12 graph drawing; SVDs

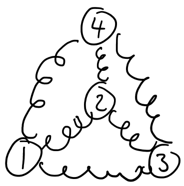
Wednesday, October 14, 2020 5:02 PM

Last time we left off defining the energy of a graph drawing.

Define: The energy of a drawing R be

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} \| \rho(v_i) - \rho(v_j) \|^2 .$$

Analogy:



Connect nodes by springs and then minimize the potential energy of the system.

"Good" drawings are ones that minimize energy (but are not trivial)

Define: The energy of a drawing R of a weighted graph $G = (V, W)$ is

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} w_{ij} \| \rho(v_i) - \rho(v_j) \|^2 .$$

(think of w_{ij} as spring stiffness)

Prop. 19.1 Let $G = (V, W)$ be a weighted graph, with $|V| = m$ and $W \in \mathbb{R}^{m \times m}$ symmetric, and let R be the matrix of a graph drawing ρ of G in $\mathbb{R}^n$ (an $m \times n$ matrix). If $L = D - W$ is the unnormalized Laplacian matrix, then

$$\mathcal{E}(R) = \operatorname{tr}(R^T L R) .$$

Proof.

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} w_{ij} \| \rho(v_i) - \rho(v_j) \|^2$$

$$= \sum_{k=1}^{n} \sum_{\{v_i, v_j\} \in E} w_{ij} (R_{ik} - R_{jk})^2$$

$$= \sum_{k=1}^{n} \cdot \frac{1}{2} \sum_{i,j=1}^{m} w_{ij} (R_{ik} - R_{jk})^2$$

$$= \sum_{k=1}^{n} (R^k)^T L R^k$$

(where R^k is the kth col of R)
(by Prop. 18.4)

$$= \operatorname{tr}(R^T L R) .$$

$\boxed{\diagup\!\diagup}$

$= \mathrm{Tr}(R^T L R).$ $\square$

So the energy $\mathcal{E}(R)$ is the sum of the (nonnegative) eigenvalues of $R^T L R$.

Note that for any invertible matrix M, $\rho(v_i) M$ is another graph drawing that conveys the same amount of information. So we may as well choose R + have pairwise orthogonal unit length cols, $R^T R = I$.

Def. 19.3 If a matrix R of a graph drawing satisfies $R^T R = I$, then the corresponding drawing is an orthogonal graph drawing (this rules out trivial drawings)

Thm 19.1/19.2 Let $G = (V, W)$ be a weighted graph connected graph with $|V| = m$. If the eigenvalues of $L = D - W$ are $0 = \lambda_1 < \lambda_2 \le \lambda_3 \le \cdots \le \lambda_m$, then the minimal energy of any balanced orthogonal graph drawing of G in $\mathbb{R}^n$ is $\lambda_2 + \cdots + \lambda_{n+1}$. The $m \times n$ matrix R consisting of any unit eigenvectors $u_2, \dots, u_{n+1}$ associated with $\lambda_2, \dots, \lambda_{n+1}$ yields a balanced orthogonal graph drawing of minimal energy.

proof. By the Poincare separation theorem/eigenvalue interlacing,

$$\lambda_k = \lambda_k(L) \le \lambda_k(R^T L R)$$

$$\Rightarrow \sum_{k=1}^{n} \lambda_k \le \mathrm{Tr}(R^T L R).$$

And if $R = [u_1 \cdots u_n]$, then $R^T L R = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$, so $\mathrm{Tr}(R^T L R) = \sum_{k=1}^{n} \lambda_k$.

But $\lambda_1 = 0$ and $\lambda_2 > 0$, so $u_1 = \frac{\vec{1}}{\sqrt{m}}$.

Because $\langle u_i, u_1 \rangle = 0 \ \forall i \ne 1$, $u_i^T \vec{1} = 0$.

Thus, we can get a balanced orthogonal drawing in $\mathbb{R}^{n-1}$ by removing u_1 and just using $R = [u_2 \cdots u_n]$, which has the same energy

i.e. Balanced orthogonal drawing in $\mathbb{R}^n$ $\Longleftrightarrow$ orthogonal drawing in $\mathbb{R}^{n+1}$, with energy $\lambda_2 + \cdots + \lambda_{n+1}$. $\square$

Aside, using the first eigenvector $u_1 = \frac{\vec{1}}{\sqrt{m}}$ is undesirable because it means all pts have the same first coordinate, another reason to remove it.

Singular value decomposition and polar forms

Def. 20.3 A triple (V, D, U) s.t. $A = VDU^T$, where the columns of V and U are orthonormal (i.e. $V^TV = I$, $U^TU = I$), and D is a diagonal matrix with nonnegative entries is called a singular value decomposition (SVD) of A.

To understand this, let's take a step back and consider maps composed with their adjoints.

Euclidean spaces

Prop. 20.1 Given any linear map $f: E \to F$, the eigenvalues of $f^* \circ f$ and $f \circ f^*$ are nonnegative.

proof. Suppose $f^* \circ f(u) = \lambda u$.

$$\lambda \langle u, u \rangle = \langle (f^* \circ f)(u), u \rangle = \langle f(u), f(u) \rangle$$

$$\Rightarrow \lambda \geq 0.$$

Prop. Given linear maps $f: E \to F$, $g: F \to E$, $g \circ f: F \to F$ and $f \circ g: E \to E$ have the same nonzero eigenvalues. $(\dim(E) = n, \dim(F) = m)$

proof. Suppose $\lambda \neq 0$ is an eigenvalue of $g \circ f$ with eigenvector u.

Then $g \circ f(u) = \lambda u$

$\Rightarrow f \circ g \circ f(u) = \lambda f(u)$

$\Rightarrow \lambda$ is an eigenvalue of $f \circ g$ with eigenvector $f(u)$, unless $f(u) = 0$.

Suppose $f(u) = 0$. Then $g \circ f(u) = 0$, a contradiction because $\lambda \neq 0$.

Thus, $f(u) \neq 0$, so λ is an eigenvalue of $f \circ g$.

Def 20.1 Given $f: E \to F$ linear, the square roots $\sigma_i > 0$ of the positive eigenvalues of $f \circ f^*$ (or $f^* \circ f$) are called the singular values of f.

positive eigenvalues ... singular values of f.

Prop. 20.2 Given $f: E \to F$,

$$\mathrm{Ker}\, f = \mathrm{Ker}\,(f^* \circ f)$$
$$\mathrm{Ker}\, f^* = \mathrm{Ker}\,(f \circ f^*)$$
$$\mathrm{Ker}\, f = (\mathrm{Im}\, f^*)^\perp$$
$$\mathrm{Ker}\, f^* = (\mathrm{Im}\, f)^\perp$$
$$\dim(\mathrm{Im}\, f) = \dim(\mathrm{Im}\, f^*)$$
$$\mathrm{rank}(f) = \mathrm{rank}(f^*) = \mathrm{rank}(f \circ f^*) = \mathrm{rank}(f^* \circ f).$$

proof omitted

Thm 20.1/20.3 For every real $n \times n$ matrix A, $\exists\, U$ and V orthogonal and a diagonal matrix D s.t. $A = VDU^T$ and

$$D = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix},$$

where $\sigma_1, \dots, \sigma_r$ are the singular values of A and $\sigma_{r+1} = \dots = \sigma_n = 0$.
The cols of U are the eigenvectors of A^TA (a.k.a. right singular vectors)
and cols of V are the eigenvectors of AA^T (a.k.a. left singular vectors).

proof. A^TA is symmetric pos semidefinite, so by spectral theorem

$$A^TA = UD^2U^T, \quad D = \mathrm{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0),$$

where $\sigma_1^2, \dots, \sigma_r^2$ are the nonzero eigenvalues of A^TA,

and $r = \mathrm{rank}(A)$.

Then $U^TA^TAU = D^2$.

Let f_i be the ith col of AU.

Then $\langle f_i, f_i \rangle = \sigma_i^2$

and $\langle f_i, f_j \rangle = 0 \quad \forall\, i \neq j$.

For $1 \le i \le r$, let $v_i = \sigma_i^{-1} f_i$, an orthonormal set of vectors.

For $1 \le i \le r$, let $v_i = \sigma_i^{-1} f_i$, an orthonormal set of vectors.
Complete into an orthonormal basis $(v_i)_{1 \le i \le n}$, and note

$$\langle v_i, f_j \rangle = \sigma_j \langle v_i, v_j \rangle = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j. \end{cases}$$

Then $V^T A U = D$ by design.

$\Rightarrow A = V D U^T$.

Then $A^T A = U D^2 U^T$, $A A^T = V D^2 V^T$.

So $A^T A U = U D^2$, $A A^T V = V D^2$, proving that cols of U and V are eigenvectors of $A^T A$ and $A A^T$.

∎

Aside: proof also holds for complex $n \times n$ matrices $A = V D U^*$.

Def. 20.4 A pair (R, S) such that $A = RS$ with R orthogonal and S symmetric positive semidefinite is called a polar decomposition of A.

$\forall$ $n \times n$ real matrix A, $A = V D U^T$, let $\underbrace{R = V U^T}_{\text{orthogonal}}$ and $S = \underbrace{U D U^T}_{\text{SPSD}}$, so $A = RS$.

Aside: $A = UH$, where A is complex,
U is unitary,
H is Hermitian pos. semi-def.

Thm 20.2/20.4 (Weyl's inequalities, 1949)
For any complex $n \times n$ matrix A, if $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ are the eigenvalues of A and $\sigma_1, \dots, \sigma_n \in \mathbb{R}_+$ are the singular values of A, listed so that $|\lambda_1| \ge \dots \ge |\lambda_n|$ and $\sigma_1 \ge \dots \ge \sigma_n \ge 0$, then

$$|\lambda_1| \cdots |\lambda_n| = \sigma_1 \cdots \sigma_n \quad \text{and}$$

$$|\lambda_1| \cdots |\lambda_k| \le \sigma_1 \cdots \sigma_k \quad \text{for } k = 1, \dots, n-1.$$

proof omitted

But in general, the relationship b/t singular values and eigenvalues is not obvious.

But what do singular values mean? We see a hint in the RS decomposition, that somehow the SVD is decomposing actions into a "rotation" and a symmetric positive semidefinite transform, so $S = UDU^T$ is a scaling in the appropriate basis, with scaling magnitude determined by the singular values. That suggests another motivation for SVDs.

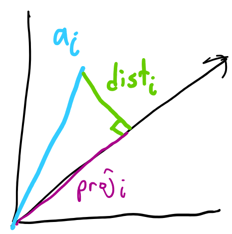
Suppose we have n points in $\mathbb{R}^d$.

Let's try to find the best-fitting 1-dim subspace (i.e fit a line)

One idea is to minimize squared distance from the line.

How should we measure squared distance?

- Linear regression: measure distance only along a privileged axis
- SVD: measure ordinary Euclidean dist. from line.



Consider projecting a single pt $\vec{a}_i$ to a line.

$$\underbrace{\|a_i\|^2}_{\text{independent of line.}} = \underbrace{dist_i^2}_{\text{minimizing distance}} + \underbrace{proj_i^2}_{\text{maximizing projection}}.$$

minimizing distance $\Longleftrightarrow$ maximizing projection

Let $a_1, \dots, a_n \in \mathbb{R}^d$, and let $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \in \mathbb{R}^{n \times d}$.

Let $u \in \mathbb{R}^d$ be a unit vector, defining a line (1D subsp) $U \subseteq \mathbb{R}^d$.

Let $P_U : \mathbb{R}^n \to U$ be the projection operator, $P_U(x) = (x \cdot u)\,u$.

Then $\|P_U(a_i)\| = |a_i \cdot u|$, and $\sum_{i=1}^n |a_i \cdot u|^2 = \|Au\|^2$.

Def. The first singular vector u_1 of A is

$$u_1 = \arg\max_{|u|=1} \|Au\|.$$

u_1 defines the best-fit line in terms of minimizing squared distance.

Def. The first right singular value $\sigma_1(A) = \|Au_1\|$

Note that $\sigma_1^{\,2} = \sum_{i=1}^{n} (a_i \cdot u_1)^2$.

u_1 is the line capturing the most variance of $a_1, \dots, a_n$.

This gives the best-fit 1D subspace. What about higher-dim subspaces?

Let's be greedy.

Def. The kth right singular vector u_k of A is

$$u_k = \arg\max_{\substack{u \in \{u_1, \dots, u_{k-1}\}^\perp \\ |u|=1}} |Au|$$

and $\sigma_k(A) = |Au_k|$ is the kth right singular value.

Thm Let $A \in \mathbb{R}^{n \times d}$ with right singular vectors $u_1, \dots, u_r$. For $1 \le k \le r$, let $U_k = \text{span}\{u_1, \dots, u_k\}$. For each k, U_k is the best-fit k-dim subsp. of A.

(by best-fit, we mean minimizing squared dist)

proof. By induction. By def. for $k=1$.

Assume U_{k-1} is a best-fit $(k-1)$-dim subsp.

Suppose W is a best-fit k-dim. subsp.

Choose an orthonormal basis $w_1, \dots, w_k$ of W so that $w_k \perp U_{k-1}$.

Then $\sum_{i=1}^{k-1} \|Aw_i\|^2 \le \sum_{i=1}^{k-1} \|Au_i\|^2$ because U_{k-1} is optimal.

Furthermore, $\|Aw_k\|^2 \le \|Au_k\|^2$ because $u_k = \arg\max_{\substack{u \perp U_{k-1} \\ |u|=1}} \|Au\|^2$.

$\Rightarrow \sum_{i=1}^{k} \|Aw_i\|^2 \le \sum_{i=1}^{k} \|Au_i\|^2 \Rightarrow U_k$ is also optimal. $\square$

$$\Rightarrow \sum_{i=1}^{k} \|Aw_i\|^2 \le \sum_{i=1}^{k} \|Au_i\|^2 \Rightarrow U_k \text{ is also optimal.}$$

$\boxed{/\!/}$

Def. The left singular vectors are defined as $v_1, \dots, v_r$, where

$$v_i = \frac{1}{\sigma_i(A)} A u_i .$$

Aside: $v_i = \arg\max_{\substack{v \in \{v_1, \dots, v_{i-1}\}^\perp \\ |v| = 1}} |v^T A|$, are are also orthogonal. (proof later)

Thm. Let $A \in \mathbb{R}^{n \times d}$ with right-singular vectors $u_1, \dots, u_r$
left-singular vectors $v_1, \dots, v_r$
and singular values $\sigma_1, \dots, \sigma_r$

Let $U = [u_1 \cdots u_r] \in \mathbb{R}^{d \times r}$

$V = [v_1 \cdots v_r] \in \mathbb{R}^{n \times r}$

$D = \operatorname{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$

Then $A = VDU^T = \sum_{i=1}^{r} \sigma_i v_i u_i^T \in \mathbb{R}^{n \times d}$

proof. Extend $u_1, \dots, u_r$ to an orthonormal basis $u_1, \dots, u_d$ of $\mathbb{R}^d$
and let $\sigma_{r+1} = \dots = \sigma_d = 0$.

Then $$\sum_{i=1}^{r} \sigma_i v_i u_i^T u_j = \sum_{i=1}^{n} \sigma_i v_i \underbrace{\langle u_i, u_j \rangle}_{\substack{0 \text{ if } i \neq j \\ 1 \text{ if } i = j}} = \sigma_j v_j = A u_j .$$

Thus VDU^T applied to a basis $u_1, \dots, u_d$ gives the same thing as A,
so $A = VDU^T$.

$\boxed{/\!/}$

Thm Let A be a rank r matrix. The left singular vectors $v_1, \dots, v_r$ are orthogonal.

proof. $A^TA = UDV^TVDU^T$

$$\Rightarrow \quad U^TA^TAU = DV^TVD$$

$$\Rightarrow \quad D^{-1}U^TA^TAUD^{-1} = V^TV$$

By Rayleigh-Ritz $\max_{|x|=0} x^TA^TAx = \lambda_{max}(A^TA)$

$$\Rightarrow \quad \lambda_{max}(A^TA) = \max_{|x|=0} \|Ax\|^2 = \sigma_1^2, \text{ achieved for } x = u_1.$$

Easy to show from this that $u_1, \dots, u_r$ are eigenvectors of A^TA corresponding to the largest r eigenvalues $\sigma_1^2, \dots, \sigma_r^2$ of A^TA.

Thus, $U^TA^TAU = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_r^2 \end{bmatrix} \Rightarrow D^{-1}U^TA^TAUD^{-1} = I$

$$\Rightarrow V^TV = I, \text{ so}$$

V has orthogonal columns $(v_1, \dots, v_r)$.

Low-rank approximations

Lemma The rows of $A_k = \sum_{i=1}^{k} \sigma_i v_i u_i^T$ are the projections of the rows of A onto $U_k = \text{span}\{u_1, \dots, u_k\}$.

proof. Let $A = \begin{bmatrix} a^1 \\ \vdots \\ a^n \end{bmatrix}$

$P_{U_k}(a^i) = \sum_{i=1}^{k} (a^i \cdot u_i) u_i$ then $A_k = \sum_{i=1}^{k} \sigma_i v_i u_i^T = \sum_{i=1}^{k} A u_i u_i^T$

$$= \sum_{i=1}^{k} \begin{bmatrix} a^1 \\ \vdots \\ a^n \end{bmatrix} u_i u_i^T = \begin{bmatrix} P_{U_k}(a^1) \\ \vdots \\ P_{U_k}(a^n) \end{bmatrix}.$$

Thm For any matrix B of rank at most k,

$$\|A - A_k\|_F \le \|A - B\|_F \quad .$$

proof. Let $B = \underset{B' \mid \text{rank}(B') \le k}{\arg\min} \|A - B'\|_F^2$, $\quad B = \begin{bmatrix} b^1 \\ \vdots \\ b^n \end{bmatrix}$.

Let $V = \text{span}\{b^1, \ldots, b^k\}$. $\quad \dim(V) \le k$.

Suppose $b^i \ne P_V(a^i)$.

Then replace b^i with $P_V(a^i)$ in B, forming B'.

$P_V(a^i) \in V$, so $\text{span}\{\text{rows of } B'\} \subseteq V$, so $\text{rank}(B') \le k$.

But $\|A - B'\|_F \le \|A - B\|_F^2$ because $P_V(a^i) = \underset{v \in V}{\arg\min} \|a^i - v\|$.

($P_V(a^i)$ is the closest V gets to row a^i).

WLOG, all rows of B are projections of rows of A onto V.

But A_k minimizes squared distances to any k-dim subspace (from proof of SVD),

so $\quad \|A - A_k\|_F \le \|A - B\|_F$.

Lemma: $\|A\|_2 = \sigma_1(A)$

proof. $\sigma_1(A) = \max_{|u|=1} \|Au\| = \|A\|_2$.

Lemma: $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

proof. Let $A = \sum_{i=1}^{r} \sigma_i v_i u_i^T$ be the SVD of A.

Then $A - A_k = \sum_{i=k+1}^{r} \sigma_i v_i u_i^T$, and by uniqueness of the SVD, this is also a SVD.

$$\Rightarrow \|A - A_k\|_2 = \sigma_1(A - A_k) = \sigma_{k+1}$$

Thm Let $A \in \mathbb{R}^{n \times d}$. For any matrix B of rank at most k,

$$\|A - A_k\|_2 \le \|A - B\|_2 \quad .$$

proof. If $\text{rank}(A) \le k$, then $A - A_k = 0$, $\|A - A_k\|_2 = 0$, so trivially true.

Assume $\text{rank}(A) > k$. Then $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.

$\dim(\mathrm{Ker}\, B) \geq d-k$.

Let $u_1, \ldots, u_{k+1}$ be the first $k+1$ sing. vec. of A.

Then $\exists z \neq 0$ s.t. $z \in \mathrm{Ker}\, B \cap \mathrm{span}(u_1, \ldots, u_{k+1})$, and $\|z\|_2 = 1$.

Then $\|A-B\|_2^2 \geq \|(A-B)z\|_2^2 = \|Az\|_2^2$

$$= \left\| A \sum_{i=1}^{n} (z \cdot u_i) u_i \right\|_2^2 = \left\| \sum_{i=1}^{n} (z \cdot u_i) \sigma v_i \right\|_2^2$$

(Pythagorean thm)
$$= \sum_{i=1}^{n} \sigma_i^2 (z \cdot u_i)^2 = \sum_{i=1}^{k+1} \sigma_i^2 (z \cdot u_i)^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (z \cdot u_i)^2 \underset{\|z\|=1}{=} \sigma_{k+1}^2$$

Applications of SVDs.

Theorem 21.1 Every linear system $Ax = b$, $A \in \mathbb{R}^{m \times n}$ $(m > n)$, has a least squares solution x^+ of smallest norm,

i.e. $$x^+ = \arg\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

proof sketch: $\mathrm{Im}(A)$ is a subspace of $\mathbb{R}^m$, and $P_{\mathrm{Im}(A)}(b)$ (projection) is the closest $\mathrm{Im}(A)$ gets to b. Then solve for x^+ by $Ax^+ = P_{\mathrm{Im}(A)}(b)$.

Thm 21.2 $x^+ = A^+ b = U D^+ V^T b$, where $A = VDU^T$ is an SVD,

and $$(D^+)_{ij} = \begin{cases} 1/D_{ij} & \text{if } D_{ij} \neq 0 \\ 0 & \text{if } D_{ij} = 0 \end{cases}.$$

Aside: $A^+ = UD^+V^T = (A^TA)^{-1}A^T$ is the Moore-Penrose pseudo-inverse of A.
(if A has full rank)

Aside: Sometimes D is defined to be rectangular, which is why we use D^+. Easier if D is square with nonzero diagonals.