathcal{U} \cap \mathcal{M}
\]

an n-dimensional vector

the set of all unimodal vectors

\[
x_1 \leq x_2 \leq \cdots \leq x_s \leq x_{s+1} \leq \cdots \leq x_n \geq 0
\]

Goal: Design low complexity algorithms

cost function for the estimation problem

a 3D unimodal cone \(\mathcal{U}\) intersected a sphere

non-convex!
Projection will be expensive when we need it for very update!

Prior work mainly focused on projections:

- For simple objectives (e.g., least-squares, \( L_1 \), \( L \)-infinity norm):
  - Fast isotonic projection: Németh&Németh’10
  - Prefix isotonic regression: Stout’10
  - Complexity: roughly \( O(n) \) – \( O(n^2) \)

- For general objective: use projection
  - Alternating least-squares with unimodal projection: Bro&Sidiropoulos’98
  - Projected gradient: Chen&Mitra’17

\[
\mathbf{x}(t+1) = \mathcal{P}_{\mathcal{U}} \left[ \mathbf{x}(t) + \lambda_t \nabla f(\mathbf{x}(t)) \right]
\]
Can we design low complexity projection-free methods?

If the constraint set is convex, then the Frank-Wolfe update can guarantee to stay inside the constraint set.

Not the case here!

The Frank-Wolfe update procedure (no projection required)

\[ \mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \lambda_t (\hat{\mathbf{y}} - \mathbf{x}^{(t)}) \]

\[ \hat{\mathbf{y}} = \arg \min_{\mathbf{y} \in \mathcal{M}} \nabla f(\mathbf{x}^{(t)})^T \mathbf{y} \]
Proposed design: successive linear approximation could be a way to handle the non-convex constraint

\[
x^{(t+1)} = x^{(t)} + \lambda_t (\hat{y} - x^{(t)})
\]

Frank-Wolfe update

\[
\begin{align*}
\text{minimize} & \quad f(x^{(t)}) + \nabla f(x^{(t)})^T y \\
\text{subject to} & \quad y \in U(x^{(t)})
\end{align*}
\]

dynamically construct a convex constraint set
New challenges: need to dynamically design the convex constraint set $U(x^{(t)})$

$$x^{(t+1)} = x^{(t)} + \lambda_t (\hat{y} - x^{(t)})$$

minimize $y \in \mathbb{R}^n$

subject to $y \in U(x^{(t)})$

**Challenge 1:** The sub-problems need to be solved efficiently

$O(n)$ complexity or better

**Challenge 2:** Needs to justify the convergence

Convex local constraint set

$\mathcal{U} \cap \mathcal{M}$

Original constraint set (non-convex)
Property: The union of two adjacent components are convex but three are non-convex.

Choice 1:
\[ U(x^{(t)}) = U_4 \cup U_5 \]

Choice 2:
\[ U(x^{(t)}) = U_4 \cup U_5 \cup U_6 \]

\[ U_s = \left\{ x \in \mathbb{R}^n : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_s \quad x_s \geq x_{s+1} \geq \cdots \geq x_n \geq 0 \right\} \]

set of unimodal vectors with the sth element being the largest.
$U(x^{(t)})$ needs to be convex such that the sub-problems can be solved efficiently.

Choice 2:

$$U(x^{(t)}) = \text{conv}(U_4 \cup U_5 \cup U_6)$$?

$$U_s = \left\{ x \in \mathbb{R}^n : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_s \leq x_{s+1} \geq \cdots \geq x_n \geq 0 \right\}$$

set of unimodal vectors with the $s$th element being the largest.
The update always stay inside the constraint set, i.e., being unimodal

\[ U(\mathbf{x}) := \text{conv}(\tilde{U}_{S(\mathbf{x})}) \cap L(\mathbf{x}) \]

\[ L(\mathbf{x}) \triangleq \{ \mathbf{y} \in \mathbb{R}^n : a(\mathbf{x}) \leq \|\mathbf{y}\|_1 \leq b(\mathbf{x}) \} \]

where

\[ a(\mathbf{x}) = \min\{1, \hat{\alpha}(\mathbf{x})\}\|\mathbf{x}\|_1 \]

\[ b(\mathbf{x}) = \max\{1, \hat{\alpha}(\mathbf{x})\}\|\mathbf{x}\|_1 \]

\[ \hat{\alpha} = \arg\min_{\alpha \geq 0, \alpha \mathbf{x} \in \mathcal{M}} f(\alpha \mathbf{x}) \]

**Property:** \( U(\mathbf{x}) \) is a convex polytope with at most \( 2n \) extreme points

→ one being the solution to the LP
→ found in \( 2n \) steps by a simplex algorithm
Why the sub-problem can be computed efficiently?

Each $U(x^{(t)})$ is a convex polytope with at most $2n$ extreme points, so a simplex method can find the optimal solution via at most $2n$ steps.

\[
\min_{y \in \mathbb{R}^n} f(x^{(t)}) + \nabla f(x^{(t)})^T y
\]
subject to \( y \in U(x^{(t)}) \)
The algorithm converges, and global convergence is also possible

- **Theorem 1.** The algorithm converges under step size \( \lambda_t = 2/(t + 2) \).

- **Theorem 2.** Under adaptive step size, the gap
  \[ g(x) = \max_{w \in U(x)} - \nabla f(x)^T (w - x) \]
  converges to zero at a rate \( O(1/\sqrt{t + 1}) \).

  - \( g(x) = 0 \) defines a stationary point
  - \( g(x) \) gives a lower bound the duality gap

- **Theorem 3.** If \( f \) is strictly convex and its critical point \( \nabla f(x^*) = 0 \) satisfying \( f(x_1) < f(x_2) \) for any \( \|x_1 - x^*\| < \|x_2 - x^*\| \), and in addition, \( x^* \in \mathcal{U} \cap \mathcal{M} \), then the algorithm converges to the global optimal solution \( x^* = x^* \) from any initial point.
Example of Estimating an n-dimensional unimodal signal: the convergence

- Gaussian noise corrupted observation $z = c + n, \ n \sim \mathcal{N}(0, \sigma^2 I)$.
- Choice of objective $f(x) = \|x - z\|^2_2$

High SNR,
Theorem 3 applies,
global convergence

Low SNR, $\sigma=10$,
multiple stationary points
The recovery performance: enforcing unimodality enhances the estimation performance

\[
\text{minimize} \quad f(x) = \|x - z\|_2^2 \\
\text{subject to} \quad x \text{ being unimodal (and non-negative)}
\]

Unimodal projection [Stout’10] + non-negative projection
Application to tensor factorization: multimodal data $\rightarrow$ tensor model $\rightarrow$ unimodal structure for each layer $\rightarrow$ unimodal FW algorithm

- Multimodal data for estimating a source
- Signals at each layer are unimodal (peaks are assumed aligned)
- Sparse observation at each layer $N(\log N)^2 \sim M$

Tensor $\mathbf{X} \in \mathbb{R}^{N \times N \times K}$
The dominant vectors from least-squares rank-1 tensor approximation are unimodal

- **Theorem** (Chen&Mitra’18). The optimal solutions $w_1$ and $w_2$ of $P_0$ are unimodal, with their peak locations correspond to the source location.

$$P_0 : \begin{align*}
\text{minimize} & \quad \| \mathbf{X} - \alpha \times_1 w_1 \times_2 w_2 \times_3 w_3 \|^2_F \\
\text{subject to} & \quad \alpha > 0, \|w_1\| = \|w_2\| = \|w_3\| = 1
\end{align*}$$ (full observation)

- $\| \mathbf{X} \|^2_F \triangleq \sum_i \sum_j \sum_k \mathbf{X}(i, j, k)^2$

- $\mathbf{X} \times_p \mathbf{A}$ denotes the mode-$p$ multiplication
Tensor factorization enforcing unimodality constraints

\[ P_{\text{UTF}} : \min_{\alpha, w_1, w_2, w_3} \| W \odot (X - \alpha \times_1 w_1 \times_2 w_2 \times_3 w_3) \|_F^2 \]
subject to \( \alpha > 0, \| w_1 \|_1 = \| w_2 \|_1 = \| w_3 \|_1 = 1 \).
\( w_1, w_2 \in \mathcal{U} \)

**Proposed modified Frank-Wolfe algorithm** applies to update \( w_1, w_2 \), and \( w_3 \) alternatively using the gradients:

\[
\begin{align*}
\frac{1}{2} \frac{\partial f}{\partial w_1} &= -a \langle X^w_{(1)} \rangle^T (w_3 \otimes w_2) + a^2 \left[ W^T_{(1)} (w_3^2 \otimes w_2^2) \right] \odot w_1 \\
\frac{1}{2} \frac{\partial f}{\partial w_2} &= -a \langle X^w_{(2)} \rangle^T (w_3 \otimes w_1) + a^2 \left[ W^T_{(2)} (w_3^2 \otimes w_1^2) \right] \odot w_2 \\
\frac{1}{2} \frac{\partial f}{\partial w_3} &= -a \langle X^w_{(3)} \rangle^T (w_2 \otimes w_1) + a^2 \left[ W^T_{(3)} (w_2^2 \otimes w_1^2) \right] \odot w_3
\end{align*}
\]
Localization: the same source emitting two types of signals being captured by RSS and TOA sensors

- How data is generated:
  - 50% for RSS of the EM signal
    \[ P_{dB}(d) = 70 - 36 \times \log_{10}(\max\{10, d\}) + \mathcal{N}(0, \sigma_s^2) \]
  - 50% for TOA of the acoustic signal
    \[ t(d) = \frac{d}{340 \text{ m/s}} + \mathcal{N}(0, \sigma_t^2) \quad \sigma_t = 100 \text{ ms} \]

- Preprocessing: Data normalization
  \[ h_1(d) = \exp(-\beta_1 10^{-P_{dB}(d)/10}) \quad h_2(d) = \exp(-\beta_2 t(d)^2) \]

  Betas are chosen such that the normalized data is roughly uniform; \( N \) is the largest number satisfying \( 1.5N(\log N)^2 \leq \sum_k |\mathcal{M}_k| \)
Enforcing unimodality indeed improves the estimation

Unimodality-Non-Aware
(weighted centroid)

\[
\hat{s}_{RSS} = \frac{\sum_{m \in \mathcal{R}_{RSS}} q^{(m)} z^{(m)}}{\sum_{m \in \mathcal{R}_{RSS}} q^{(m)}}
\]

Matrix-based methods

Enforcing unimodality

tensor-based methods
Tensor factorization strategy fuses multimodal data better

When one of the signal modes (TOA signal) is corrupted...

Matrix-based methods: sensitive to the deterioration of the signals

Tensor-based methods: robust to the deterioration of the signals
Substantial complexity reduction by the proposed unimodal-FW algorithm

Projected gradient based on unimodal regression ([Stout’08], state-of-the-art)

Substantial complexity reduction

Proposed Unimodal Frank-Wolfe
In conclusion, we developed a unimodal-FW algorithm to solve unimodality-constrained problems.

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in \mathcal{U} \cap \mathcal{M}
\end{align*}
\]

- main idea: construct a sequence of linear sub-problems, each constrained by a convex polytope
- complexity 2n for each sub-problem
- shown to converge (global convergence possible)
- Demonstration for a data fusion problem using tensor model

Thank you & Questions?
Bibliography


