Löwdin Orthogonalization - A Natural Supplement to Gram-Schmidt

The SVD is the most generally applicable of the orthogonal-diagonal-orthogonal type matrix decompositions

==

The SVD contains a great deal of information and is very useful as a theoretical and practical tool

==

Its importance in numerical linear algebra, data compression, and least-squares problem is widely known

==

Perhaps less well-known is that the SVD yields a mathematically beautiful orthogonalization technique

1 Preliminaries

We'll assume that $A \in \mathbb{R}^{m\times n}$ with $m \geq n$.

==

Everything that follows has an obvious dual counterpart for the case $m < n$

All that follows holds, with appropriate modifications, for complex-valued matrices

==

Definition 1.1 Let $A \in \mathbb{R}^{m\times n}$. Then the *full* singular value decomposition of A is

$$A = U\Sigma V^T = \left(\begin{array}{c|c|c|c} & & & \\ & & & \\ U_1 & U_2 & \cdots & U_m \\ & & & \\ & & & \end{array}\right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \sigma_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \left(\begin{array}{c} (V_1)^T \\ \hline (V_2)^T \\ \hline \vdots \\ \hline (V_n)^T \end{array}\right)$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal

The σ_i's are the *singular values* of A, by convention arranged in nonincreasing order

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0;$$

The columns U_j of U are called *left singular vectors* of A; the columns V_j of V are called *right singular vectors* of A

Another incarnation of the SVD is the *reduced* SVD

$$A = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & & \vdots \\ \vdots & & \ddots & \\ u_{m1} & \cdots & & u_{mn} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \sigma_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & & \vdots \\ \vdots & & \ddots & \\ v_{n1} & \cdots & & v_{nn} \end{pmatrix}$$

where the matrix U is no longer square (so it can't be orthogonal) but still has orthonormal columns, Σ is square and diagonal, and V is still orthogonal

==

It is the reduced SVD which we'll use for our orthogonalization technique

==

The *Frobenius norm* of A is $\|A\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$

==

Lemma 1.2 Let $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $x \in \mathbb{R}^n$. Let P be a matrix in $\mathbb{R}^{n \times m}$ with orthonormal rows, and Q be a matrix in $\mathbb{R}^{m \times n}$ with orthonormal columns. Then

$$\begin{aligned} \|AP\|_F &= \|A\|_F \qquad \text{and} \qquad (1) \\ \|Qx\|_2 &= \|x\|_2 \qquad (2) \end{aligned}$$

Note that although the singular values of A are uniquely determined, the left (or right) singular vectors are only determined up to sign

==

If we fix signs for V_j, then the signs for U_j are determined

****

2 Löwdin (Symmetric) Orthogonalization

For nonsingular A with reduced SVD $A = U\Sigma V^T$, the matrix $L := UV^T$ is called the *Löwdin orthogonalization* of the matrix A

==

Discovered (in a non-SVD form) by a Swedish chemist, Per-Olov Löwdin, for the purpose of orthogonalizing hybrid electron orbitals

==

L is unique since any sequence of sign choices for the columns of V determines a sequence of signs for the columns of U

$$\begin{aligned} L_{ij} &= U_{i1}(V^T)_{1j} + U_{i2}(V^T)_{2j} + U_{i3}(V^T)_{3j} + \cdots + U_{in}(V^T)_{nj} \\ &= U_{i1}V_{j1} + U_{i2}V_{j2} + U_{i3}V_{j3} + \cdots + U_{in}V_{jn} \end{aligned}$$

==

Like Gram-Schmidt orthogonalization, it takes as input a linearly independent set (the columns of A) and outputs an orthonormal set (the columns of UV^T)

==

(Classical) Gram-Schmidt is unstable due to repeated subtractions; Modifed Gram-Schmidt (usually) remedies this

==

But occasionally we want to disturb the original set of vectors as little as possible

****

Theorem 2.1 Let $m \geq n$, $A \in \mathbb{R}^{m\times n}$, and suppose that A has full rank. Over all matrices $Q \in \mathbb{R}^{m\times n}$ with orthonormal columns, $\|A - Q\|_F$ is minimized when $Q = UV^T$.

Proof: Let $Q \in \mathbb{R}^{m\times n}$ with $Q^TQ = I_{n\times n}$. Fix the reduced SVD of A be $A = U\Sigma V^T$ by fixing a sequence of signs for the columns of V. By Lemma 1.2, we have

$$\begin{aligned} \|A - Q\|_F &= \|U\Sigma V^T - Q\|_F \\ &= \|U\Sigma - QV\|_F \end{aligned}$$

===

The problem we must solve is to specify

$$\arg\left\{\min\left\{\ \|U\Sigma - QV\|_F\ \middle|\ Q^TQ = I_{n\times n}\right\}\right\} \tag{3}$$

or, equivalently (because $f(x) = x^2$ is increasing),

$$\arg\left\{\min\left\{\ \|U\Sigma - QV\|_F^2\ \middle|\ Q^TQ = I_{n\times n}\right\}\right\}$$

===

Denote $X := QV$ and note that

$$\begin{aligned} &\arg\left\{\min\left\{\ \|U\Sigma - QV\|_F^2\ \middle|\ Q^TQ = I_{n\times n}\right\}\right\} \\ =\ & V^T\ \left(\arg\left\{\min\left\{\ \|U\Sigma - X\|_F^2\ \middle|\ X^TX = I_{n\times n}\right\}\right\}\right) \end{aligned}$$

Thus we seek to solve

$$\arg\left\{\min\left\{\ \|U\Sigma - X\|_F^2\ \middle|\ X^T X = I_{n\times n}\right\}\right\} \tag{4}$$

==

We have

$$\begin{aligned}\|U\Sigma - X\|_F^2 &= \|(U\Sigma - X)_1\|_2^2 + \|(U\Sigma - X)_2\|_2^2 + \cdots + \|(U\Sigma - X)_n\|_2^2 \\ &= \|(\sigma_1 U_1 - X_1)\|_2^2 + \|(\sigma_2 U_2 - X_2)\|_2^2 + \cdots + \|(\sigma_n U_n - X_n)\|_2^2.\end{aligned}$$

==

Suppose we minimize each of the $\|\sigma_j U_j - X_j\|_2^2$ individually. Will the column-wise concatenation of such solutions yield a solution to (4)? Yes, if the constraint

$$X^T X = I_{n\times n} \qquad \text{is satisfied.} \tag{5}$$

Consider the j^{th} column in $U\Sigma - X$:

$$(U\Sigma - X)_j = (\sigma_j U_j - X_j) = \begin{pmatrix} \sigma_j u_{1j} - x_{1j} \\ \sigma_j u_{2j} - x_{2j} \\ \vdots \\ \sigma_j u_{nj} - x_{nj} \end{pmatrix}$$

==

Now

$$\begin{aligned} \|(U\Sigma - X)_j\|_2^2 &= \sum_{k=1}^{n} (\sigma_j u_{kj} - x_{kj})^2 \\ &= \sigma_j^2 \sum_{k=1}^{n} u_{kj}^2 - 2\sigma_j \sum_{k=1}^{n} u_{kj} x_{kj} + \sum_{k=1}^{n} x_{kj}^2 \\ &= \sigma_j^2 - 2\sigma_j \sum_{k=1}^{n} u_{kj} x_{kj} + 1 \quad \text{(by Lemma 1.2).} \end{aligned}$$

==

Since σ_j, 1, and 2 are positive constants,

$$\begin{aligned} & \arg\left\{ \min\left\{ \sigma_j^2 - 2\sigma_j \sum_{k=1}^{n} u_{kj} x_{kj} + 1 \;\middle|\; \|X_j\|_2 = 1 \right\} \right\} \\ = \; & \arg\left\{ \max\left\{ \sum_{k=1}^{n} u_{kj} x_{kj} \;\middle|\; \|X_j\|_2 = 1 \right\} \right\}. \end{aligned}$$

This is clearly maximized when $X_j = U_j$, so the constraint $X^T X = I_{n\times n}$ is satisfied and

$$X = QV = U \quad \text{solves the arg-min problem (4),} \quad \text{so}$$

$$Q = UV^T \quad \text{solves the arg-min problem (3).}$$

$\square$

==

In the case that rank$(A) < n$, L still solves (3) but is not the *unique* minimizer.

==

Example 2.2

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & -1 \end{pmatrix} \qquad L = \begin{pmatrix} \dfrac{\frac{\sqrt{2}}{2}}{\sqrt{4-2\sqrt{2}}} & \dfrac{-1+\frac{\sqrt{2}}{2}}{\sqrt{4-2\sqrt{2}}} & -\dfrac{1}{\sqrt{2}} \\ \dfrac{\frac{2}{\sqrt{2}}-1}{\sqrt{4-2\sqrt{2}}} & \dfrac{1}{\sqrt{4-2\sqrt{2}}} & 0 \\ \dfrac{-\frac{\sqrt{2}}{2}}{\sqrt{4-2\sqrt{2}}} & \dfrac{1-\frac{\sqrt{2}}{2}}{\sqrt{4-2\sqrt{2}}} & -\dfrac{1}{\sqrt{2}} \end{pmatrix}$$

$$U = \begin{pmatrix} \dfrac{1}{2} & -\dfrac{1}{\sqrt{2}} & -\dfrac{1}{2} \\ \dfrac{1}{\sqrt{2}} & 0 & \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{2} & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{2} \end{pmatrix} \qquad V^T = \begin{pmatrix} \dfrac{1}{\sqrt{4-2\sqrt{2}}} & \dfrac{-1+\sqrt{2}}{\sqrt{4-2\sqrt{2}}} & 0 \\ 0 & 0 & 1 \\ \dfrac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} & \dfrac{1}{\sqrt{4-2\sqrt{2}}} & 0 \end{pmatrix}$$

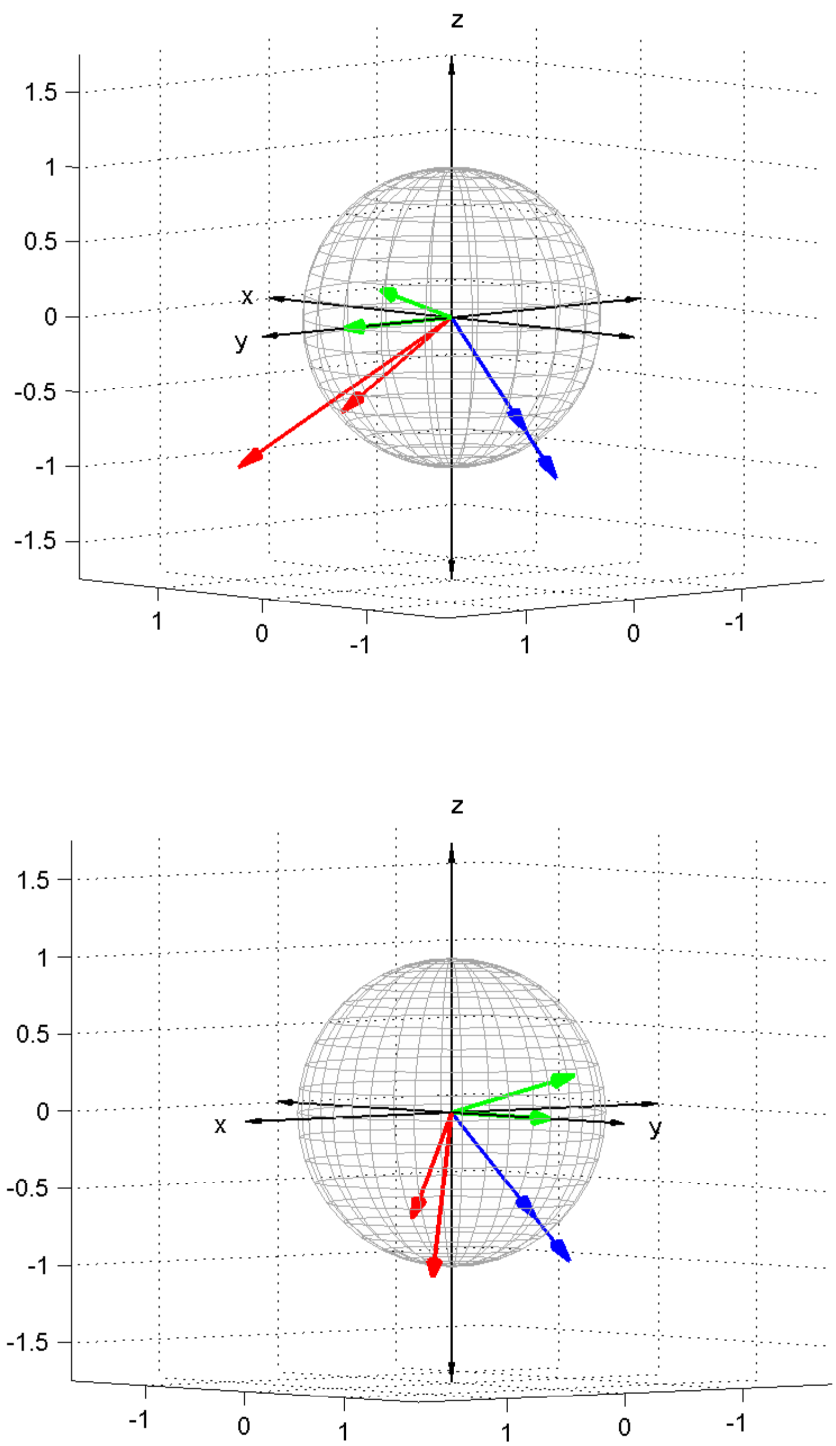

Figure 1: The columns of $L = UV^T$ and the columns of A

3 Why Include This In Your Linear Algebra Course?

There are *a lot* of orthogonalization techniques - in fact, U from the reduced $A = U\Sigma V^T$ is a perfectly good orthogonalization of A

==

Gram-Schmidt requires the choice of distinguished (initial) vector, but Löwdin orthogonalization is egalitarian in the sense that it gives all vectors equal footing

==

The Löwdin orthogonalization L of a matrix A with linearly independent columns optimally resembles A (and of course $-L$ is maximally distant from A)

==

The proof of Theorem 2.1 uses simple optimization and is just plain fun; it's slightly simpler in the case of square A

==

Can present in class the proof of the square case, then assign a project in which students find where in the non-square case the proof breaks down, and repair it

==

Time permitting, investigation into the rank-deficient case is worthwhile

Scott Beaver

Western Oregon University

beavers@wou.edu

http://www.wou.edu/~beavers