香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

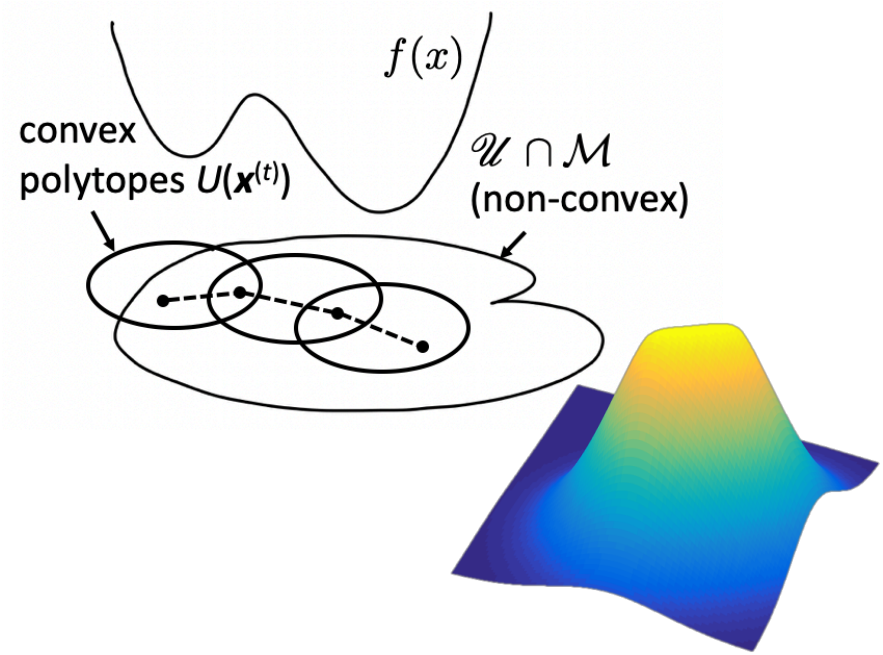
# A Modified Frank-Wolfe Algorithm for Tensor Factorization with Unimodal Signals


Junting Chen

The Chinese University of Hong Kong, Shenzhen
Guangdong, China

Urbashi Mitra

University of Southern California
CA, USA

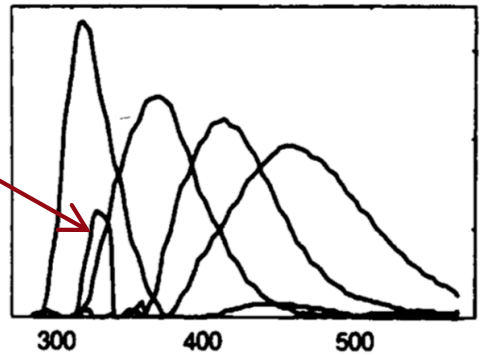


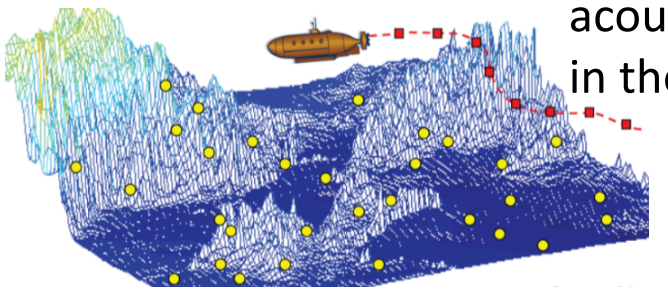
**Acknowledgement**

This research has been funded in part by one or more of the following grants:
ONR N00014-15-1-2550, NSF CNS-1213128, NSF CCF-1718560, NSF CCF-1410009 , NSF CPS-1446901, and AFOSR FA9550-12-1-0215.

# Many signals to be estimated have unimodal properties

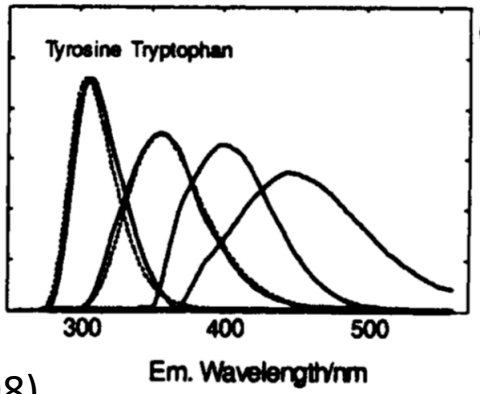
False estimate (known as outlier) when *not* exploiting unimodality



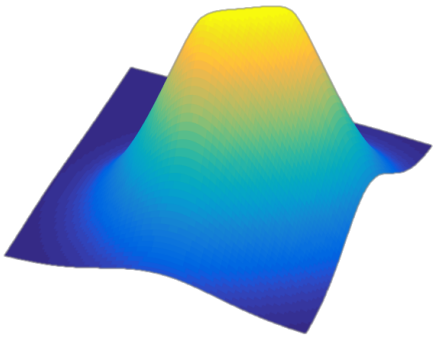



acoustic sensors in the ocean

Known spectrum curves (unimodal)




(Bro&Sidiropoulos'98)



Signal propagation (spatially unimodal)

Estimate spectra of different chemicals from compound samples

Source localization exploiting only unimodality in an unknown environment (e.g., underwater)

# Formulation: Add a unimodality constraint to improve the estimation

cost function for the estimation problem

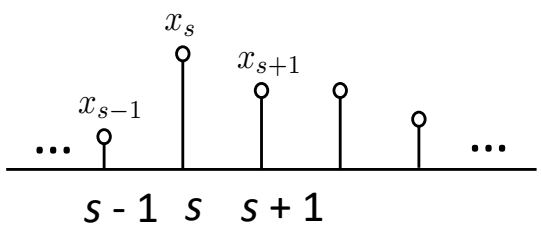
$$\mathscr{P}: \quad \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad f(\boldsymbol{x})$$
$$\text{subject to} \quad \boldsymbol{x} \in \mathscr{U} \cap \mathcal{M}$$

an n-dimensional vector

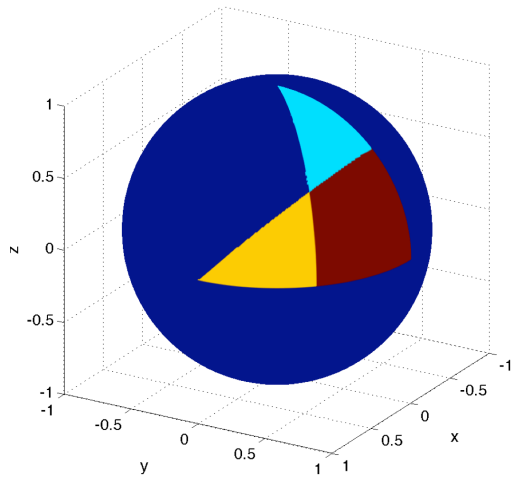
the set of all unimodal vectors

some other constraints. Here, we assume:

$$\mathcal{M} = \{\boldsymbol{x} : a \leq \|\boldsymbol{x}\|_1 \leq b\}$$




$$0 \leq x_1 \leq x_2 \cdots \leq x_s$$
$$x_{x+1} \geq x_{s+2} \geq \cdots \geq x_n \geq 0$$




a 3D unimodal cone $\mathscr{U}$ intersected a sphere

non-convex!

Goal: Design low complexity algorithms

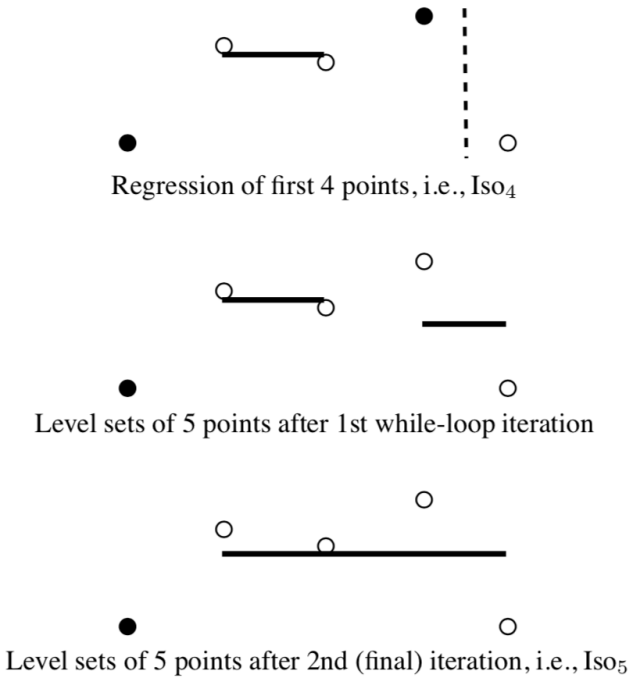
# Projection will be expensive when we need it for very update!

Prior work mainly focused on projections:

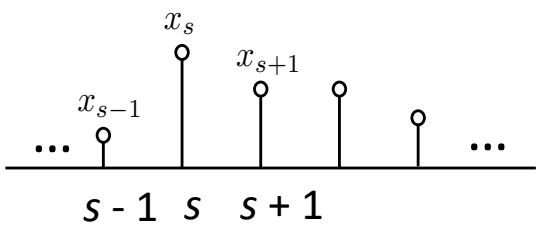
- For simple objectives (e.g., least-squares, $L_1$, $L$-infinity norm):
  - Fast isotonic projection: Németh&Németh'10
    Prefix isotonic regression: Stout'10
  - Complexity: roughly $O(n) - O(n^2)$

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

$$\boldsymbol{x}^{(t+1)} = \mathcal{P}_{\mathscr{U}}\left[\boldsymbol{x}^{(t)} + \lambda_t \nabla f(\boldsymbol{x}^{(t)})\right]$$



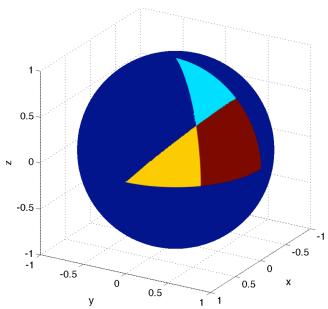

prefix isotonic projection



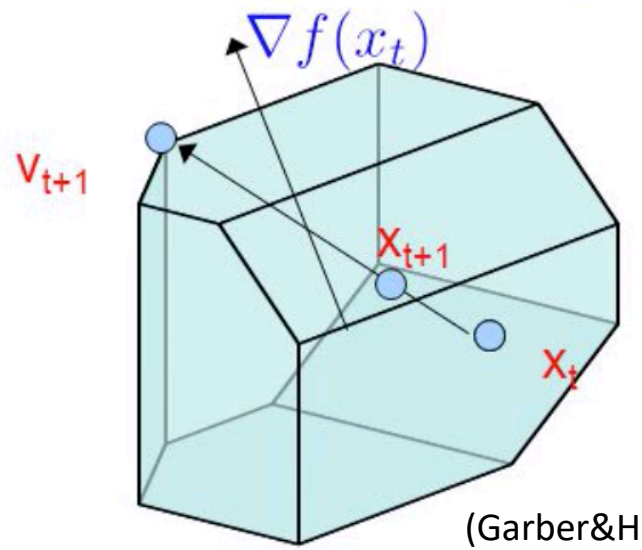
# Can we design low complexity projection-free methods?

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)




(Garber&Hazan)

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

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}\in\mathcal{M}} \nabla f(\boldsymbol{x}^{(t)})^{\mathrm{T}}\boldsymbol{y}$$

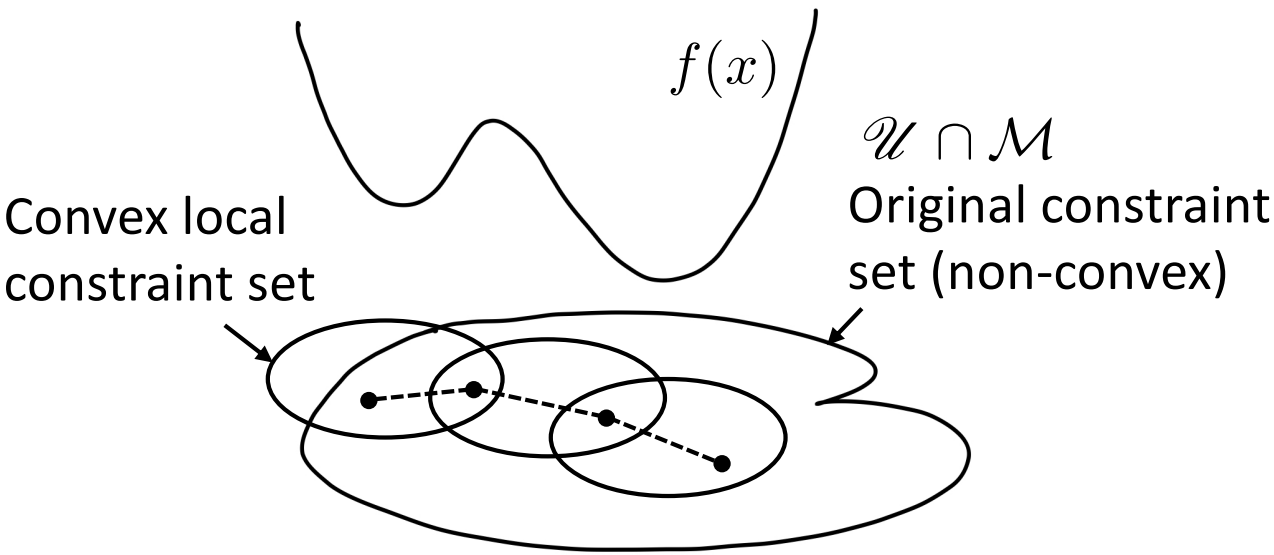
# Proposed design: successive linear approximation could be a way to handle the non-convex constraint

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

Frank-Wolfe update

$$\begin{aligned} \underset{\boldsymbol{y}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}^{(t)}) + \nabla f(\boldsymbol{x}^{(t)})^{\mathrm{T}}\boldsymbol{y} \\ \text{subject to} \quad & \boldsymbol{y} \in U(\boldsymbol{x}^{(t)}) \end{aligned}$$

dynamically construct a convex constraint set



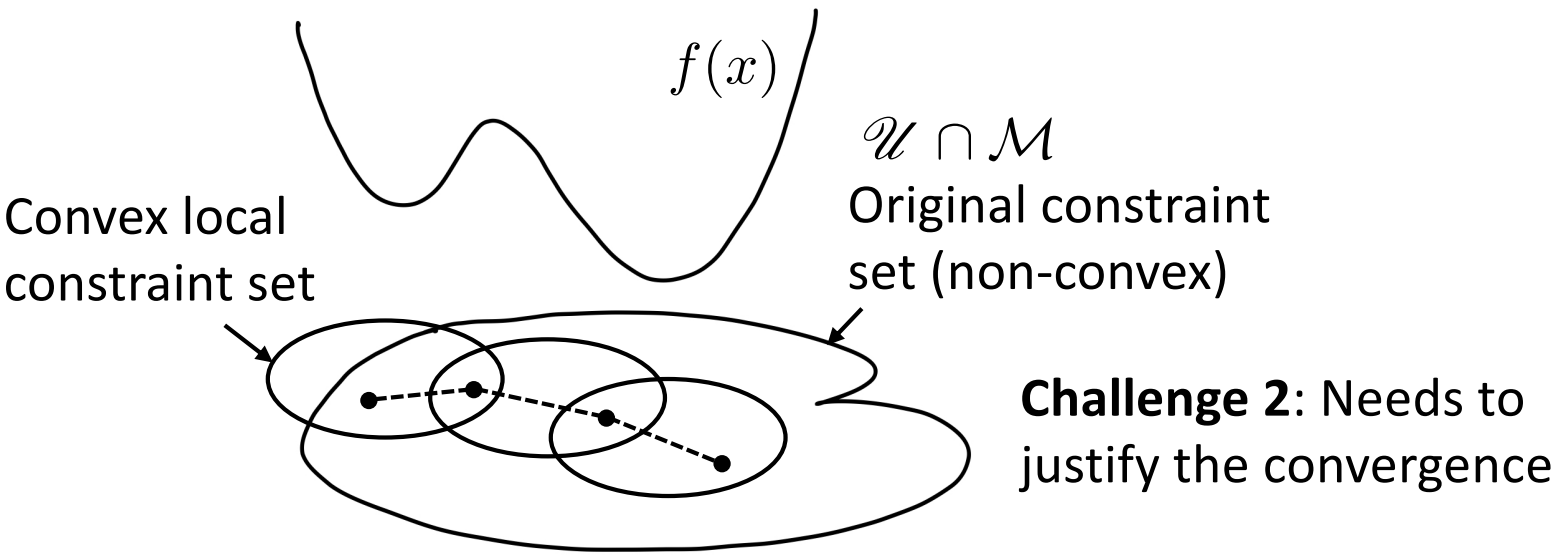
# New challenges: need to dynamically design the convex constraint set $U(\mathrm{x}^{(t)})$

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

$$\begin{array}{ll} \underset{\boldsymbol{y}\in\mathbb{R}^n}{\text{minimize}} & f(\boldsymbol{x}^{(t)}) + \nabla f(\boldsymbol{x}^{(t)})^{\mathrm{T}}\boldsymbol{y} \\ \text{subject to} & \boldsymbol{y} \in U(\boldsymbol{x}^{(t)}) \end{array}$$

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

$O(n)$ complexity or better

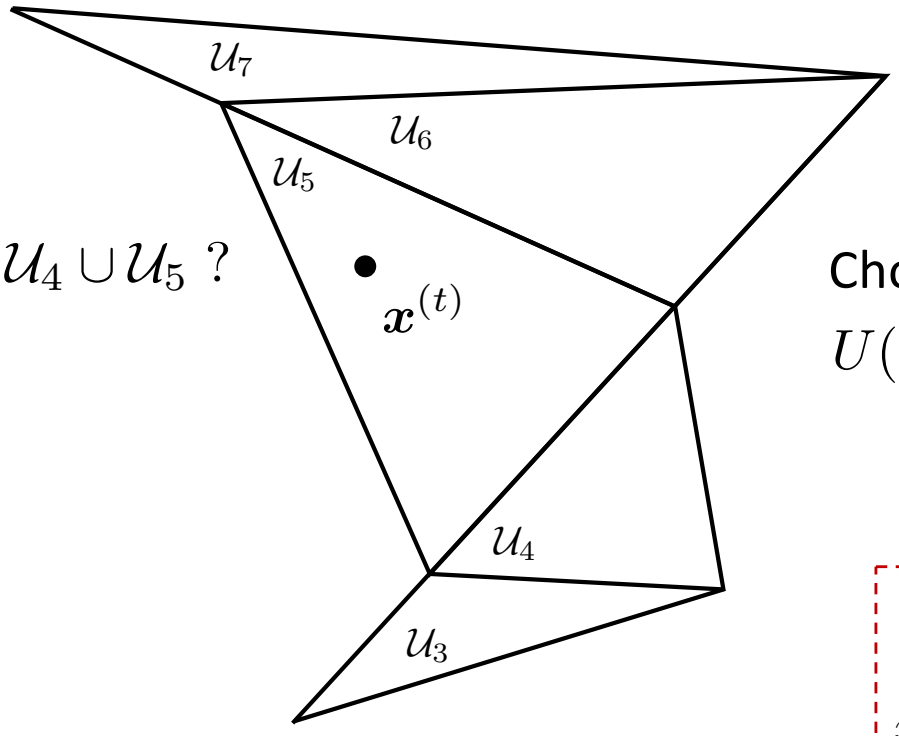



**Challenge 2**: Needs to justify the convergence

# Property: The union of two adjacent components are convex but three are non-convex

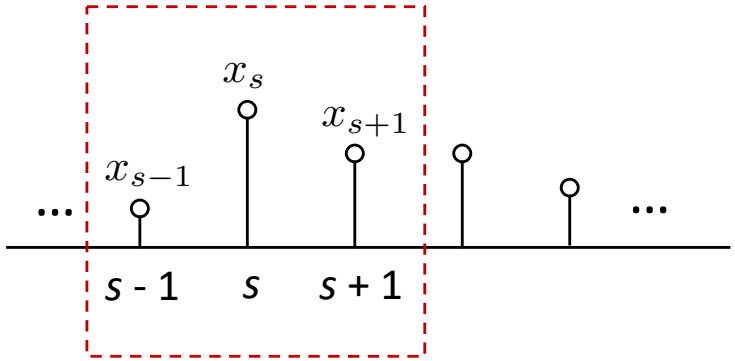
Choice 1:

$U(\boldsymbol{x}^{(t)}) = \mathcal{U}_4 \cup \mathcal{U}_5$ ?




Choice 2:

$U(\boldsymbol{x}^{(t)}) = \mathcal{U}_4 \cup \mathcal{U}_5 \cup \mathcal{U}_6$ ?



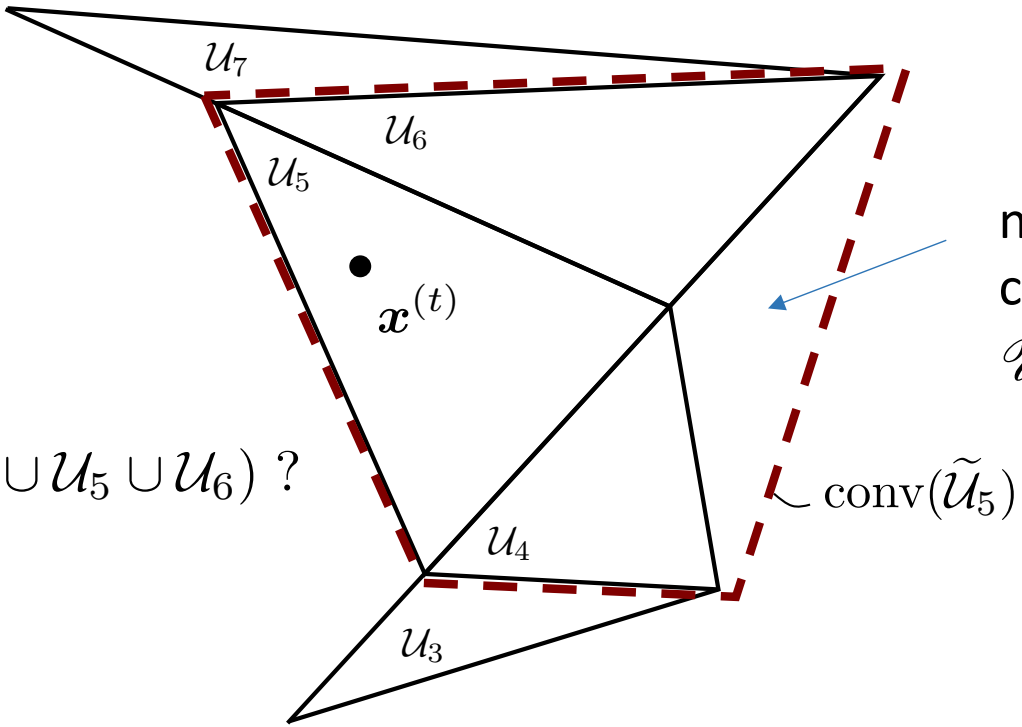

$$\mathcal{U}_s = \left\{ \begin{array}{r} \boldsymbol{x} \in \mathbb{R}^n : \; 0 \leq x_1 \leq x_2 \leq \cdots \leq x_s \\ x_s \geq x_{s+1} \geq \cdots \geq x_n \geq 0 \end{array} \right\}$$

set of unimodal vectors with the *s*th element being the largest

# $U(x^{(t)})$ needs to be convex such that the sub-problems can be solved efficiently

Choice 2':

$$U(\boldsymbol{x}^{(t)}) = \operatorname{conv}(\mathcal{U}_4 \cup \mathcal{U}_5 \cup \mathcal{U}_6)\ ?$$




not included in the constraint set $\mathscr{U} \cap \mathcal{M}$

$$\mathcal{U}_s = \left\{ \begin{array}{l} \boldsymbol{x} \in \mathbb{R}^n : \ 0 \leq x_1 \leq x_2 \leq \cdots \leq x_s \\ \qquad\qquad x_s \geq x_{s+1} \geq \cdots \geq x_n \geq 0 \end{array} \right\}$$

set of unimodal vectors with the *s*th element being the largest

# The update always stay inside the constraint set, i.e., being unimodal

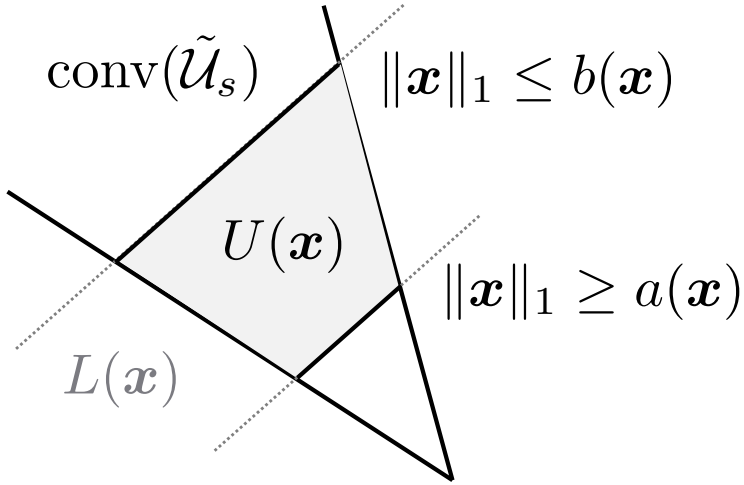
$$U(\boldsymbol{x}) := \mathrm{conv}\big(\widetilde{\mathcal{U}}_{S(\boldsymbol{x})}\big) \cap L(\boldsymbol{x})$$

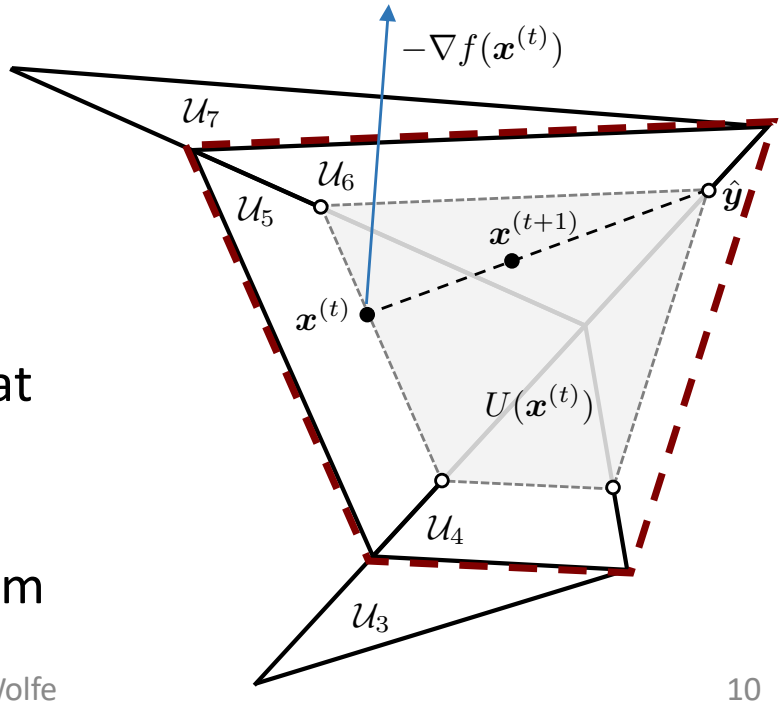
$$L(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{R}^n : a(\boldsymbol{x}) \leq \|\boldsymbol{y}\|_1 \leq b(\boldsymbol{x})\}$$

where

$$a(\boldsymbol{x}) = \min\{1, \hat{\alpha}(\boldsymbol{x})\}\|\boldsymbol{x}\|_1$$
$$b(\boldsymbol{x}) = \max\{1, \hat{\alpha}(\boldsymbol{x})\}\|\boldsymbol{x}\|_1$$

$$\hat{\alpha} = \underset{\alpha \geq 0, \alpha\boldsymbol{x} \in \mathcal{M}}{\mathrm{argmin}} \; f(\alpha\boldsymbol{x})$$




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

→ one being the solution to the LP

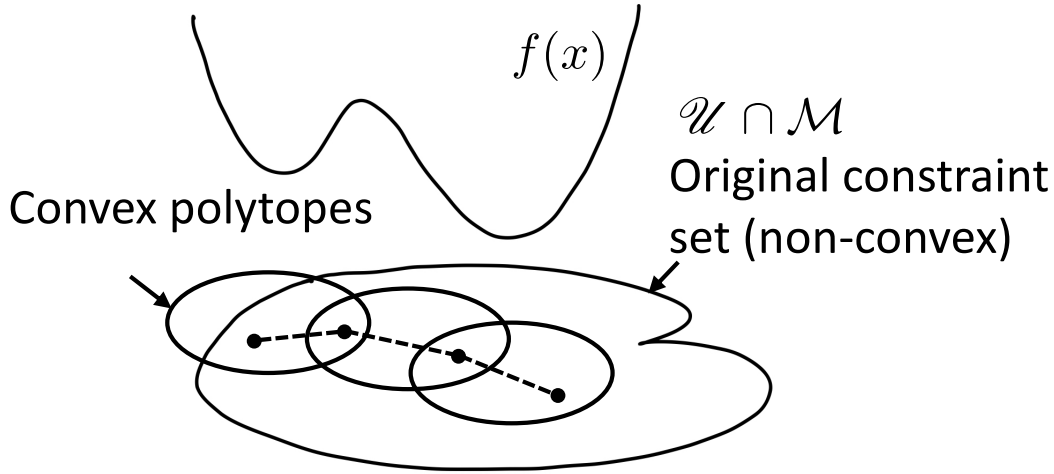
→ found in $2n$ steps by a simplex algorithm

# Why the sub-problem can be computed efficiently?

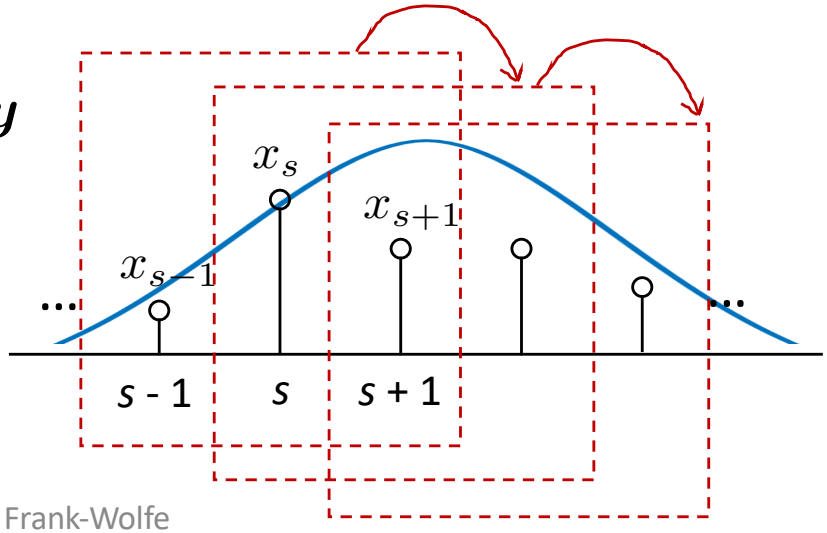
Each $U(\boldsymbol{x}^{(t)})$ is a convex polytope with at most $2n$ extreme points

↓

a simplex method can find the optimal solution via at most $2n$ steps




$$\begin{aligned} \underset{\boldsymbol{y} \in \mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}^{(t)}) + \nabla f(\boldsymbol{x}^{(t)})^{\mathrm{T}} \boldsymbol{y} \\ \text{subject to} \quad & \boldsymbol{y} \in U(\boldsymbol{x}^{(t)}) \end{aligned}$$



# The algorithm converges, and global convergence is also possible

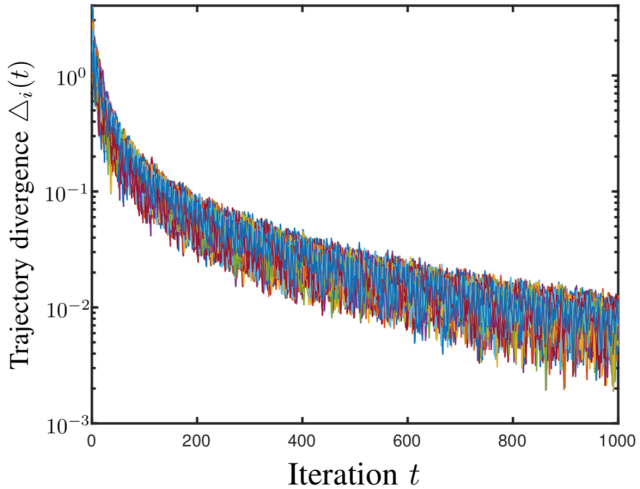
- **Theorem 1**. The algorithm converges under step size $\lambda_t = 2/(t+2)$.
- **Theorem 2**. Under <u>adaptive step size</u>, the gap

$$g(\boldsymbol{x}) = \max_{\boldsymbol{w} \in U(\boldsymbol{x})} -\nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{w} - \boldsymbol{x})$$

  converges to zero at a rate $\mathcal{O}(1/\sqrt{t+1})$.
  - $g(\boldsymbol{x})$ = 0 defines a stationary point
  - $g(\boldsymbol{x})$ gives a lower bound the duality gap
- **Theorem 3**. If f is strictly convex and its critical point $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$ satisfying $f(\boldsymbol{x}_1) < f(\boldsymbol{x}_2)$ for any $\|\boldsymbol{x}_1 - \boldsymbol{x}^\star\| < \|\boldsymbol{x}_2 - \boldsymbol{x}^\star\|$, and in addition, $\boldsymbol{x}^\star \in \mathscr{U} \cap \mathcal{M}$, then the algorithm converges to the <u>global optimal solution</u> $\boldsymbol{x}^* = \boldsymbol{x}^\star$ from any initial point.

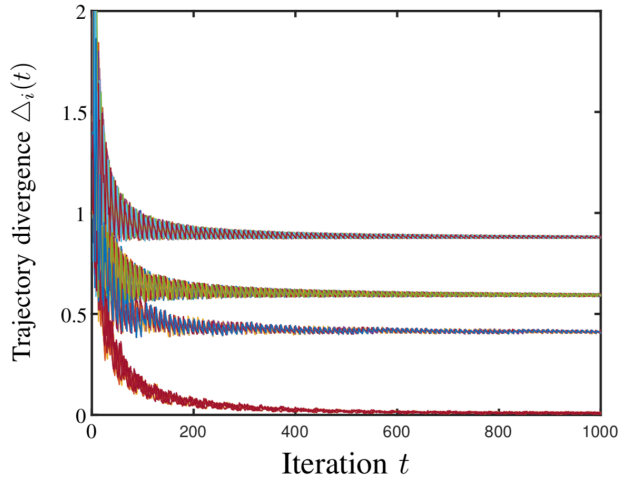
# Example of Estimating an n-dimensional unimodal signal: the convergence

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




High SNR,

Theorem 3 applies, global convergence

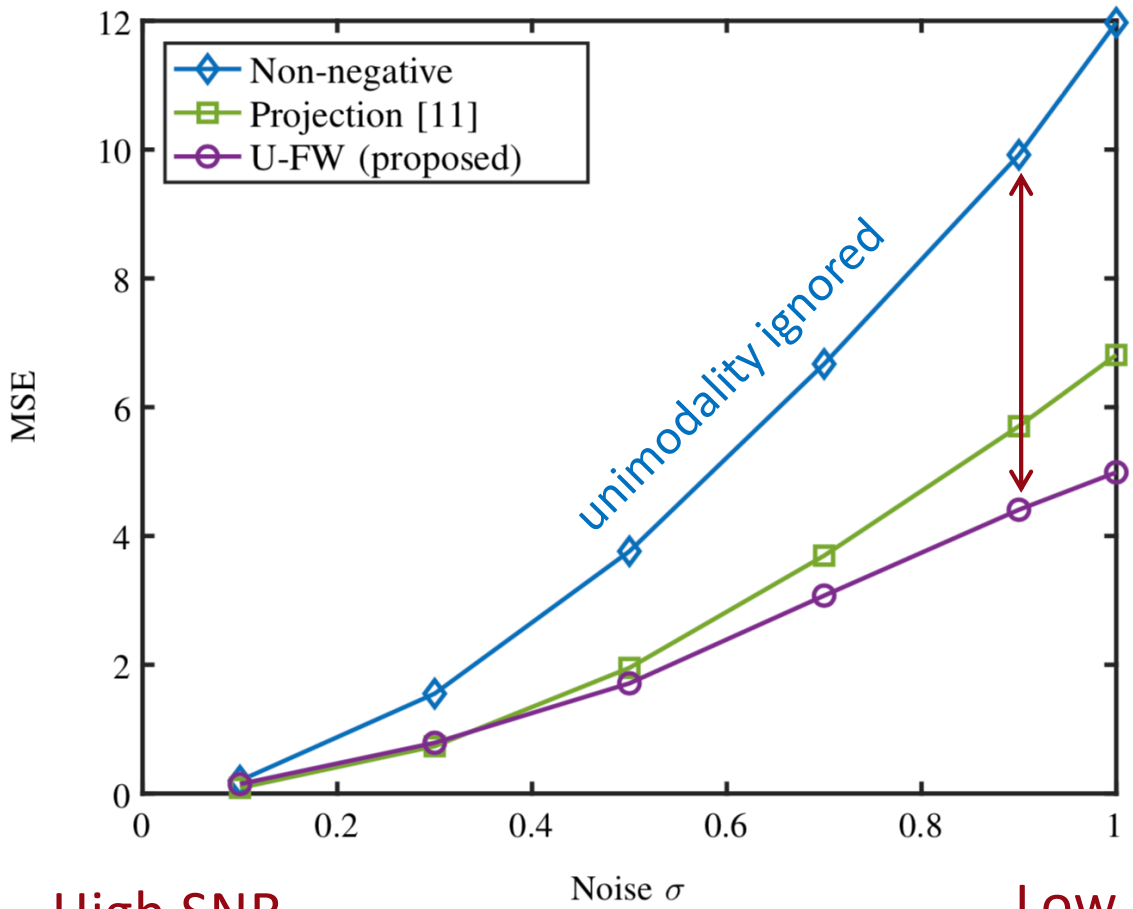



Low SNR, sigma=10, multiple stationary points

# The recovery performance: enforcing unimodality enhances the estimation performance

minimize $f(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{z}\|_2^2$

subject to ***x*** being unimodal (and non-negative)

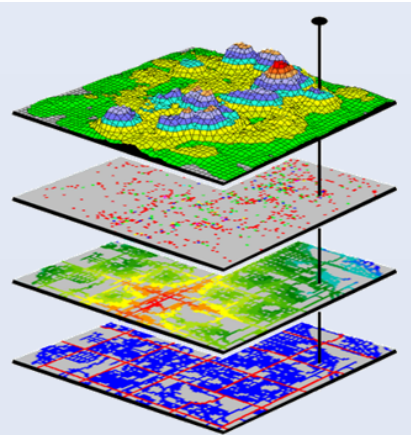



Unimodal projection [Stout'10] + non-negative projection

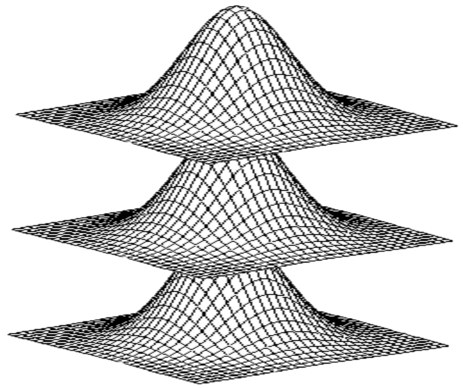
High SNR

Low SNR

# Application to tensor factorization: multimodal data → tensor model → unimodal structure for each layer → unimodal FW algorithm

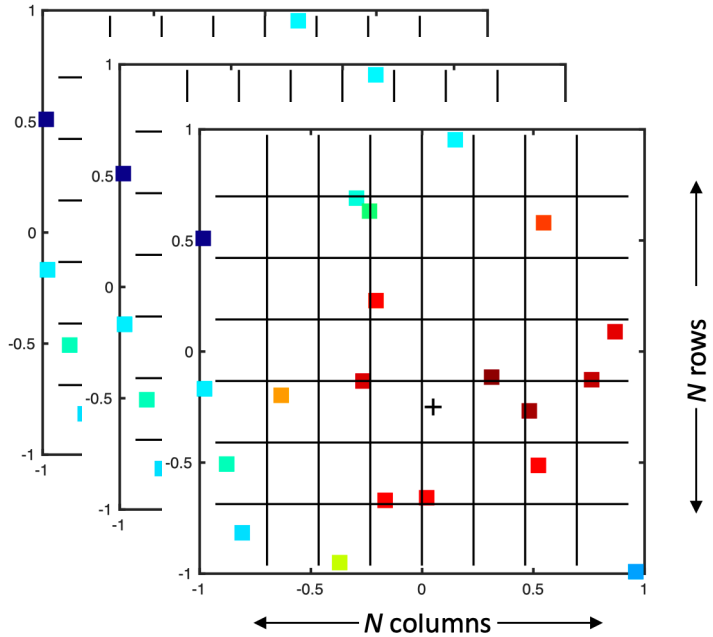


multimodal data for estimating a source



signals at each layer are unimodal (peaks are assumed aligned)

tensor $\mathfrak{X} \in \mathbb{R}^{N \times N \times K}$




sparse observation at each layer $N(\log N)^2 \sim M$

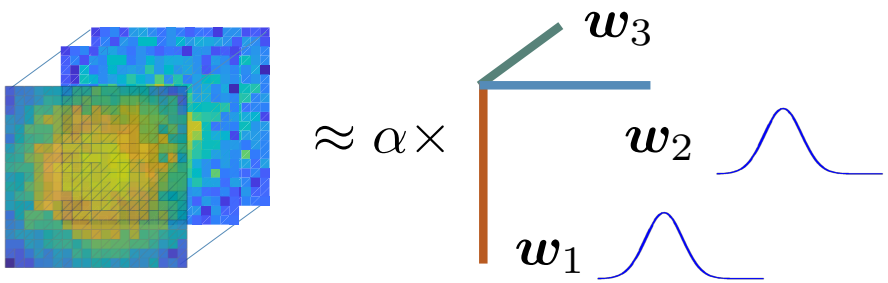
# The dominant vectors from least-squares rank-1 tensor approximation are unimodal

- **Theorem** (Chen&Mitra'18). The optimal solutions $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ of $P_0$ are unimodal, with their <u>peak locations</u> correspond to the <u>source location</u>.

$$\mathscr{P}_0: \quad \underset{\boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{w}_3,\alpha}{\text{minimize}} \quad \|\mathfrak{X} - \alpha \times_1 \boldsymbol{w}_1 \times_2 \boldsymbol{w}_2 \times_3 \boldsymbol{w}_3\|_{\mathrm{F}}^2 \quad \text{(full observation)}$$

$$\text{subject to} \quad \alpha > 0, \|\boldsymbol{w}_1\| = \|\boldsymbol{w}_2\| = \|\boldsymbol{w}_3\| = 1$$

- $\|\mathfrak{X}\|_{\mathrm{F}}^2 \triangleq \sum_i \sum_j \sum_k \mathfrak{X}(i,j,k)^2$
- $\mathfrak{X} \times_p \boldsymbol{A}$ denotes the mode-*p* multiplication



# Tensor factorization enforcing unimodality constraints

$$\mathscr{P}_{\mathrm{UTF}}: \quad \underset{\alpha, \boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3}{\text{minimize}} \quad \|\boldsymbol{\mathcal{W}} \odot (\boldsymbol{\mathcal{X}} - \alpha \times_1 \boldsymbol{w}_1 \times_2 \boldsymbol{w}_2 \times_3 \boldsymbol{w}_3)\|_{\mathrm{F}}^2$$

$$\text{subject to} \quad \alpha > 0, \|\boldsymbol{w}_1\|_1 = \|\boldsymbol{w}_2\|_1 = \|\boldsymbol{w}_3\|_1 = 1.$$

$$\underline{\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathscr{U}}$$

**Proposed modified Frank-Wolfe algorithm** applies to update $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ alternatively using the gradients:

$$\frac{1}{2}\frac{\partial f}{\partial \boldsymbol{w}_1} = -a(\boldsymbol{X}_{(1)}^{\mathrm{w}})^{\mathrm{T}}(\boldsymbol{w}_3 \otimes \boldsymbol{w}_2) + a^2\left[\boldsymbol{W}_{(1)}^{\mathrm{T}}(\boldsymbol{w}_3^2 \otimes \boldsymbol{w}_2^2)\right] \odot \boldsymbol{w}_1$$

$$\frac{1}{2}\frac{\partial f}{\partial \boldsymbol{w}_2} = -a(\boldsymbol{X}_{(2)}^{\mathrm{w}})^{\mathrm{T}}(\boldsymbol{w}_3 \otimes \boldsymbol{w}_1) + a^2\left[\boldsymbol{W}_{(2)}^{\mathrm{T}}(\boldsymbol{w}_3^2 \otimes \boldsymbol{w}_1^2)\right] \odot \boldsymbol{w}_2$$

$$\frac{1}{2}\frac{\partial f}{\partial \boldsymbol{w}_3} = -a(\boldsymbol{X}_{(3)}^{\mathrm{w}})^{\mathrm{T}}(\boldsymbol{w}_2 \otimes \boldsymbol{w}_1) + a^2\left[\boldsymbol{W}_{(3)}^{\mathrm{T}}(\boldsymbol{w}_2^2 \otimes \boldsymbol{w}_1^2)\right] \odot \boldsymbol{w}_3$$

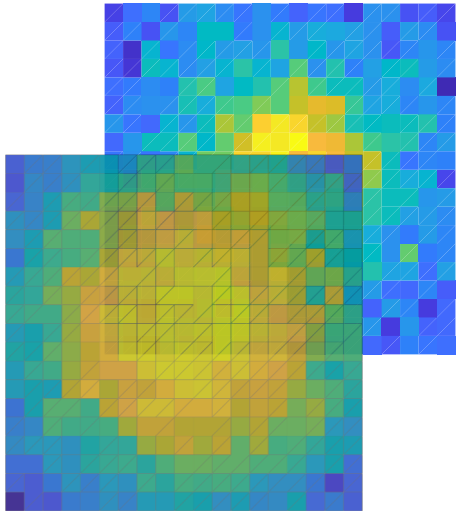
# Localization: the same source emitting two types of signals being captured by RSS and TOA sensors

- How data is generated:
  50% for RSS of the EM signal

$$P_{\mathrm{dB}}(d) = 70 - 36 \times \log_{10}(\max\{10, d\}) + \mathcal{N}(0, \sigma_{\mathrm{s}}^2)$$

  50% for TOA of the acoustic signal $\sigma_{\mathrm{s}} = 10 \text{ dB}$

$$t(d) = \frac{d}{340 \text{ m/s}} + \mathcal{N}(0, \sigma_{\mathrm{t}}^2) \qquad \sigma_{\mathrm{t}} = 100 \text{ ms}$$

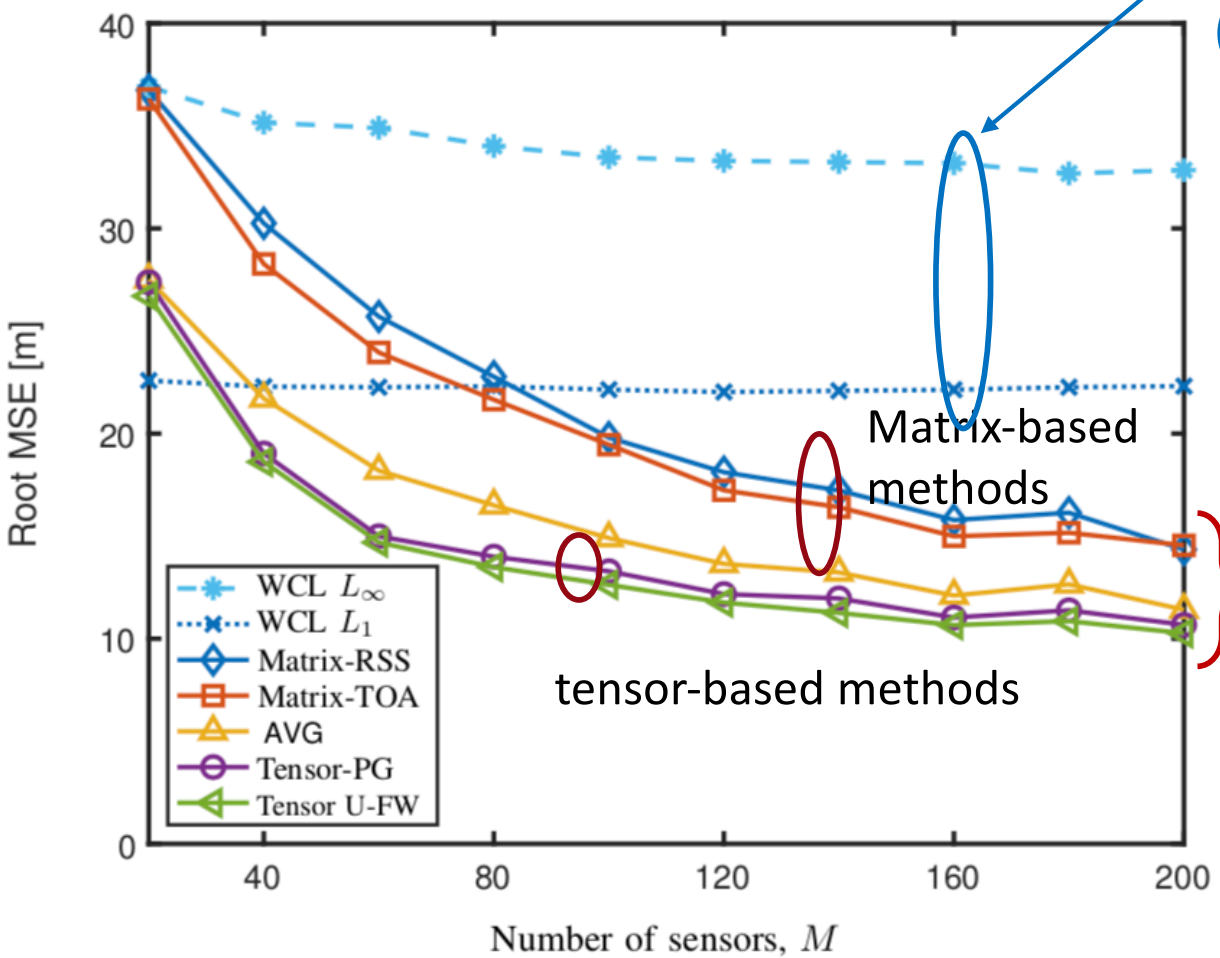


- Preprocessing: Data normalization

$$h_1(d) = \exp(-\beta_1 10^{-P_{\mathrm{dB}}(d)/10}) \qquad h_2(d) = \exp(-\beta_2 t(d)^2)$$

  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



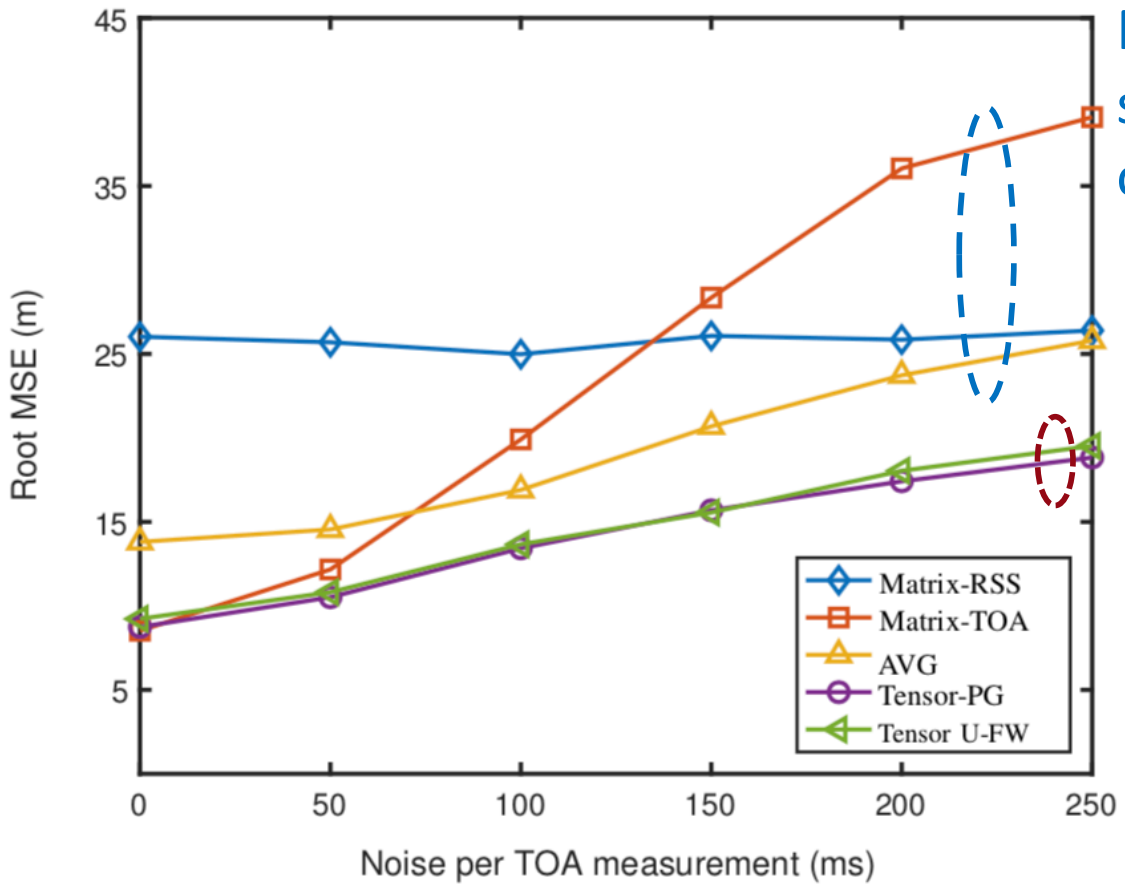

Unimodaityl-non-aware (weighted centroid)

$$\hat{\mathbf{s}}_{\mathrm{RSS}} = \frac{\sum_{m \in \mathcal{R}_{\mathrm{RSS}}} \varrho^{(m)} \boldsymbol{z}^{(m)}}{\sum_{m \in \mathcal{R}_{\mathrm{RSS}}} \varrho^{(m)}}$$

Enforcing unimodality

# Tensor factorization strategy fuses multimodal data better

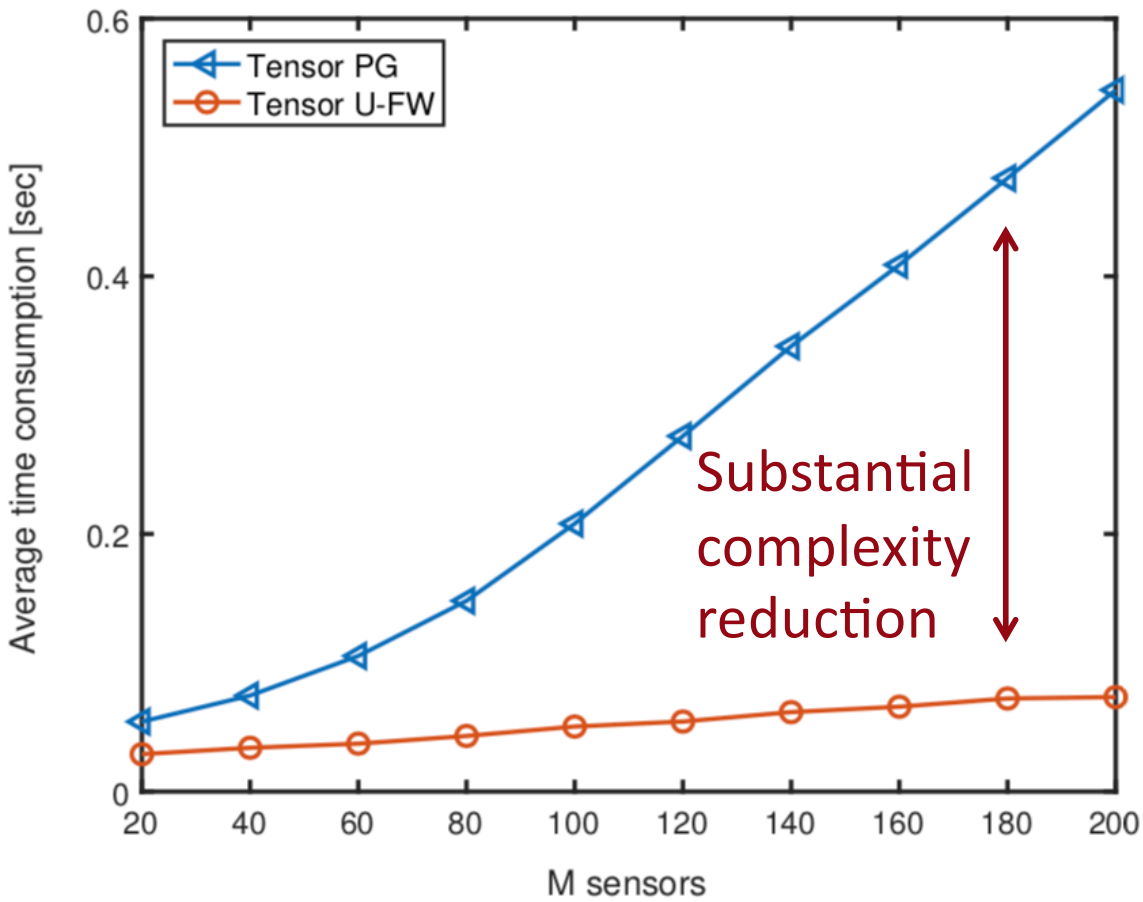
When one of the signal modes (TOA signal) is corrupted…




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

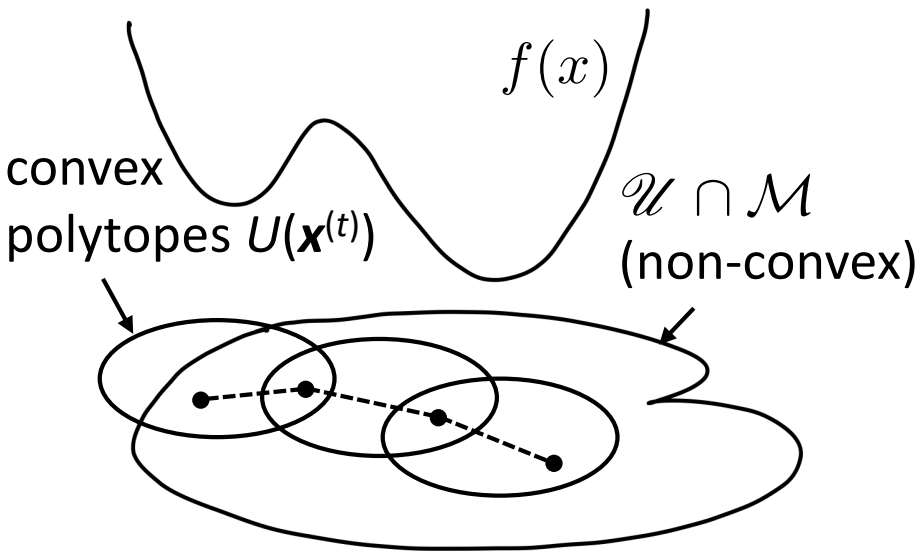



Projected gradient based on unimodal regression ([Stout'08], state-of-the-art)

Proposed Unimodal Frank-Wolfe

# In conclusion, we developed a unimodal-FW algorithm to solve unimodality-constrained problems

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathscr{U} \cap \mathcal{M} \end{aligned}$$




- 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


- [Bro&Sidiropoulous'98] R. Bro and N. D. Sidiropoulos, "Least squares algorithms under unimodality and non-negativity constraints," *Journal of Chemometrics*, vol. 12, no. 4, pp. 223–247, 1998.
- [Stout'08] Q. F. Stout, "Unimodal regression via prefix isotonic regression," *Computational Statistics & Data Analysis*, vol. 53, no. 2, pp. 289–297, 2008.
- [Németh&Németh'10] A. Németh and S. Németh, "How to project onto an isotone projection cone," *Linear Algebra and its Applications*, vol. 433, no. 1, pp. 41–51, 2010.
- [Gunn&Dunson'05] L. H. Gunn and D. B. Dunson, "A transformation approach for incorporating monotone or unimodal constraints," *Biostatistics*, vol. 6, no. 3, pp. 434–449, 2005.
- [Köllman et.al.'14] C. Köllmann, B. Bornkamp, and K. Ickstadt, "Unimodal regression using bernstein–schoenberg splines and penalties," *Biometrics*, vol. 70, no. 4, pp. 783–793, 2014.
- [Chen&Mitra'17] J. Chen and U. Mitra, "Unimodality-Constrained Matrix Factorization for Non-Parametric Source Localization," 2017, submitted to *IEEE Trans. Signal Process.*, preprint arXiv:1711.07457.
- [Chen&Mitra'18] J. Chen and U. Mitra, "A tensor decomposition technique for source localization from multimodal data," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Calgary, Alberta, Canada, Apr. 2018.