Approximability of subspace approximation

Amit Deshpande

Microsoft Research India

Credits

Based on joint work with Madhur Tulsiani and Nisheeth Vishnoi,

Credits

Based on joint work with Madhur Tulsiani and Nisheeth Vishnoi,

and disjoint (?) work of

- Varadarajan, Venkatesh, Ye, and Zhang (SICOMP, 2007)
- Kindler, Naor, and Schechtman (Math of OR, 2010)
- Guruswami, Raghavendra, Saket, and Wu (preprint)
- ...
- the anonymous heroes who discovered eigenvalues, eigenvectors, gaussians etc.

Subspace approximation

Find a low-dimensional representation of high-dimensional data up to a small error.

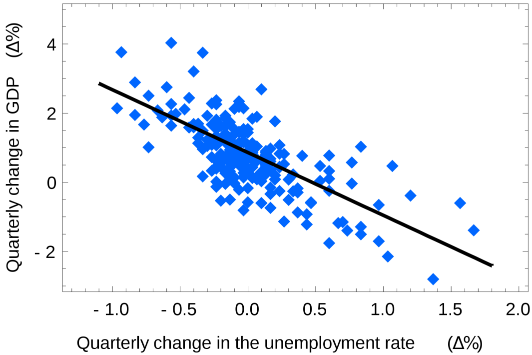
or for simplicity,

Given n points $a_1, a_2, \ldots, a_n \in \mathbb{R}^d$, find a k-dimensional linear subspace V that minimizes

$$\left(\sum_{i=1}^{n} d(a_i, V)^p \right)^{1/p}.$$

Special cases

$p = 2$ (ordinary least squares) and $p = \infty$ (radii of point sets).



http://en.wikipedia.org/wiki/File:Okuns_law_quarterly_differences.svg

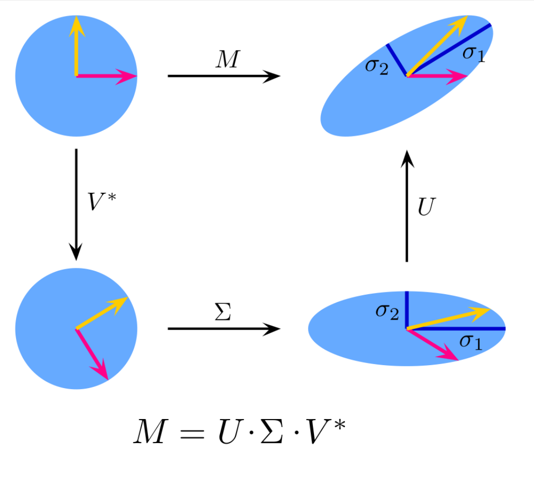
Simplest case: $\dim(V) = d - 1$ and $p = 2$

Subspace V is uniquely identified by its unit normal x.

$$\min_{\dim(V)=d-1} \left(\sum_{i=1}^{n} d(a_i, V)^2 \right)^{1/2} = \min_{\|x\|_2=1} \left(\sum_{i=1}^{n} \langle a_i, x \rangle^2 \right)^{1/2}$$
$$= \min_{\|x\|_2=1} \|Ax\|_2 \,.$$

So the optimal x is the smallest singular vector of $A \in \mathbb{R}^{n \times d}$, which has $a_1, \ldots, a_n$ as its rows.

Singular Value Decomposition (SVD)



http://commons.wikimedia.org/wiki/File:Singular-Value-Decomposition.svg

Beyond SVD?

- Convex programs: minimize a convex function over a convex domain.

Beyond SVD?

- Convex programs: minimize a convex function over a convex domain.

- Minimizing $\|Ax\|_p$ over $\|x\|_2 \geq 1$ is not a convex program.

Beyond SVD?

- Convex programs: minimize a convex function over a convex domain.
- Minimizing $\|Ax\|_p$ over $\|x\|_2 \geq 1$ is not a convex program.
- Magic of SVD: But we can do this *efficiently* for $p = 2$ (Ref. Matrix Computations, Golub and Van Loan).

Convex relaxation and randomized rounding

Using $\langle a_i, x\rangle^p = (a_i^T x x^T a_i)^{p/2}$,

$$\min_{\|x\|_2=1} \left(\sum_{i=1}^{n} \langle a_i, x\rangle^p \right)^{1/p} \xrightarrow{\text{relax}} \min_{\substack{X\,:\,X\succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \left(\sum_{i=1}^{n} \left(a_i^T X a_i \right)^{p/2} \right)^{1/p}.$$

Convex relaxation and randomized rounding

Using $\langle a_i, x\rangle^p = (a_i^T x x^T a_i)^{p/2}$,

$$\min_{\|x\|_2=1} \left(\sum_{i=1}^n \langle a_i, x\rangle^p \right)^{1/p} \xrightarrow{\text{relax}} \min_{\substack{X \,:\, X \succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \left(\sum_{i=1}^n \left(a_i^T X a_i \right)^{p/2} \right)^{1/p} .$$

- X may not have rank one (and thus no expression as xx^T); SVD gives $X = V \Sigma V^T = \sum_{i=1}^n \sigma_i v_i v_i^T$.

Convex relaxation and randomized rounding

Using $\langle a_i, x\rangle^p = (a_i^T x x^T a_i)^{p/2}$,

$$\min_{\|x\|_2=1} \left(\sum_{i=1}^{n} \langle a_i, x\rangle^p \right)^{1/p} \xrightarrow{\text{relax}} \min_{\substack{X\,:\,X \succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \left(\sum_{i=1}^{n} \left(a_i^T X a_i \right)^{p/2} \right)^{1/p}.$$

- X may not have rank one (and thus no expression as xx^T); SVD gives $X = V\Sigma V^T = \sum_{i=1}^{n} \sigma_i v_i v_i^T$.
- Compute SVD of X to get its singular values $\sigma_1, \ldots, \sigma_n$ and singular vectors $v_1, \ldots, v_n$. Output as vector x the (normalized) random linear combination

$$\sum_{i=1}^{n} r_i \sqrt{\sigma_i} v_i, \quad \text{where } r_i\text{'s are i.i.d. } N(0,1).$$

Approximation factor

We show that

$$\text{Relax} \leq \text{OPT} \leq \text{Round}$$

Approximation factor

We show that

$$\text{Relax} \leq \text{OPT} \leq \text{Round} \leq \gamma_p \cdot \text{Relax}$$

Approximation factor

We show that

$$\text{Relax} \leq \text{OPT} \leq \text{Round} \leq \gamma_p \cdot \text{Relax} \leq \gamma_p \cdot \text{OPT},$$

giving an approximation factor of $\gamma_p = \mathrm{E}\left[r^p\right]^{1/p} \approx \sqrt{p/e}$, the p-th moment of $N(0,1)$ random variable r.

Approximation factor

We show that

$$\text{Relax} \leq \text{OPT} \leq \text{Round} \leq \gamma_p \cdot \text{Relax} \leq \gamma_p \cdot \text{OPT},$$

giving an approximation factor of $\gamma_p = \mathrm{E}\left[r^p\right]^{1/p} \approx \sqrt{p/e}$, the p-th moment of $N(0,1)$ random variable r.

More generally, the convex relaxation and rounding (with tiny modifications) give an approximation factor of $\sqrt{2}\gamma_p$ for k-dimensional subspace approximation, for any k and $p \geq 2$. For $p = \infty$, the approximation factor becomes $O(\sqrt{\log n})$.

Continuous analog and integrality/rank gap

Continuous analog of subspace approximation

$$\min_{\|x\|_2=1} \sum_{i=1}^{n} \langle a_i, x \rangle^p \quad \xrightarrow{\text{relax}} \quad \min_{\substack{X \,:\, X \succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \sum_{i=1}^{n} \left(a_i^T X a_i\right)^{p/2}$$

Continuous analog and integrality/rank gap

Continuous analog of subspace approximation

$$\min_{\|x\|_2=1} \sum_{i=1}^n \langle a_i, x\rangle^p \xrightarrow{\text{relax}} \min_{\substack{X \,:\, X \succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \sum_{i=1}^n \left(a_i^T X a_i\right)^{p/2}$$

$$\Big\downarrow \text{continuous} \qquad\qquad\qquad\qquad \Big\downarrow \text{continuous}$$

$$\min_{\|x\|_2=1} \int_{\mathbb{R}^d} \langle g, x\rangle^p \, \mu(g) dg \xrightarrow{\text{relax}} \min_{\substack{X \,:\, X \succcurlyeq 0 \\ X \text{ symmetric} \\ \text{trace}(X)=1}} \int_{\mathbb{R}^d} \left(g^T X g\right)^{p/2} \mu(g) dg$$

Continuous analog and integrality/rank gap

Continuous analog of subspace approximation

$$\min_{\|x\|_2=1} \sum_{i=1}^n \langle a_i, x\rangle^p \quad \xrightarrow{\text{relax}} \quad \min_{\substack{X\,:\,X\succcurlyeq 0\\ X \text{ symmetric}\\ \text{trace}(X)=1}} \sum_{i=1}^n \left(a_i^T X a_i\right)^{p/2}$$

$\downarrow$ continuous $\qquad\qquad$ $\downarrow$ continuous

$$\min_{\|x\|_2=1} \int_{\mathbb{R}^d} \langle g, x\rangle^p \,\mu(g) dg \quad \xrightarrow{\text{relax}} \quad \min_{\substack{X\,:\,X\succcurlyeq 0\\ X \text{ symmetric}\\ \text{trace}(X)=1}} \int_{\mathbb{R}^d} \left(g^T X g\right)^{p/2} \mu(g) dg$$

$$\text{LHS} = \int_{\mathbb{R}} g_1^p e^{-g_1^2/2} dg_1 \quad \text{vs.} \quad \text{RHS} \leq \frac{1}{d^{p/2}} \int_{\mathbb{R}^d} \|g\|^p \, e^{-\|g\|^2/2} dg$$

Dictatorship test

$$\text{IsDictator(x)} : x \mapsto \mathrm{E}\left[\langle a, x\rangle^{p}\right] = \frac{1}{2^d} \sum_{a \in \{-1,1\}^d} \langle a, x\rangle^{p} .$$

Dictatorship test

$$\mathrm{IsDictator(x)} : x \mapsto \mathrm{E}\left[\langle a, x\rangle^p\right] = \frac{1}{2^d} \sum_{a \in \{-1,1\}^d} \langle a, x\rangle^p .$$

- Dictator: If $x = (1, 0, \ldots, 0)$ then $\mathrm{E}\left[\langle a, x\rangle^p\right] = 1$, for p even.

Dictatorship test

$$\text{IsDictator(x)} : x \mapsto \mathrm{E}\left[\langle a, x\rangle^p\right] = \frac{1}{2^d} \sum_{a \in \{-1,1\}^d} \langle a, x\rangle^p \,.$$

- Dictator: If $x = (1, 0, \ldots, 0)$ then $\mathrm{E}\left[\langle a, x\rangle^p\right] = 1$, for p even.
- Far-from-dictator: If all the coordinates of x are small, then

$$\begin{aligned}
\mathrm{E}\left[\langle a, x\rangle^p\right] &\approx \mathrm{E}\left[\langle g, x\rangle^p\right] && \text{by invariance principle} \\
&= \mathrm{E}\left[\langle g, (1, 0, \ldots, 0)\rangle^p\right] && \text{by spherical symmetry} \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g_1^p e^{-g_1^2/2} dg_1 = \gamma_p^p .
\end{aligned}$$

Thank you. Any questions?