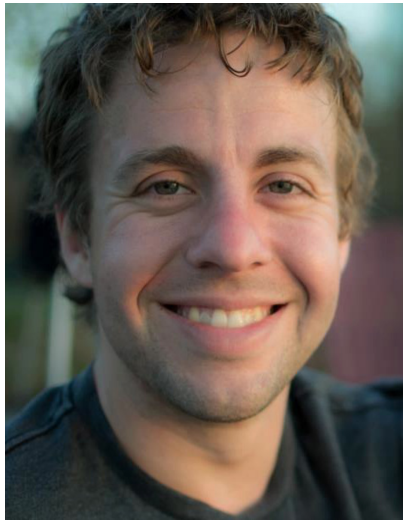
Principal Component Projection and Regression in Nearly Linear Time through Asymmetric SVRG

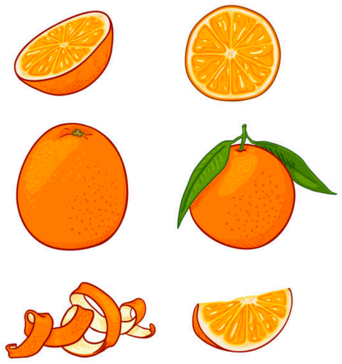
Yujia Jin, Aaron Sidford



Stanford University

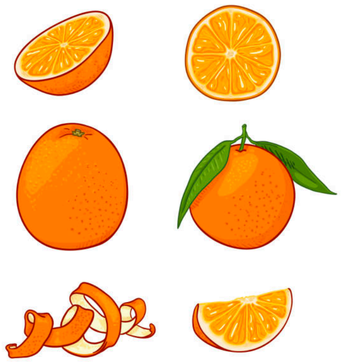
How to make juice?

Peel & cut



How to make juice?

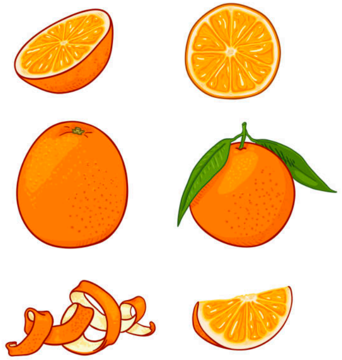
Peel & cut



Put in juicer

How to make juice?

Peel & cut

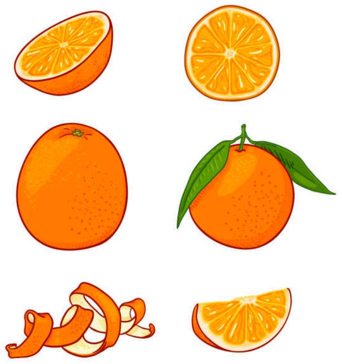


Put in juicer

Juice pulper

How to make juice?

Peel & cut

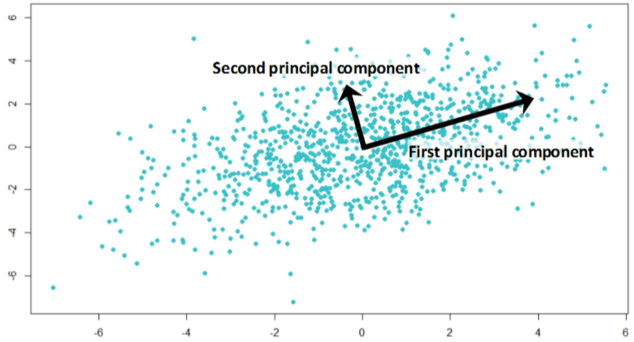


Put in juicer

Juice pulper

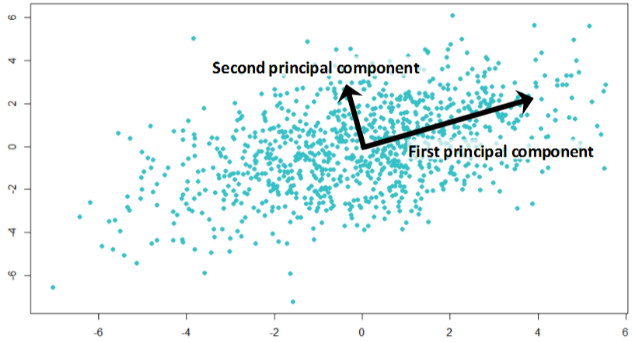
Efficient!

Motivation

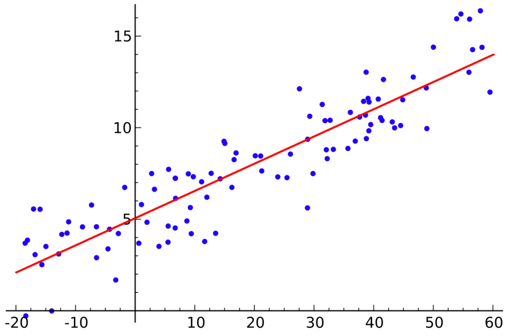


Data preprocessing: PCA

Motivation

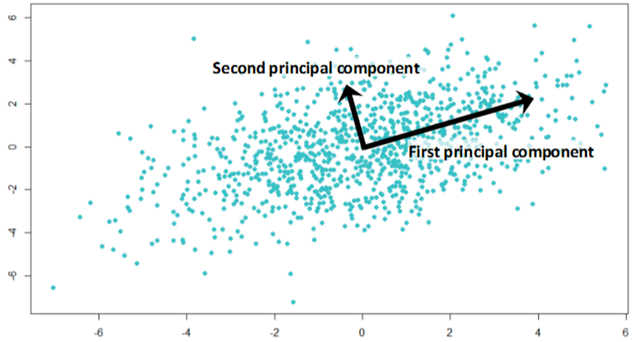


Data preprocessing: PCA

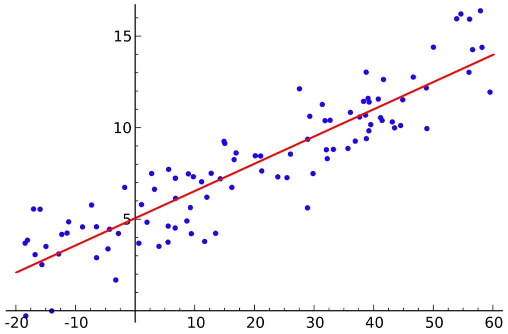


Data analysis tasks: Projection, regression, etc.

Motivation

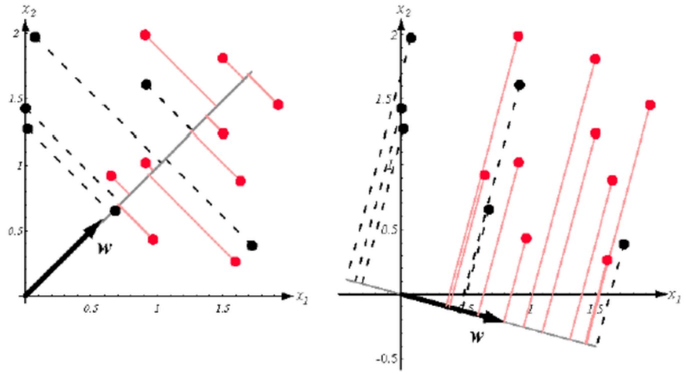


Data preprocessing: PCA

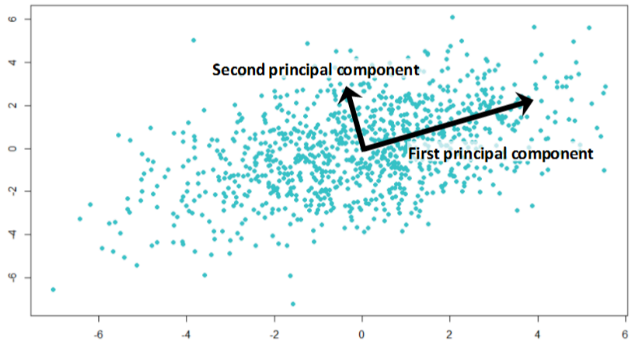


Data analysis tasks: Projection, regression, etc.

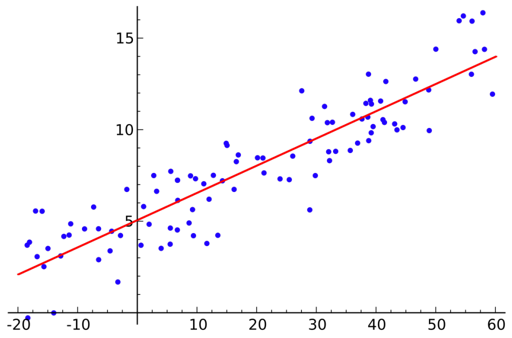
Principal Component Projection & Regression



Motivation

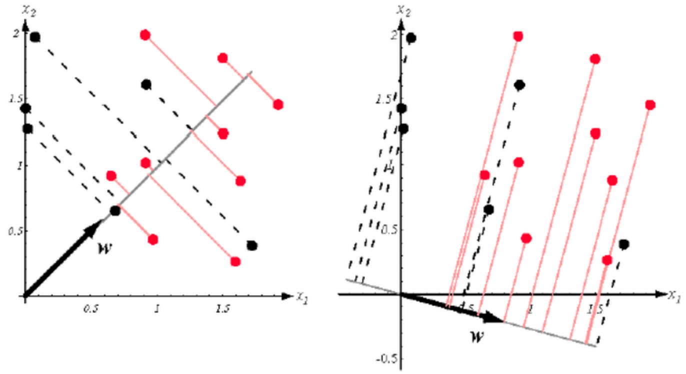


Data preprocessing:
PCA



Data analysis tasks:
Projection, regression, etc.

Principal Component
Projection & Regression



Combining tasks
in one-step

Problems: PCP and PCR

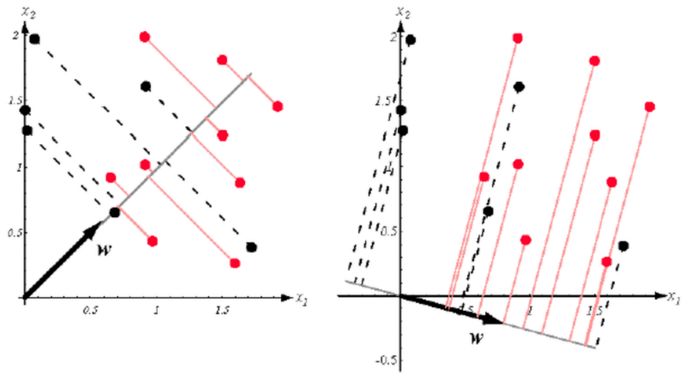
A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

Problems: PCP and PCR

A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

PCR: do regression restricted on $\mathcal{V}^\star$,
= solve $\min_{x \in \mathbb{R}^d} \|AP_\lambda x - b\|$ given b

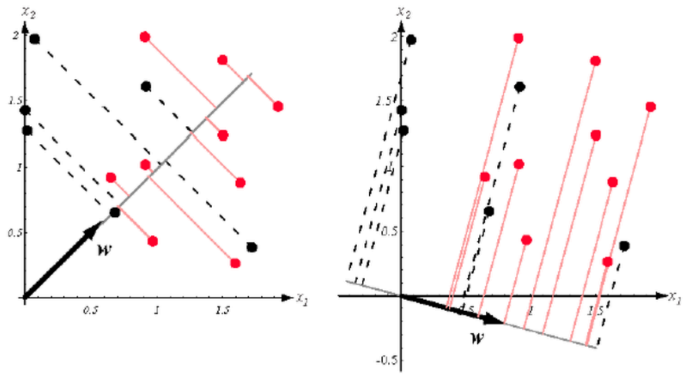


Problems: PCP and PCR

A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

PCR: do regression restricted on $\mathcal{V}^\star$,
= solve $\min_{x\in\mathbb{R}^d} \|AP_\lambda x - b\|$ given b



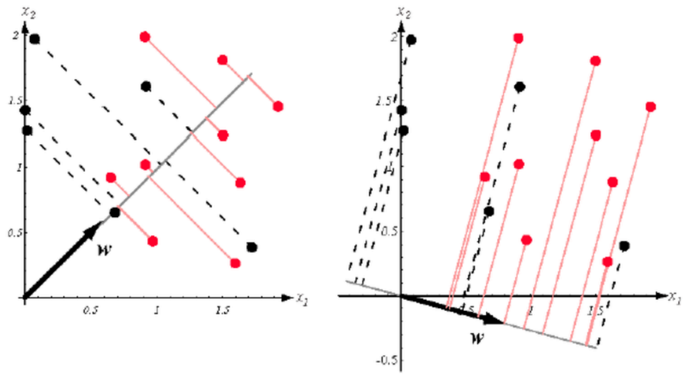
$$\text{PCR} \xrightarrow[\text{[FMMS '16]}]{\tilde{O}(1)\text{ poly. red.}} \text{PCP}$$

Problems: PCP and PCR

A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

PCR: do regression restricted on $\mathcal{V}^\star$,
= solve $\min_{x\in\mathbb{R}^d} \|AP_\lambda x - b\|$ given b



$$\text{PCR} \xrightarrow{\tilde{O}(1)\text{ poly. red.}} \text{PCP}$$

[FMMS '16]

Results: State of the Art

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

linear in k

Classic method:
through computing top-k eigenvectors explicitly

$\tilde{O}(k \cdot \mathrm{nnz} + \ldots)$

e.g. power method

λ - eigenvalue threshold;
γ - eigengap;
k - number of top eigenvalues;
nnz - number of nonzeros;

Results: State of the Art

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

linear in k

Classic method:
through computing top-k eigenvectors explicitly

$\tilde{O}(k \cdot \mathrm{nnz} + \ldots)$

e.g. power method

$\rightarrow$

linear in γ^{-1}

Recent method:
through polynomial approximation of sign(x)

$\tilde{O}(\gamma^{-1}\mathrm{nnz} + \ldots)$

[AL '17, MMS '18]

λ - eigenvalue threshold;
γ - eigengap;
k - number of top eigenvalues;
nnz - number of nonzeros;

Results: State of the Art

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

linear in k

Classic method:
through computing top-k eigenvectors explicitly

$\tilde{O}(k \cdot \mathrm{nnz} + \ldots)$

e.g. power method

→

linear in γ^{-1}

Recent method:
through polynomial approximation of sign(x)

$\tilde{O}(\gamma^{-1}\mathrm{nnz} + \ldots)$

[AL ’17, MMS ’18]

→

nearly-linear!

Our method:
through rational approximation of sign(x)

$\tilde{O}(\mathrm{nnz} + \ldots)$

λ - eigenvalue threshold;
γ - eigengap;
k - number of top eigenvalues;
nnz - number of nonzeros;

Our results: PCP (PCR)

A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

PCP(PCR) solver runtime:

(unaccelerated) $$\tilde{O}(\boxed{\mathrm{nnz}} + d \cdot \mathrm{sr}(A) \cdot \kappa^2 \gamma^{-2})$$

(accelerated) $$\tilde{O}(\boxed{\mathrm{nnz}} + \sqrt{\mathrm{nnz} \cdot d \cdot \mathrm{sr}(A)} \cdot \kappa \gamma^{-1})$$

λ - eigenvalue threshold;
γ - eigengap;
k - number of top eigenvalues;
nnz - number of nonzeros;
$\kappa \triangleq \lambda_1/\lambda;\ \mathrm{sr}(A) \triangleq \|A\|_F^2/\lambda_1.$

Our results: PCP (PCR)

A - n-by-d data matrix;
$\mathcal{V}^\star$ - interested top-eigenspace;
P_λ - projection onto $\mathcal{V}^\star$.

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

PCP(PCR) solver runtime:

(unaccelerated) $\tilde{O}(\boxed{\mathrm{nnz}} + d \cdot \mathrm{sr}(A) \cdot \kappa^2 \gamma^{-2})$

(accelerated) $\tilde{O}(\boxed{\mathrm{nnz}} + \sqrt{\mathrm{nnz} \cdot d \cdot \mathrm{sr}(A)} \cdot \kappa \gamma^{-1})$

λ - eigenvalue threshold;
γ - eigengap;
k - number of top eigenvalues;
nnz - number of nonzeros;
$\kappa \triangleq \lambda_1/\lambda;\ \mathrm{sr}(A) \triangleq \|A\|_F^2/\lambda_1.$

Low-order for large n

Technique #1: Rational Approximation

"Stronger linear algebraic primitives solve problem in fewer steps"

Technique #1: Rational Approximation

"Stronger linear algebraic primitives solve problem in fewer steps"

- Reduce problem to find function $r(x) \approx_\epsilon \mathrm{sign}(x)$ and apply $\frac{1}{2}\left[r\left(A^\top A - \lambda I\right)v + v\right]$

Technique #1: Rational Approximation

"Stronger linear algebraic primitives solve problem in fewer steps"

- Reduce problem to find function $r(x) \approx_{\epsilon} \mathrm{sign}(x)$ and apply $\frac{1}{2}\left[r\left(A^{\top}A - \lambda I\right)v + v\right]$

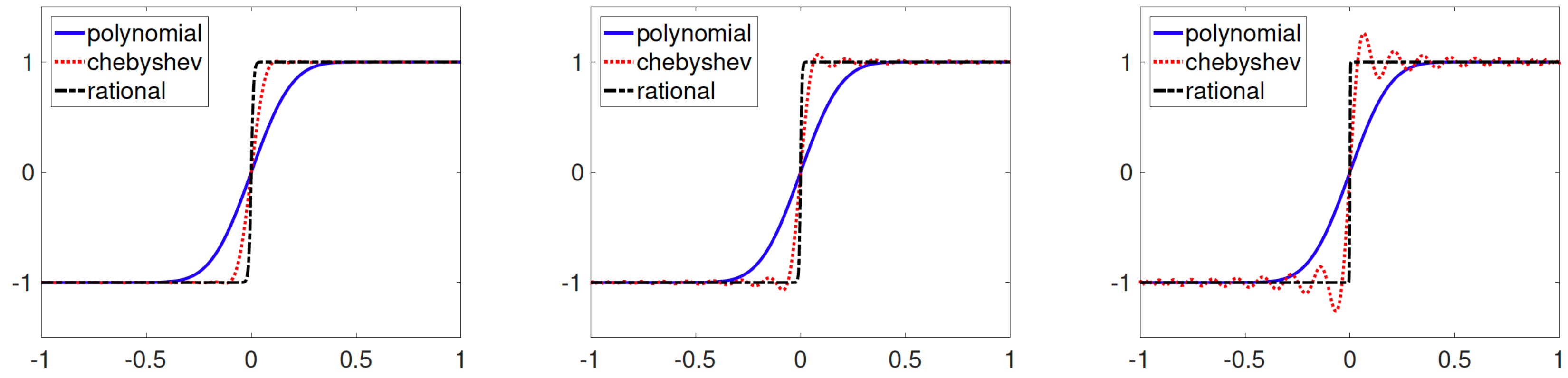
desired Zolotarev rational: $r_k^{\lambda\gamma}(x) = Cx \prod_{i \in [k]} \frac{x^2 + c_{2i}}{x^2 + c_{2i-1}} \approx \mathrm{sign}(x)$

[Zolotarev 1877]

Technique #1: Rational Approximation

"Stronger linear algebraic primitives solve problem in fewer steps"

- Same degree in comparison with standard polynomials [FMMS16], [AL17]



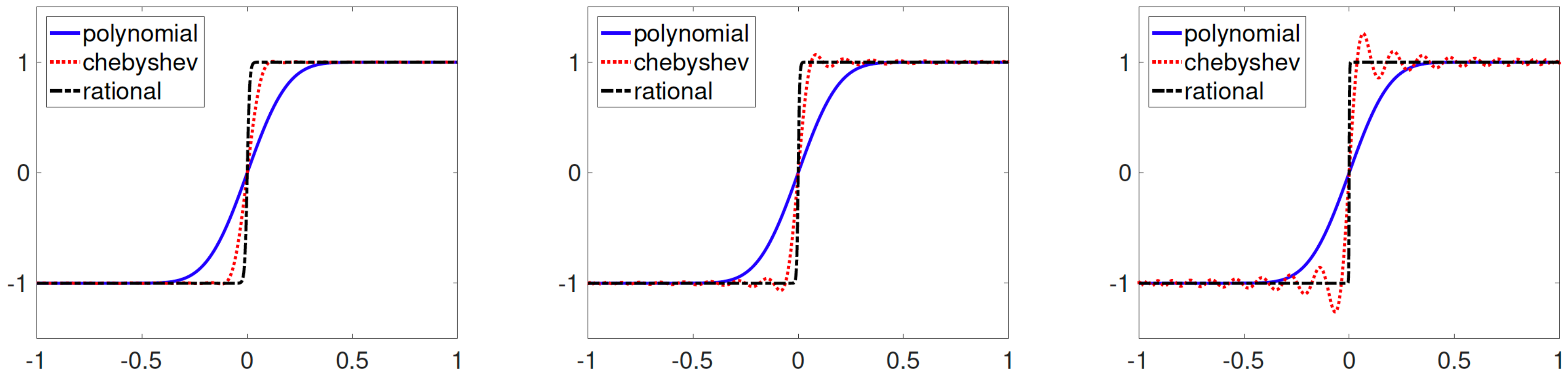
(a) γ=0.1 (b) γ=0.05 (c) γ=0.01

Figure 1: same degree = 21, different γ

Technique #1: Rational Approximation

"Stronger linear algebraic primitives solve problem in fewer steps"

- Compared with standard polynomials [FMMS16], [AL17]



(a) γ=0.1 (b) γ=0.05 (c) γ=0.01

Figure 1: same degree = 21, different γ

Better Approximation Quality under same degree

Technique #2: Asymmetric SVRG solver

"Variance reduction for non-convex problems"

Technique #2: Asymmetric SVRG solver

"Variance reduction for non-convex problems"

Solving squared system $\boxed{(M^2 + \mu^2 I)}x = v$

PSD

Technique #2: Asymmetric SVRG solver

"Variance reduction for non-convex problems"

Solving squared system $\boxed{(M^2 + \mu^2 I)}x = v$

PSD

$\Downarrow$

Solving asymmetric system $\boxed{\begin{pmatrix} I & -\frac{1}{\mu}M \\ \frac{1}{\mu}M & I \end{pmatrix}} z = \begin{pmatrix} 0 \\ v/\mu^2 \end{pmatrix}$

non-PSD

Technique #2: Asymmetric SVRG solver

"Variance reduction for non-convex problems"

Solving squared system $\boxed{(M^2 + \mu^2 I)x = v}$ PSD

$\Downarrow$

Solving asymmetric system $\boxed{\begin{pmatrix} I & -\frac{1}{\mu}M \\ \frac{1}{\mu}M & I \end{pmatrix}} z = \begin{pmatrix} 0 \\ v/\mu^2 \end{pmatrix}$ non-PSD

non-convex
BUT
smaller variance!

Technique #2: Asymmetric SVRG solver

"Variance reduction for non-convex problems"

Solving squared system $\boxed{(M^2 + \mu^2 I)}x = v$ PSD

$\Downarrow$

Solving asymmetric system $\boxed{\begin{pmatrix} I & -\frac{1}{\mu}M \\ \frac{1}{\mu}M & I \end{pmatrix}} z = \begin{pmatrix} 0 \\ v/\mu^2 \end{pmatrix}$. non-PSD

non-convex
BUT
smaller variance!

Our method for solving asymmetric

SVRG (variance-reduced stochastic gradient descent) $\widetilde{O}(\mathrm{nnz} + d \cdot \mathrm{sr}(A) \cdot 1/\mu^2)$

Savings:
Dimensional-related factor in runtime compared with solving squared systems directly

Results: main ideas

PCP: project a given vector onto $\mathcal{V}^\star$,
= compute $P_\lambda v$

1. Rational function approx.
reduce to solve $\tilde{O}(1)$
linear systems in form
$((A^\top A - \lambda I)^2 + cI)x = v$

2. Asymmetric SVRG applied
to nonconvex problem
Runtime: $\tilde{O}(\mathrm{nnz} + d \cdot \mathrm{sr}(A)\lambda^{-2}\gamma^{-2})$

Results: main ideas

PCP: project a given vector onto $\mathcal{V}^{\star}$,
= compute $P_{\lambda} v$

1. Rational function approx.
reduce to solve $\tilde{O}(1)$
linear systems in form
$((A^{\top}A - \lambda I)^2 + cI)x = v$

2. Asymmetric SVRG applied
to nonconvex problem
Runtime: $\tilde{O}(\mathrm{nnz} + d \cdot \mathrm{sr}(A)\lambda^{-2}\gamma^{-2})$

Overall complexity
(unaccelerated):

$$\tilde{O}(\mathrm{nnz} + d \cdot \mathrm{sr}(A) \cdot \frac{\kappa^2}{\gamma^2})$$

Takeaways:

- **Idea:**
  I. rational function allows lower degree approximation

- **Solver:**

Takeaways:

- **Idea:**
  - **I.** rational function allows lower degree approximation
  - **II.** better variance reduction by extending the problem to larger dimension space
- **Solver:**

Takeaways:

- **Idea:**
  - **I.** rational function allows lower degree approximation
  - **II.** better variance reduction by extending the problem to larger dimension space
- **Solver:**
  - **I.** for structured squared / asymmetric / non-PSD systems

Takeaways:

- **Idea:**
  - **I.** rational function allows lower degree approximation
  - **II.** better variance reduction by extending the problem to larger dimension space
- **Solver:**
  - **I.** for structured squared / asymmetric / non-PSD systems
  - **II.** for sign and square-root matrix function approximation

Takeaways:

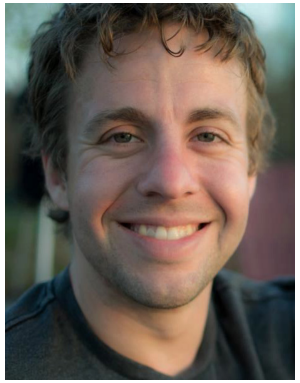
- **Idea:**
  - **I.** rational function allows lower degree approximation
  - **II.** better variance reduction by extending the problem to larger dimension space

- **Solver:**
  - **I.** for structured squared / asymmetric / non-PSD systems
  - **II.** for sign and square-root matrix function approximation
  - **III.** for nearly-linear time PCP and PCR solver

Thank you!

Yujia Jin

Aaron Sidford



Questions?

- Welcome to our poster @ 10:45am – 11:45am, Wednesday (12/11)
  @ East Exhibition Hall B + C #162
- arxiv: 1910.06517
- Email: yujiajin@stanford.edu