Tensor principal component analysis via sum-of-squares proofs

Sam Hopkins
Cornell

Jonathan Shi
Cornell

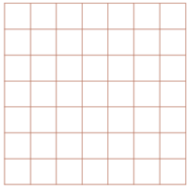
David Steurer
Cornell

COLT – Paris, July 2015

principal component analysis (PCA)

vanilla PCA

2-wise correlation data
= matrix



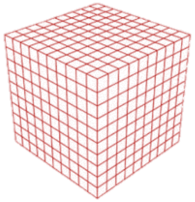
- basic data analysis technique
- given noisy pairwise correlation data $A \in \mathbb{R}^{n \times n}$, find direction of maximum empirical variance

$$\text{maximize } \langle x, Ax \rangle \text{ over all unit vectors } x \in \mathbb{R}^n$$

- computationally efficient (take x to be top eigenvector of A)

variants of PCA

3-wise correlation data
= 3-tensor



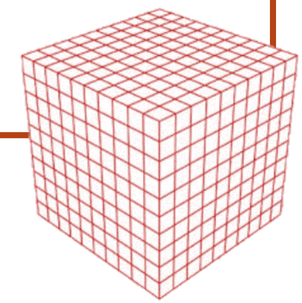
- restrict to sparse directions (SPARSE PCA) or exploit higher-order correlation data $A \in \mathbb{R}^{n \times n \times \cdots \times n}$ (TENSOR PCA)
- better statistical properties in important applications; huge body of works
- *but:* **computationally challenging** (NP-hard in worse case; unclear complexity in stochastic setting)

tensor principal component analysis

(k,τ)-stochastic model [Montanari-Richard]

given k-order tensor A as below, recover v (approximately)

$$A = \tau \cdot v^{\otimes k} + Z \in \mathbb{R}^{n^k} \text{ with } Z \sim N(0,1)^{\otimes k}$$



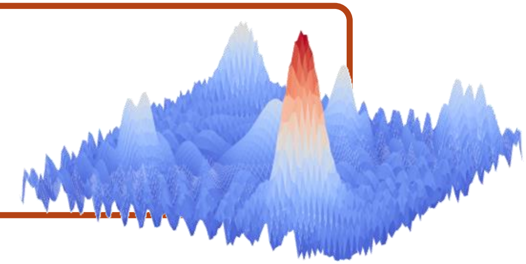
signal-to-noise ratio

noise: random tensor

signal: rank-1 tensor of unit vector $v \in \mathbb{R}^n$

maximum likelihood estimation (MLE)

maximize $\langle A, x^{\otimes k} \rangle$ over all unit vectors $x \in \mathbb{R}^n$

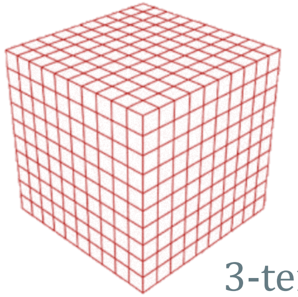


$k = 2$: computationally efficient (eigenvalue problem; even in worst case)

$k = 3$: appears to capture difficulty of general k in stochastic model
(also NP-hard in worst case, but no bearing on stochastic model)

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



previous results [Montanari-Richard=MR]

information-theoretic recovery

(*) works as long as $\tau \geq \tilde{O}(n^{1/2})$ (tight)

computational recovery

MR algorithm: reshape A to n^2-by-n matrix; output top right singular vector

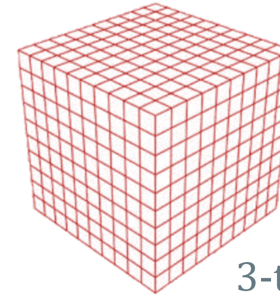
theoretical guarantee: algorithm works as long as $\tau \geq \tilde{O}(n)$

empirical performance: algorithm works as long as $\tau \geq \tilde{O}(n^{3/4})$

tension: theoretical analysis of MR *tight* in many ways **but** empirical performance *should be predictive* for mathematical truth (average-case problem & large input sizes)

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



3-tensor

this work

techniques: sum-of-squares meta-algorithm & proof system;

powerful general approach to unsupervised learning

[Barak-Kelner-S.'12+15,Potechin-Meka-Wigderson'15, Barak-Moitra, Ge-Ma, Ma-Wigderson,…]

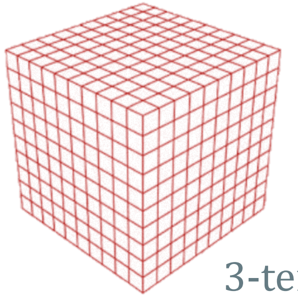
recovery guarantee: theoretical analysis matches empirical performance of MR,
$\tau \gg n^{3/4}$ — one algorithm very similar to MR

nearly-linear time: informed by theoretical analysis; exploit knowledge about eigenvalues to speed up eigenvector computation

lower bounds: rule out better recovery guarantees by algorithms based on broad set of techniques (deg-4 sum-of-squares proof system)

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



relaxation & rounding approach

relaxation: tractable (convex) optimization problem associated with (*); optimal value gives upper bound on optimal value of (*)

rounding: transform solution for relaxation to solution for (*) with approximately same objective value

failure of this approach: for $n^{3/4} \ll \tau \ll n$,

opt. value of MR relaxation (top singular value of A)
is far from opt. value of (*)

but (empirically)

opt. solution of MR relaxation
is close to opt. solution of (*)

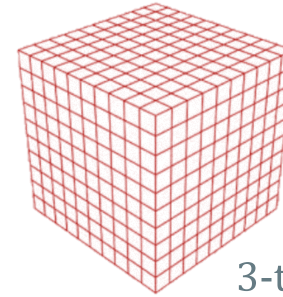
→ ***no rounding analysis possible*** *(in the usual sense)*

our explanation *(for variant of MR relaxation)*

second-order effect in opt. value of relaxation drives recovery

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



sum-of-squares upper bounds

warm-up: upper bounds for homogeneous n-var. deg.-4 polynomial $p(x)$

consider affine linear subspace $H_{p(x)}$ of *matrix representations* of $p(x)$

$$H_{p(x)} \stackrel{\text{def}}{=} \left\{ P \in \mathbb{R}^{n^2 \times n^2} \,\middle|\, p(x) = \langle x^{\otimes 2}, P x^{\otimes 2} \rangle \right\}$$

$\lambda_{\max}(P) = \max_{\|y\|=1} \langle y, Py \rangle$

then, $\max_{\|x\|=1} p(x) \le \lambda_{\max}(P)$ for every $P \in H_{p(x)}$

→ find *best upper bound* $\lambda_{\max}(P)$ with $P \in H_{p(x)}$ *(semidefinite programming)*

deg-d sum-of-squares upper bounds for general polynomial $p(x)$

find *best upper bound* $\lambda_{\max}(P)$ with

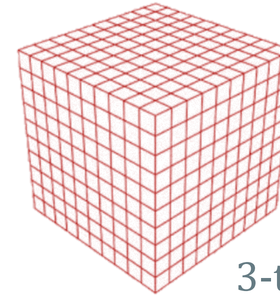
$$P \in \bigcup_{\deg q(x) \le d-2} H_{p(x)+q(x)\cdot(\|x\|^2-1)}$$

different polynomials but same function as $p(x)$ on unit sphere

run time $n^{O(d)}$ *(semidefinite programming)*

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



3-tensor

efficient upper bounds on random polynomials

can show: deg-4 sum-of-squares gives upper bound $\tau_0 = \widetilde{\Theta}(n)^{3/4}$ for random deg-3 polynomial $z(x) = \langle Z, x^{\otimes 3} \rangle$ over unit sphere

concretely: $z(x) + \tau_0/2 \cdot (\|x\|^4 - \|x\|^2)$ has matrix representation with $\lambda_{\max}(\cdot) \leq \tau_0$

approach for recovery: for $\tau \gg \tau_0$, corresponding matrix representation of A has top eigenvector determined by signal v (eig.vec. is close to $v^{\otimes 2}$)

where does upper bound for $z(x)$ come from?

reshape Z to n^2-by-n matrix so that $z(x) = \langle Zx, x^{\otimes 2} \rangle$

tempting but poor Cauchy-Schwarz bound: $\langle Zx, x^{\otimes 2} \rangle \leq \sqrt{\|Zx\|^2 \cdot \|x\|^4}$

tight Cauchy-Schwarz bound: $\langle x, Z^\top x^{\otimes 2} \rangle \leq \sqrt{\|x\|^2 \cdot \|Z^\top x^{\otimes 2}\|^2}$

$Z_i \in \mathbb{R}^{n \times n}$
i-th slice of Z

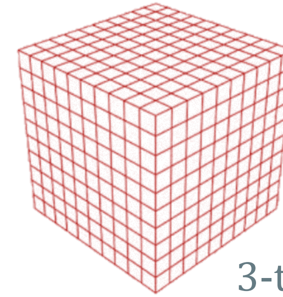
poor matrix representation for $\|Z^\top x^{\otimes 2}\|^2$: $ZZ^\top$ (only rank-n)

best matrix representation for — $= \sum_i \langle x, Z_i x \rangle^2$: $\sum_i Z_i \otimes Z_i$

use matrix
Bernstein

$A = \tau \cdot v^{\otimes 3} + Z \in \mathbb{R}^{n^3}$ with $Z \sim N(0,1)^{\otimes 3}$

MLE: maximize $\langle A, x^{\otimes 3} \rangle$ over all unit vectors $x \in \mathbb{R}^n$ (*)



3-tensor

conclusion

sum-of-squares gives new perspective on spectral algorithms

status quo: focus on spectrum of single matrix associated with problem;
e.g., data matrix (matrix problems), Laplacian (graph problems)

sum-of-squares: associate hierarchy of increasingly rich families of matrices with single problem → ***better algorithms***

research directions

thank you!

faster algorithms via sum-of-squares

challenge: size of matrices increases quickly in the hierarchy (albeit poly.)

upcoming work: techniques to ***significantly compress*** matrices in higher levels of hierarchy (partial traces) [Hopkins-Schramm-Shi-S.'15]

deg-$O(\log n)$ sum-of-squares enough for $\tau \geq \tilde{O}(n)^{1/2}$? (info.-theory limit)