Homotopy Analysis for Tensor PCA

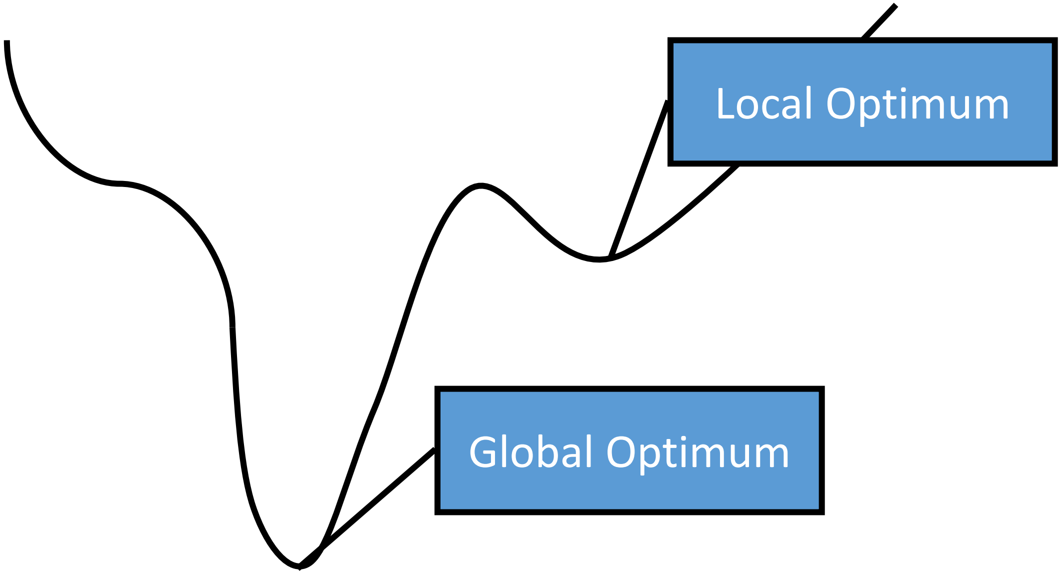
Yuan Deng

Duke University

Joint work with Anima Anandkumar, Rong Ge, Hossein Mobahi

Non-convex Optimization

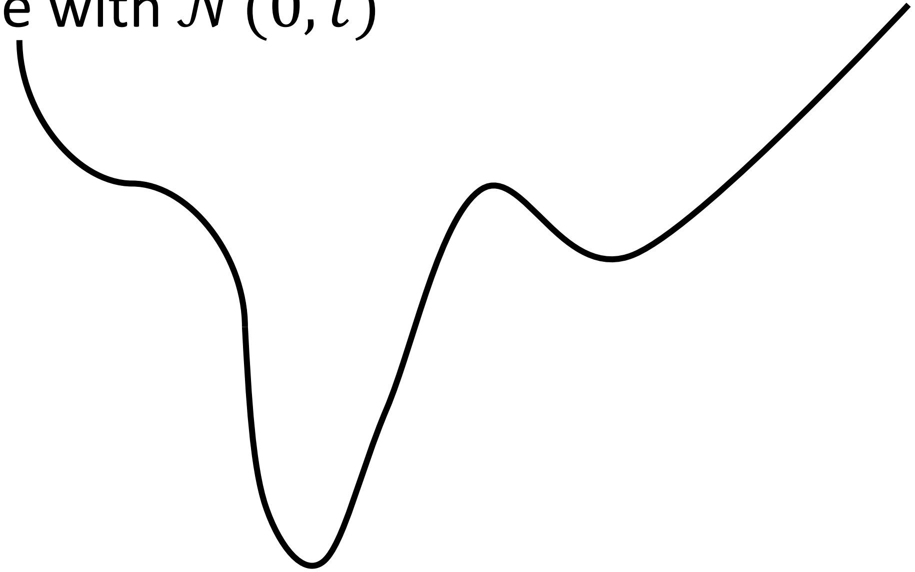
- Optimizing smooth function f(x).



How to get rid of local optima?

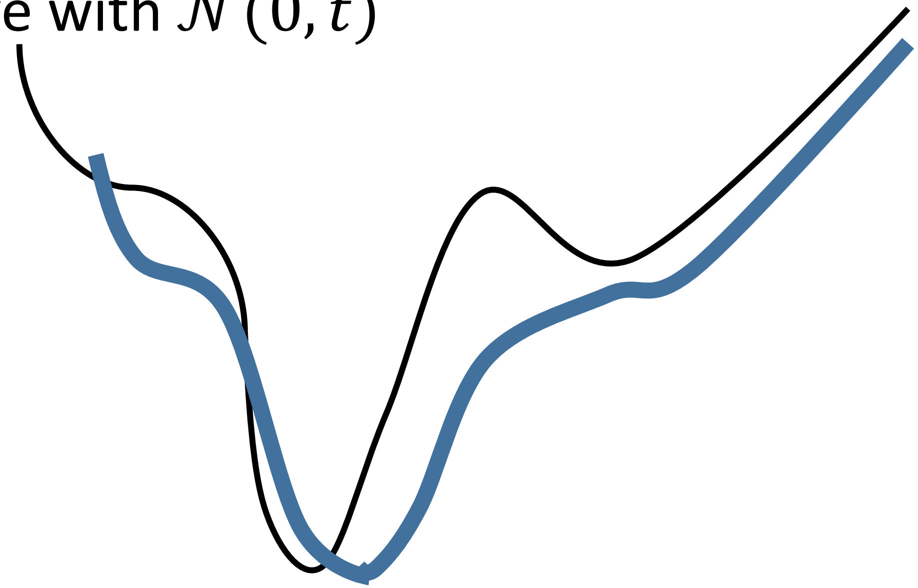
Gaussian Smoothing

- Idea: Smooth the function
  - convolve with $\mathcal{N}(0, t)$



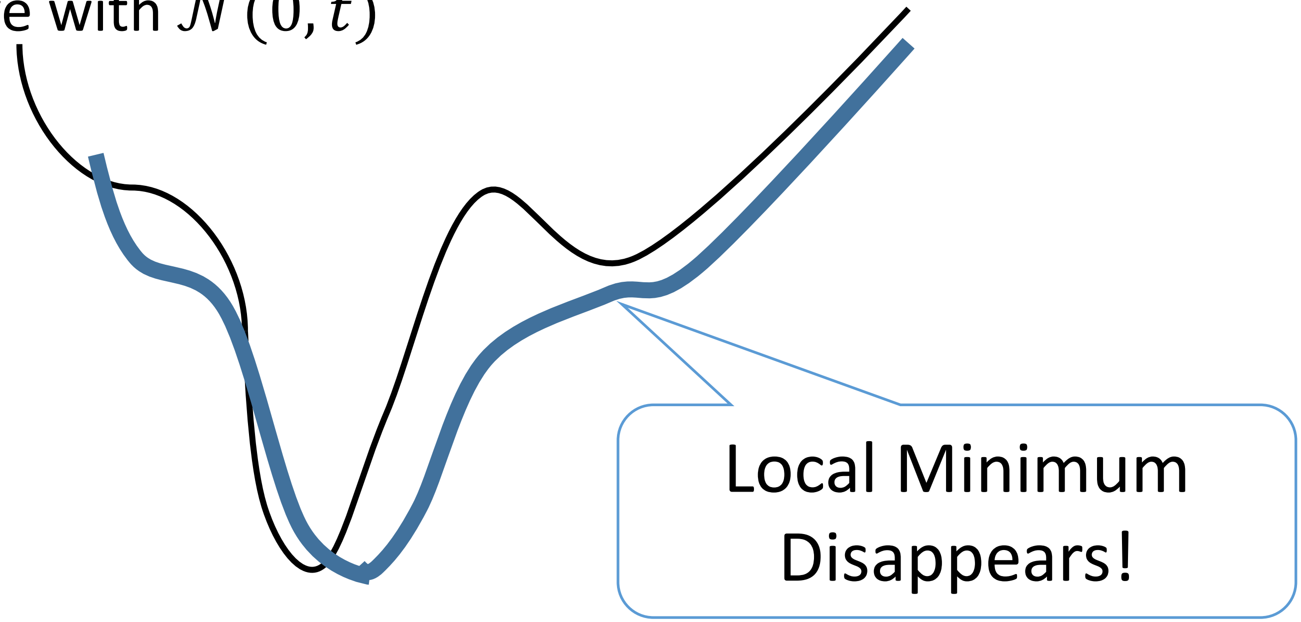
Gaussian Smoothing

- Idea: Smooth the function
  - convolve with $\mathcal{N}(0, t)$



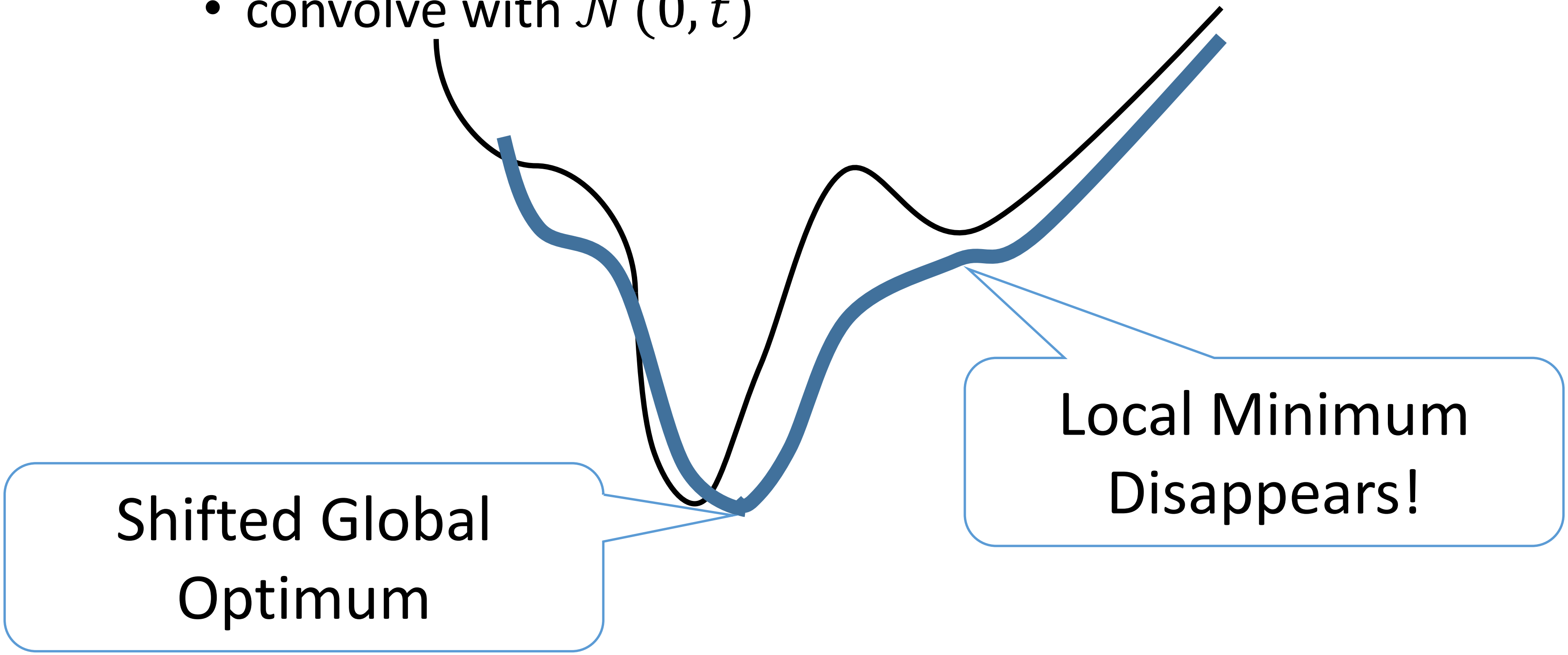
Gaussian Smoothing

- Idea: Smooth the function
  - convolve with $\mathcal{N}(0, t)$



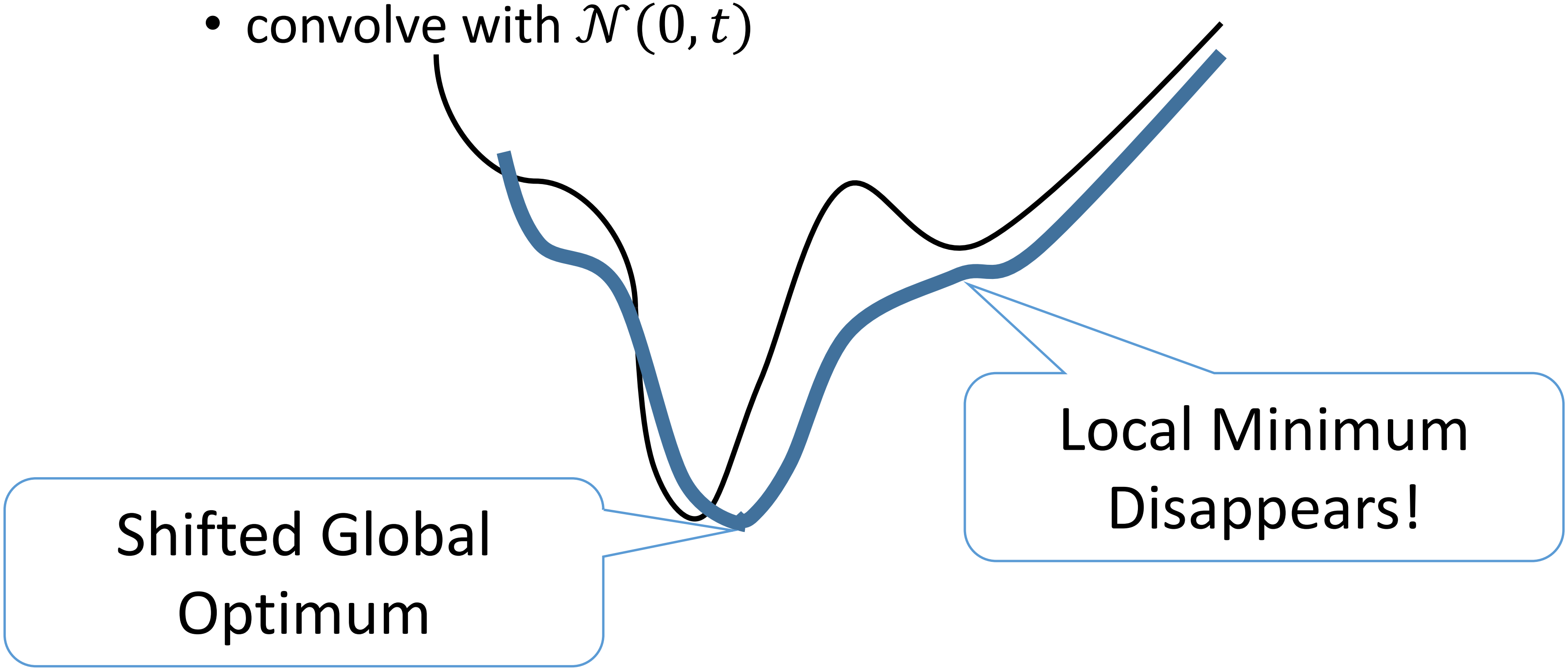
Gaussian Smoothing

- Idea: Smooth the function
  - convolve with $\mathcal{N}(0, t)$



Gaussian Smoothing

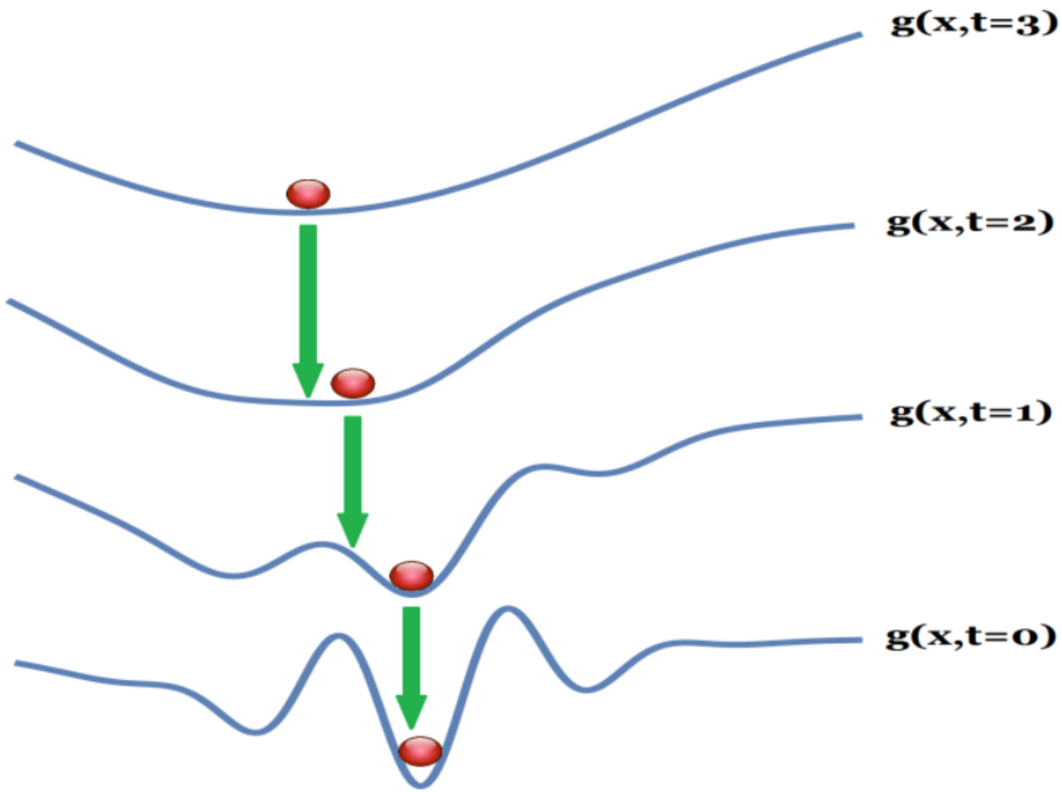
- Idea: Smooth the function
  - convolve with $\mathcal{N}(0, t)$



- How to decide how much to smooth?
- How to recover the original global optium?

Homotopy Method

- Try all level of smoothing!



Homotopy Method

Computer Vision

- image deblurring [Boccuto et al., 2002]
- image restoration [Nikolova et al., 2010]
- optical flow [Brox & Malik, 2011]

Clustering [Gold, 1994]

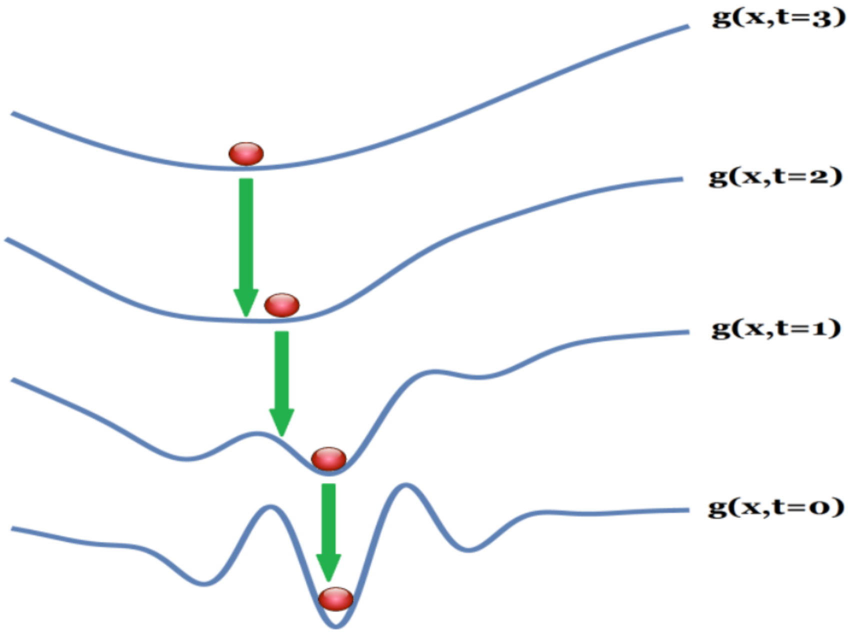
Graph matching [Zaslavskiy et al., 2009]

- ***No theoretical guarantees*** on the solution
  - too restrictive [Mobahi and Fisher III, 2015]
  - difficult to check [Hazan et al., 2016]

Homotopy Method

- **Handcrafted** the choice of smoothing levels
- **Slow**: Local search is repeated

  for each smoothing level

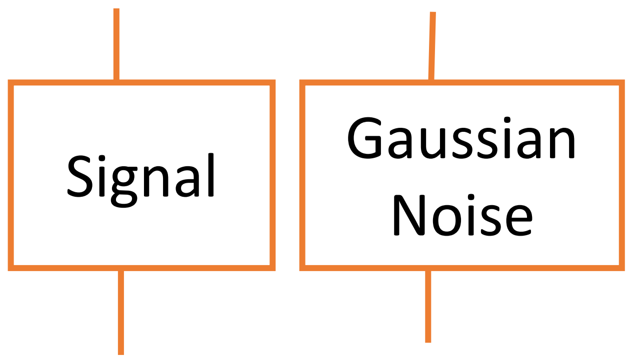


Tensor PCA [Richard and Montanari 2014]

Probabilistic model for PCA

$v \in \mathbb{R}^d, \tau \geq 0$ is the signal-to-noise ratio

$$M = \tau \boldsymbol{v} \boldsymbol{v}^\top + A$$



Tensor PCA: $T = \tau \boldsymbol{v} \otimes \boldsymbol{v} \otimes \boldsymbol{v} + A$

Objective:

- Design an efficient algorithm for as small τ as possible

Previous Work

- [Richard & Montanari 2014] Can find v when $\tau = \Omega(d)$ in poly time, and $\tau = \Omega(\sqrt{d})$ in exp. time.

- [Hopkins, Shi & Steurer 2015] Sum-of-Squares technique, can find v when $\tau = \widetilde{\Omega}(d^{3/4})$ in poly time
  - Basic Sum-of-Squares algorithm is very slow.
  - Running time can be improved $\widetilde{\Omega}(d^3)$, nearly linear

Our Results

Method	Bound on τ	Time	Extra Space
Ours	$\widetilde{\Omega}(d^{3/4})$	$\widetilde{\Omega}(d^3)$	$O(d)$
State-of-Art	$\widetilde{\Omega}(d^{3/4})$	$\widetilde{\Omega}(d^3)$	$O(d^2)$

Guarantee matches best known result

Better convergence rate when τ is closed to $d^{3/4}$

One of the first results on provably analyzing homotopy method

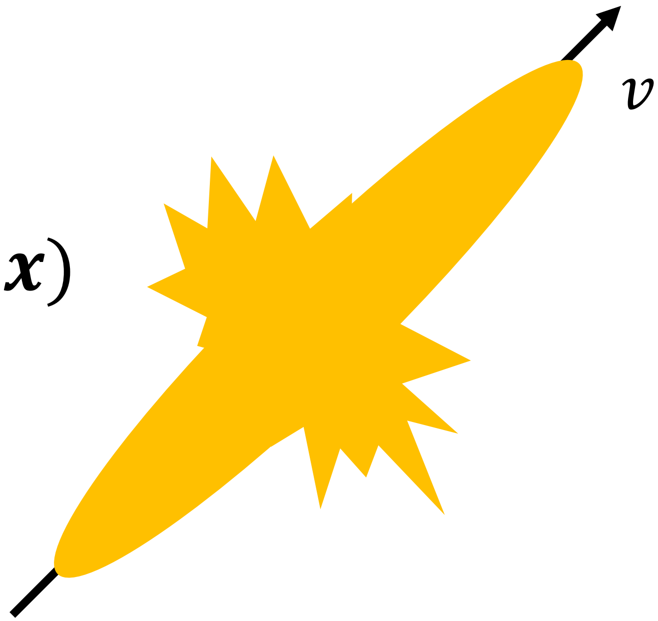
Optimization for tensor PCA

- Recall: for matrix PCA, we optimize

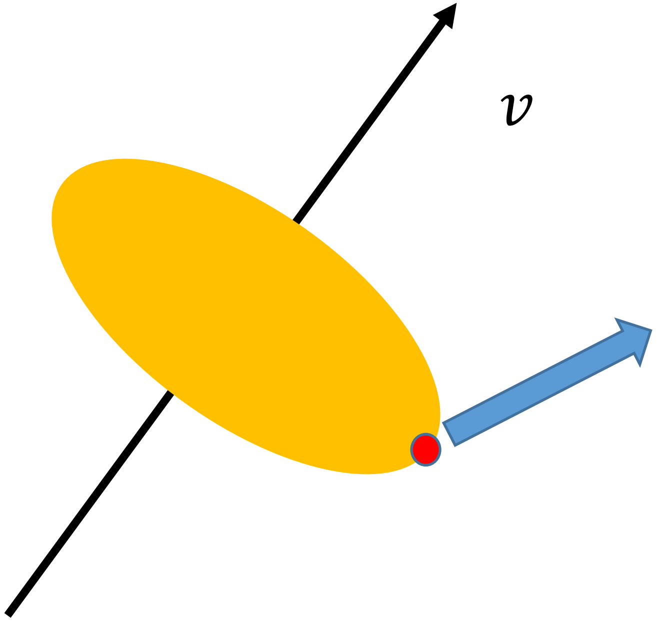
$$\max \boldsymbol{x}^\top M \boldsymbol{x} = \tau \langle \boldsymbol{v}, \boldsymbol{x} \rangle^2 + \boldsymbol{x}^\top A \boldsymbol{x}$$
$$\|\boldsymbol{x}\| = 1$$

- For tensor PCA, we optimize

$$\max T(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}) = \tau \langle v, \boldsymbol{x} \rangle^3 + A(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x})$$
$$\|\boldsymbol{x}\| = 1$$



Infinite Smoothing



unique optimum x^* :

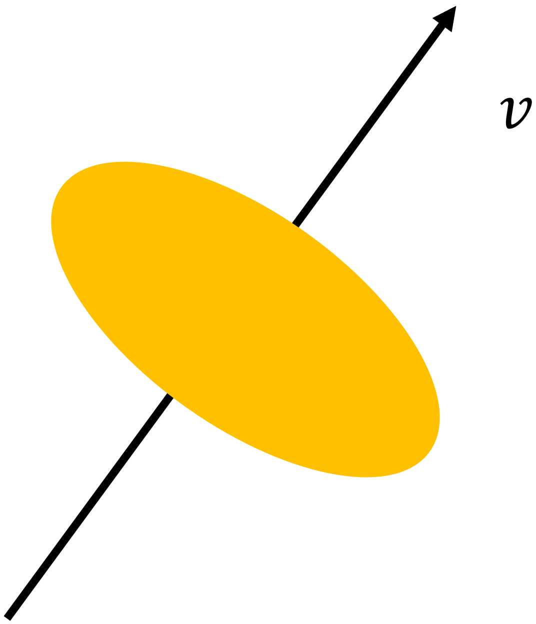
correlation $\tau / d = \Omega(d^{-0.25})$

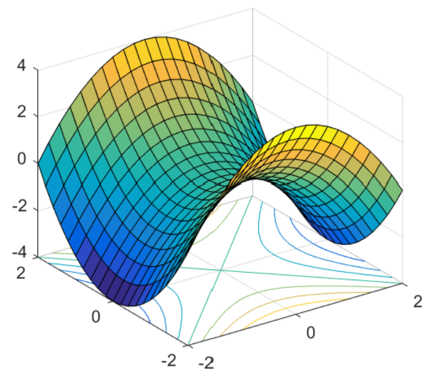
[random unit vector : $d^{-0.5}$]

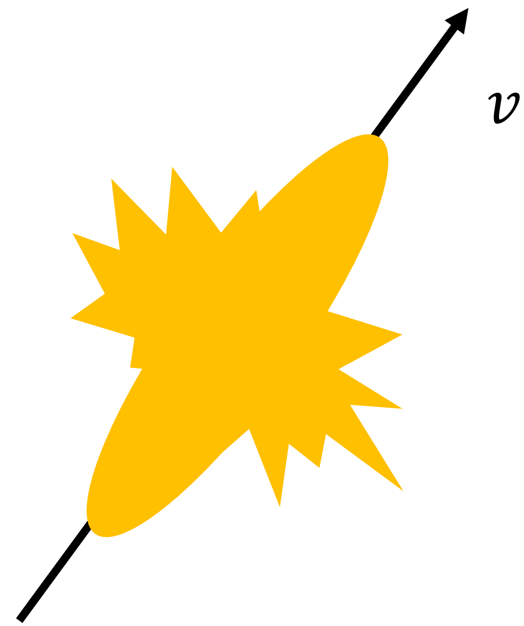
$t = \infty$

Phase Transition in Homotopy Method

- Lemma*: there is a threshold θ,



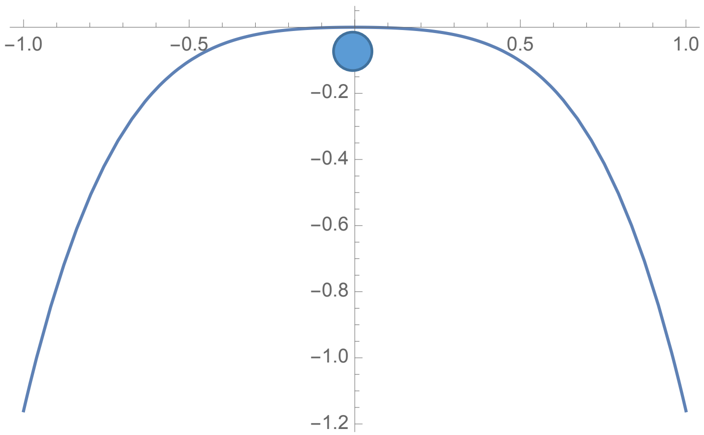




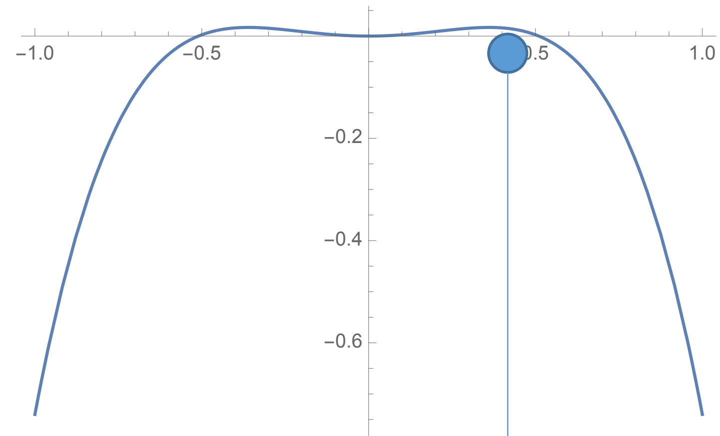
$t > \theta$ $\quad$ $t = \theta$ $\quad$ $t < \theta$

- If using infinite steps, i.e., continuously $\infty \to 0$
  - $t > \theta$, $||x^t - x^*|| \le o(1)||x^*||$
  - $t < \theta$, $\langle x^t, v \rangle = \Omega(1)$

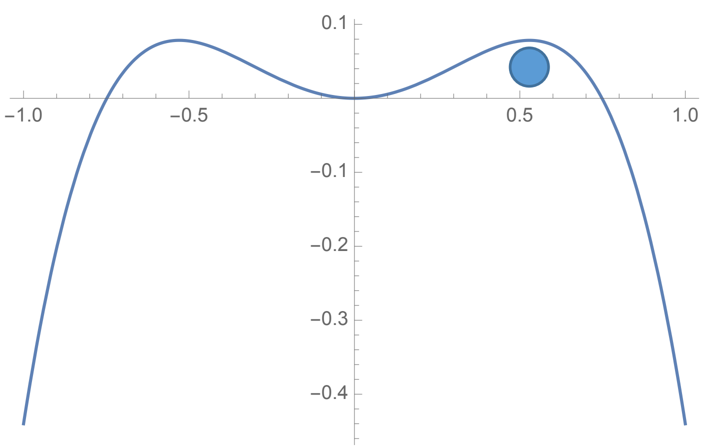
Phase Transition $f(x) = -x^4 + 0.8x^2$



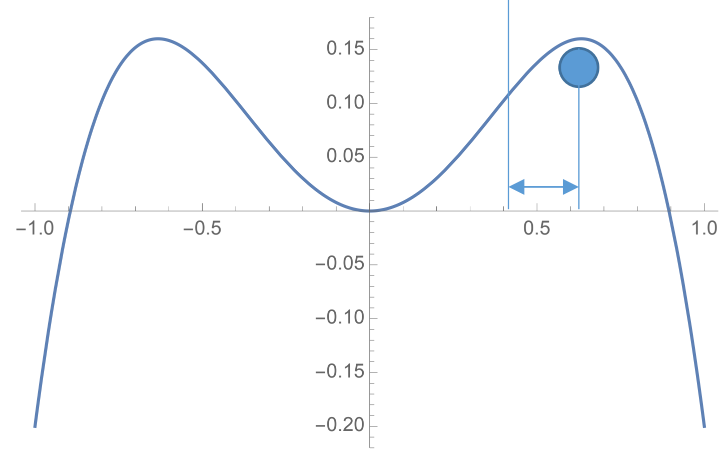
$g(x, 0.4)$



$g(x, 0.3)$



$g(x, 0.2)$



$g(x, 0)$

Phase Transition

- If using infinite steps, i.e., continuously $\infty \to 0$
  - $t > \theta, ||x^t - x^*|| \leq o(1)||x^*||$ $t_1 = \infty$
  - $t < \theta, \langle x^t, v \rangle = \Omega(1)$ $t_2 = 0$

Input: Tensor $\boldsymbol{T} = \tau \cdot \boldsymbol{v}^{\otimes 3} + \boldsymbol{A}$;
Output: Approximation of $\boldsymbol{v}$;
$m = O(\log \log n)$;
$\forall\, j, \boldsymbol{x}_j^0 = \sum_i \boldsymbol{T}_{iij} + \boldsymbol{T}_{iji} + \boldsymbol{T}_{jii}$;
$\boldsymbol{x}^0 = \boldsymbol{x}^0 / \|\boldsymbol{x}^0\|$;
for $k = 0$ *to* m **do**
    $\boldsymbol{x}^{k+1} = \boldsymbol{T}(\boldsymbol{x}^k, \boldsymbol{x}^k, :) + \boldsymbol{T}(\boldsymbol{x}^k, :, \boldsymbol{x}^k) + \boldsymbol{T}(:, \boldsymbol{x}^k, \boldsymbol{x}^k)$;
    $\boldsymbol{x}^{k+1} = \boldsymbol{x}^{k+1} / \|\boldsymbol{x}^{k+1}\|$;
end
return $\boldsymbol{x}^m$;

Infinite smoothing

Power Method at 0 smoothing

Conclusions

- Homotopy method gives near-optimal results for tensor PCA.
- Possible to analyze non-convex functions even when they really have bad local optima.

Open Problems

- More examples of Homotopy method?
  - When the tensor has higher rank?
- General results for effects of smoothing
  - What kind of local optima will disappear?
- Different way of smoothing/regularization?

Thank You!