SINGULAR VALUE DECOMPOSITIONS

Prof. Dan A. Simovici

UMB

Let $A \in \mathbb{R}^{m \times n}$ be a matrix. A singular triplet of A is a triplet $(\sigma, \mathbf{u}, \mathbf{v})$ such that

- $\sigma \in \mathbb{R}_{>0}$, $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^m$,
- $A\mathbf{u} = \sigma\mathbf{v}$ and
- $A'\mathbf{v} = \sigma\mathbf{u}$.

σ is a singular value of A, $\mathbf{u}$ is a left singular vector and $\mathbf{v}$ is a right singular vector.

For a singular triplet of A we have

$$A'A\mathbf{u} = \sigma A'\mathbf{v} = \sigma^2\mathbf{u} \text{ and } AA'\mathbf{v} = \sigma A\mathbf{u} = \sigma^2\mathbf{v}.$$

Therefore, σ^2 is both an eigenvalue of AA' and an eigenvalue of $A'A$.

Example

Let A be the matrix

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \cos\beta & \sin\beta \end{pmatrix}.$$

We have $\det(A) = \sin(\beta - \alpha)$, so the eigenvalues of $A'A$ are the roots of the equation $\lambda^2 - 2\lambda + \sin^2(\beta - \alpha) = 0$, that is, $\lambda_1 = 1 + \cos(\beta - \alpha)$ and $\lambda_2 = 1 - \cos(\beta - \alpha)$. Therefore, the singular values of A are $\sigma_1 = \sqrt{2}\left|\cos\frac{\beta-\alpha}{2}\right|$ and $\sigma_2 = \sqrt{2}\left|\sin\frac{\beta-\alpha}{2}\right|$.

It is easy to see that a unit left singular vector that corresponds to the eigenvalue $1 + \cos(\beta - \alpha)$ is

$$\mathbf{u} = \begin{pmatrix} \cos\frac{\alpha+\beta}{2} \\ \sin\frac{\alpha+\beta}{2} \end{pmatrix},$$

which corresponds to the average direction of the rows of A.

We noted that the eigenvalues of a positive semi-definite matrix are non-negative numbers.

- Since both AA' and $A'A$ are positive semi-definite matrices for $A \in \mathbb{R}^{m \times n}$, the spectra of these matrices consist of non-negative numbers $\lambda_1, \ldots, \lambda_n$.
- AA' and $A'A$ have the same rank r and therefore, the same number r of non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$.
- The singular values of A have the form $\sqrt{\lambda_1} \geqslant \cdots \geqslant \sqrt{\lambda_r}$.

We use the notation $\sigma_i = \sqrt{\lambda_i}$ for $1 \leqslant i \leqslant r$ and assume that $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$.

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a matrix having the singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_n$. If λ is an eigenvalue of A, then $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$.

Proof

Let $\mathbf{u}$ be an unit eigenvector for the eigenvalue λ. Since $A\mathbf{u} = \lambda\mathbf{u}$ it follows that $(A'A\mathbf{u}, bfu) = (A\mathbf{u}, A\mathbf{u}) = \overline{\lambda}\lambda(\mathbf{u}, bfu) = \overline{\lambda}\lambda = |\lambda|^2$. The matrix $A'A$ is Hermitian and its largest and smallest eigenvalues are σ_1^2 and σ_n^2, respectively. Thus, $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$.

The SVD Theorem

Theorem

If $A \in \mathbb{R}^{m \times n}$ *is a matrix and rank*$(A) = r$*, then* A *can be factored as* $A = UDV'$*, where* $U \in \mathbb{R}^{m \times m}$ *and* $V \in \mathbb{R}^{n \times n}$ *are orthogonal matrices, and* $D = \mathit{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{m \times n}$*, where* $\sigma_1 \geqslant \ldots \geqslant \sigma_r$ *are real positive numbers.*

Proof

The square matrix $A'A \in \mathbb{R}^{n\times n}$ has the same rank r as the matrix A and is positive semidefinite. Therefore, there are r positive eigenvalues of this matrix, denoted by $\sigma_1^2, \ldots, \sigma_r^2$, where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be the corresponding pairwise orthogonal unit eigenvectors in $\mathbb{R}^n$.

We have $A'A\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i$ for $1 \leqslant i \leqslant r$. Define $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r \ \mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$ by completing the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to an orthogonal basis

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

for $\mathbb{R}^n$. If $V_1 = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r)$ and $V_2 = (\mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$, we can write $V = (V_1 \ V_2)$.

Proof (cont'd)

The equalities involving the eigenvectors can now be written as $A'AV_1 = V_1E^2$, where $E = \text{diag}(\sigma_1, \ldots, \sigma_r)$.
Define $U_1 = AV_1E^{-1} \in \mathbb{R}^{m \times r}$. We have $U_1' = S^{-1}V_1'A'$, so

$$U_1'U_1 = S^{-1}V_1'A'AV_1E^{-1} = E^{-1}V_1'V_1E^2E^{-1} = I_r,$$

which shows that the columns of U_1 are pairwise orthogonal unit vectors.
Consequently, $U_1'AV_1E^{-1} = I_r$, so $U_1'AV_1 = E$.

Proof (cont'd)

If $U_1 = (\mathbf{u}_1 \ \cdots, \mathbf{u}_r)$, let $U_2 = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m)$ be the matrix whose columns constitute the extension of the set $\{\mathbf{u}_1 \ \cdots, \mathbf{u}_r\}$ to an orthogonal basis of $\mathbb{R}^m$. Define $U \in \mathbb{R}^{m\times m}$ as $U = (U_1 \ U_2)$. Note that

$$
\begin{aligned}
U'AV &= \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} A(V_1 \ V_2) = \begin{pmatrix} U_1'AV_1 & U_1'AV_2 \\ U_2'AV_1 & U_2'AV_2 \end{pmatrix} \\
&= \begin{pmatrix} U_1'AV_1 & U_1'AV_2 \\ U_2'AV_1 & U_2'AV_2 \end{pmatrix} = \begin{pmatrix} U_1'AV_1 & O \\ O & O \end{pmatrix} = \begin{pmatrix} E & O \\ O & O \end{pmatrix},
\end{aligned}
$$

which is the desired decomposition.

The Thin SVD Decomposition

Corollary

Let $A \in \mathbb{C}^{m \times n}$ be a matrix having non-zero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $r \leq \min\{m, n\}$. Then, A can be factored as $A = UDV^H$, where $U \in \mathbb{C}^{m \times r}$ and $V \in \mathbb{C}^{n \times r}$ are matrices having orthonormal sets of columns and D is the diagonal matrix

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

Example

The thin SVD decomposition of the matrix A

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

is

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Since U and V in the thin SVD have orthonormal columns we have $U'U = V'V = I$.

The value of a invariant norm of a matrix depends only on its singular values.

Theorem

Let $A \in \mathbb{R}^{m \times n}$ be a matrix and let $A = UDV'$ be the singular value decomposition of A. If $\| \cdot \|$ is a unitarily invariant norm, then

$$\| A \| = \| D \| = \| \mathit{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \| .$$

$\|\cdot\|_2$ and $\|\cdot\|_F$ are unitarily invariant. Therefore, the Frobenius norm can be written as

$$\| A \|_F = \sqrt{\sum_{i=1}^{r} \sigma_r^2}.$$

and $\|A\|_2 = \sigma_1$.

Matrices $A, B \in \mathbb{R}^{m \times n}$ are orthogonally equivalent (written $A \sim B$) if $A = W_1' B W_2$, where $W_1 \in \mathbb{R}^{m \times m}$, $W_2 \in \mathbb{R}^{n \times n}$ are orthogonal matrices.

Theorem

Let A and B be two matrices in $\mathbb{R}^{m \times n}$. If $A \sim B$, then they have the same singular values.

Proof

Since $A \sim B$, that is, $A = W_1' B W_2$ for some orthogonal matrices W_1 and W_2. If A has the SVD $A = U'\text{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)V$, then

$$B = W_1 A W_2' = (W_1 U')\text{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)(V W_2').$$

Since $W_1 U'$ and $V W_2'$ are both orthogonal matrices, it follows that the singular values of B are the same as the singular values of A.

Let $\mathbf{v} \in \mathbb{R}^n$ be an eigenvector of the matrix $A'A$ that corresponds to a non-zero, positive eigenvalue σ^2, that is, $A'A\mathbf{v} = \sigma^2\mathbf{v}$.
Define $\mathbf{u} = \frac{1}{\sigma}A\mathbf{v}$. We have $A\mathbf{v} = \sigma\mathbf{u}$. Also,

$$A'\mathbf{u} = A'\left(\frac{1}{\sigma}A\mathbf{v}\right) = \sigma\mathbf{v}.$$

This implies $AA'\mathbf{u} = \sigma^2\mathbf{u}$, so $\mathbf{u}$ is an eigenvector of AA' that corresponds to the same eigenvalue σ^2.
Conversely, if $\mathbf{u} \in \mathbb{R}^m$ is an eigenvector of the matrix AA' that corresponds to a non-zero, positive eigenvalue σ^2, we have $AA'\mathbf{u} = \sigma^2\mathbf{u}$. Thus, if $\mathbf{v} = \frac{1}{\sigma}A\mathbf{u}$ we have $A\mathbf{v} = \sigma\mathbf{u}$ and $\mathbf{v}$ is an eigenvector of $A'A$ for the eigenvalue σ^2.

Theorem

Let $A \in \mathbb{R}^{m\times n}$ *be a matrix such that* $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$ *is the non-increasing sequence of singular values of* A*. For* $1 \leqslant k \leqslant r$ *we have*

$$\begin{aligned}
\sigma_k &= \min_{\dim(S)=n-k+1} \max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in S \text{ and } \| \mathbf{x} \|_2 = 1\} \\
\sigma_k &= \max_{\dim(T)=k} \min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \in T \text{ and } \| \mathbf{x} \|_2 = 1\},
\end{aligned}$$

where S *and* T *range over subspaces of* $\mathbb{R}^n$.

We give the argument only for the second equality of the theorem; the first can be shown in a similar manner.
We saw that σ_k equals the square root of k^{th} largest absolute value of the eigenvalue $|\lambda_k|$ of the matrix $A'A$. By Courant-Fisher Theorem, we have

$$
\begin{aligned}
\lambda_k &= \max_{dim(T)=k} \min_{\mathbf{x}} \{\mathbf{x}' A' A \mathbf{x} \mid \mathbf{x} \in T \text{ and } \parallel \mathbf{x} \parallel_2 = 1\} \\
&= \max_{dim(T)=k} \min_{\mathbf{x}} \{\parallel A\mathbf{x} \parallel_2^2 \mid \mathbf{x} \in T \text{ and } \parallel \mathbf{x} \parallel_2 = 1\},
\end{aligned}
$$

which implies the second equality of the theorem.

The equalities established above can be rewritten as

$$\begin{aligned}
\sigma_k &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{k-1} \text{ and } \| \mathbf{x} \|_2 = 1\} \\
&= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}} \min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{n-k} \text{ and } \| \mathbf{x} \|_2 = 1\}.
\end{aligned}$$

Corollary

The smallest singular value of a matrix $A \in \mathbb{R}^{m \times n}$ *equals*

$$\min\{\parallel A\mathbf{x} \parallel_2 \mid \mathbf{x} \in \mathbb{R}^n \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\}.$$

The largest singular value of a matrix $A \in \mathbb{R}^{m \times n}$ *equals*

$$\max\{\parallel A\mathbf{x} \parallel_2 \mid \mathbf{x} \in \mathbb{R}^n \textit{ and } \parallel \mathbf{x} \parallel_2 = 1\}.$$

A prelimnary result

Lemma

Let $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r'$ be the SVD of a matrix $A \in \mathbb{R}^{m \times n}$, where $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$. For every k, $1 \leqslant k \leqslant r$ the matrix $B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i'$ has rank k.

The null space of the matrix $B(k)$ consists of those vectors $\mathbf{x}$ such that $\sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i' \mathbf{x} = \mathbf{0}$. The linear independence of the vectors $\mathbf{u}_i$ and the fact that $\sigma_i > 0$ for $1 \leqslant i \leqslant r$ implies the equalities $\mathbf{v}_i' \mathbf{x} = \mathbf{0}$ for $1 \leqslant i \leqslant r$. Thus,

$$\text{NullSp}(B(k)) = \text{NullSp}\left((\mathbf{v}_1 \ \cdots \ \mathbf{v}_k)\right).$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent it follows that $\dim(\text{NullSp}(B(k)) = n - k$, which implies $\mathit{rank}(B(k)) = k$ for $1 \leqslant k \leqslant r$.

Eckhart-Young Theorem

The SVD allows us to find the best approximation of of a matrix by a matrices of limited rank.

Theorem

Let $A \in \mathbb{R}^{m\times n}$ be a matrix whose sequence of non-zero singular values is $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$. Suppose that A can be written as

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r'.$$

Let $B(k) \in \mathbb{R}^{m\times n}$ be the matrix $B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i'$.
If $r_k = \inf\{\|A - X\|_2 \mid X \in \mathbb{R}^{m\times n}$ and rank$(X) \leqslant k\}$, then

$$\|A - B(k)\|_2 = r_k = \sigma_{k+1},$$

for $1 \leqslant k \leqslant r$, where $\sigma_{r+1} = 0$ and $B(k)$ is the best approximation of A among the matrices of rank no larger than k in the sense of the norm $\|\cdot\|_2$.

Proof

Observe that

$$A - B(k) = \sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i',$$

and the largest singular value of the matrix $\sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i'$ is σ_{k+1}. Since σ_{k+1} is the largest singular value of $A - B(k)$ we have $\|A - B(k)\|_2 = \sigma_{k+1}$ for $1 \leqslant k \leqslant r$.

We prove now that for every matrix $X \in \mathbb{R}^{m \times n}$ such that $rank(X) \leqslant k$, we have $\|A - X\|_2 \geqslant \sigma_{k+1}$. Since $\dim(\text{NullSp}(X)) = n - rank(X)$, it follows that $\dim(\text{NullSp}(X)) \geqslant n - k$. If T is the subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$, we have $\dim(T) = k + 1$. Since $\dim(\text{NullSp}(X)) + \dim(T) > n$, the intersection of these subspaces contains a non-zero vector and, without loss of generality, we can assume that this vector is a unit vector $\mathbf{x}$.

We have $\mathbf{x} = a_1\mathbf{v}_1 + \cdots a_k\mathbf{v}_k + a_{k+1}\mathbf{v}_{k+1}$ because $\mathbf{x} \in T$. The orthogonality of $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$ implies $\parallel \mathbf{x} \parallel_2^2 = \sum_{i=1}^{k+1} |a_i|^2 = 1$. Since $\mathbf{x} \in \mathsf{NullSp}(X)$, we have $X\mathbf{x} = \mathbf{0}$, so

$$(A - X)\mathbf{x} = A\mathbf{x} = \sum_{i=1}^{k+1} a_i A\mathbf{v}_i = \sum_{i=1}^{k+1} a_i \sigma_i \mathbf{u}_i.$$

Thus, we have

$$\|(A - X)\mathbf{x}\|_2^2 = \sum_{i=1}^{k+1} |\sigma_i a_i|^2 \geqslant \sigma_{k+1}^2 \sum_{i=1}^{k+1} |a_i|^2 = \sigma_{k+1}^2,$$

because $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are also orthonormal. This implies $\|A - X\|_2 \geqslant \sigma_{k+1} = \|A - B(k)\|_2$.

The matrix $B(k)$ provides an optimal approximation of A not only with respect to $\|\cdot\|_2$ but also relative to the Frobenius norm.

Theorem

$B(k)$ is the best approximation of A among matrices of rank no larger than k in the sense of the Frobenius norm.

Proof

Note that $\| A - B(k) \|_F^2 = \| A \|_F^2 - \sum_{i=1}^k \sigma_i^2$.
Let X be a matrix of rank k, which can be written as $X = \sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i'$.
Without loss of generality we may assume that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are orthonormal. If this is not the case, we can use the Gram-Schmidt algorithm to express then as linear combinations of orthonormal vectors, replace these expressions in $\sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i'$ and rearrange the terms.
The Frobenius norm of $A - X$ can be written as

$$\begin{aligned}
\| A - X \|_F^2 &= trace\left(\left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}'\right)' \left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}'\right)\right) \\
&= trace\left(A'A + \sum_{i=1}^k (\mathbf{y}_i - A'\mathbf{x}_i)(\mathbf{y}_i - A'\mathbf{x}_i)' - \sum_{i=1}^k A'\mathbf{x}_i\mathbf{x}_i' A\right).
\end{aligned}$$

Proof (cont'd)

Taking into account that $\sum_{i=1}^{k}(\mathbf{y}_i - A'\mathbf{x}_i)(\mathbf{y}_i - A'\mathbf{x}_i)'$ is a real non-negative number and that $\sum_{i=1}^{k} A'\mathbf{x}_i\mathbf{x}_i'A = \parallel A\mathbf{x}_i \parallel_F^2$ we have

$$\parallel A - X \parallel_F^2 \quad \geq \quad trace\left(A'A - \sum_{i=1}^{k} A'\mathbf{x}_i\mathbf{x}_i'A\right) = \parallel A \parallel_F^2 - trace\left(\sum_{i=1}^{k} A'\mathbf{x}_i\mathbf{x}_i'A\right)$$

Let $A = U\mathrm{diag}(\sigma_1, \ldots, \sigma_n)V'$ be the singular value decomposition of A. If $V = (V_1\ V_2)$, where V_1 has k columns $\mathbf{v}_1, \ldots, \mathbf{v}_k$, $D_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$ and $D_2 = \mathrm{diag}(\sigma_{k+1}, \ldots, \sigma_n)$, then

$$\begin{aligned} A'A &= VD'U'UDV' = (V_1\ V_2)\begin{pmatrix} D_1^2 & O \\ O & D_2^2 \end{pmatrix}\begin{pmatrix} V_1' \\ V_2' \end{pmatrix} \\ &= V_1 D_1^2 V_1' + V_2 D_2^2 V_2'. \end{aligned}$$

and $A'A = VD^2V'$.

Proof (cont'd)

These equalities allow us to write:

$$\begin{aligned}
\| A\mathbf{x}_i \|_F^2 &= trace(\mathbf{x}_i' A' A \mathbf{x}_i) \\
&= trace\left(\mathbf{x}_i' V_1 D_1^2 V_1' \mathbf{x}_i + \mathbf{x}_i' V_2 D_2^2 V_2' \mathbf{x}_i\right) \\
&= \| D_1 V_1' \mathbf{x}_i \|_F^2 + \| D_2 V_2' \mathbf{x}_i \|_F^2 \\
&= \sigma_k^2 + \left(\| D_1 V_1' \mathbf{x}_i \|_F^2 - \sigma_k^2 \| V_1' \mathbf{x}_i \|_F^2\right) \\
&\quad - \left(\sigma_k^2 \| V_2' \mathbf{x}_i \|_F^2 - \| D_2 V_2' \mathbf{x}_i \|_F^2)\right) - \sigma_k^2 (1 - \| V' \mathbf{x}_i \|).
\end{aligned}$$

Since $\| V' \mathbf{x}_i \|_F^1 = 1$ (because $\mathbf{x}_i$ is an unit vector and V is an unitary matrix) and $\sigma_k^2 \| V_2' \mathbf{x}_i \|_F^2 - \| D_2 V_2' \mathbf{x}_i \|_F^2 \geq 0$, it follows that

$$\| A\mathbf{x}_i \|_F^2 \leqslant \sigma_k^2 + \left(\| D_1 V_1' \mathbf{x}_i \|_F^2 - \sigma_k^2 \| V_1' \mathbf{x}_i \|_F^2\right).$$

Proof (cont'd)

Consequently,

$$\begin{aligned}
\sum_{i=1}^{k} \| A\mathbf{x}_i \|_F^2 &\leq k\sigma_k^2 + \sum_{i=1}^{k} \left(\| D_1 V_1' \mathbf{x}_i \|_F^2 - \sigma_k^2 \| V_1' \mathbf{x}_i \|_F^2 \right) \\
&= k\sigma_k^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} (\sigma_j^2 - \sigma_k^2) |\mathbf{v}_j' \mathbf{x}_i|^2 \\
&= \sum_{j=1}^{k} \left(\sigma_k^2 + (\sigma_j^2 - \sigma_k^2) \sum_{i=1}^{k} |\mathbf{v}_j \mathbf{x}_i|^2 \right) \\
&\leq \sum_{j=1}^{k} (\sigma_k^2 + (\sigma_j^2 - \sigma_k^2)) = \sum_{j=1}^{k} \sigma_j^2,
\end{aligned}$$

which concludes the argument.