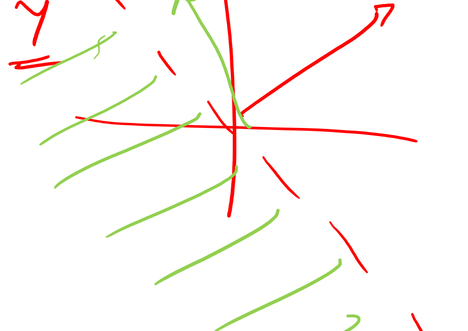
Linear Algebra & Convex Optimization – Lecture 9

References : Introduction to Linear Algebra, Gilbert Strang; Online References

Positive Semi-Definite (PSD) Matrices



A matrix is **positive semi-definite** if it is ~~symmetric~~ and all ~~eigen values are~~ non-negative

$x^T \lambda x$

$= \lambda x^T x$

- A matrix M is positive semi definite if $x^T M x \geq 0$ for all eigenvectors of M.

  $x^T M x = x^T \lambda x = \lambda x^T x$ (if x is an eigen vector)

  $\lambda \|x\|^2 \geq 0$ for all $x \neq 0$ & $\lambda \geq 0$

$M = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T \cdots \lambda_n e_n e_n^T$

$x^T M x = x^T \left[\lambda_1 e_1 e_1^T + \cdots \lambda_n e_n e_n^T \right] x$

$= \lambda_1 x^T e_1 e_1^T x + \cdots \lambda_n x^T e_n e_n^T x$

- A matrix M is positive semi definite if $x^T M x \geq 0$ for all non-zero vectors

$x^T (A + B) x$

Positive Definite Matrices:

A matrix is **positive definite** if it is symmetric and all eigen values are positive

$\lambda > 0$ for all λ of $B \implies x^T B x > 0$ for all $x \neq 0$

Poll:

If A & B are Positive definite matrices will A+B a positive definite matrix ?

yes

$A v_{max} = \lambda_{max} v_{max}$

Rayleigh Quotient

Given a symmetric Matrix A and any vector x , Rayleigh Quotient is defined as $R(A,x) = \dfrac{x^T A x}{x^T x}$

Among all x with $\|x\| = 1$, the x that maximizes , Rayleigh Quotient is the eigen vector , v_{max} that corresponds to maximum eigen value, λ_{max} of A

$$R(A, v_{\max}) = v_{max}^T A v_{\max} = v_{max}^T \lambda_{max} v_{\max} = \lambda_{max} v_{max}^T v_{\max} = \lambda_{max}$$

e_i^T e_i

$$R(A,x) \in [\lambda_{min}, \lambda_{max}]$$

$(0 \cdots 1 \cdots 0)$ A $\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$

$x^T A x > 0$

$e_i^T [A] e_i$

$$\begin{vmatrix} 1-\lambda & 10 \\ 10 & 1-\lambda \end{vmatrix}$$

$(1-\lambda)^2 - 100 = 0$

$1-\lambda = 10$

$\lambda = -9$

Poll:

Can positive definite matrices have negative diagonal entries?

No

Poll:

If all the entries in a matrix A is positive will it be positive definite matrix ?

No

$A = \begin{pmatrix} 1 & 10 \\ 10 & 1 \end{pmatrix}$ is a counter example with $\lambda = -9$.

Positive definite matrices must have stronger diagonals

Power Iteration Method to Find Maximum Eigen Value/Vector

if A is diagonalizable and has dominant eigenvalue,
then power iteration sequence $Ax, A^2x, A^3x, \ldots$ converges to the dominant eigenvector (scaled)

since A is diagonalizable, it has n linearly independent eigenvectors $v_1, \ldots, v_n$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$.
let v_1 and λ_1 be dominant.

any vector x_0 can be represented as $x_0 = c_1v_1 + \ldots + c_nv_n$

$$Ax_0 = c_1Av_1 + \ldots + c_nAv_n = c_1\lambda_1v_1 + \ldots + c_n\lambda_nv_n$$

$$A^kx_0 = c_1\lambda_1^kv_1 + \ldots + c_n\lambda_n^kv_n$$

$$A^kx_0 = \lambda_1^k\left[c_1v_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 \ldots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right]$$

since λ_1 is dominating, the ratios $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$ as $k \to \infty$ for all i

$A^kx_0 = \lambda_1^kc_1v_1$ and it gets better as k grows

How do we find eigen value of the dominant eigen vector?

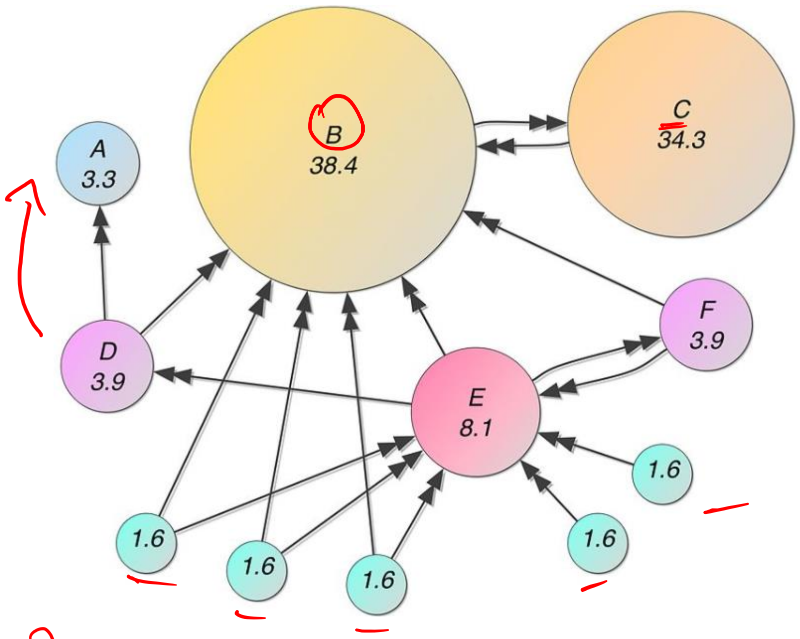
$$\frac{p^TAp}{p^Tp}$$

Dominant Eigen Vector Application : Page Rank Algorithm (Google)

$L_{ij} = 1$ if webpage j links to webpage i (written $j \to i$),

$$L = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1n} \\ L_{21} & L_{22} & \dots & L_{2n} \\ \vdots & & & \\ L_{n1} & L_{n2} & \dots & L_{nn} \end{pmatrix}$$

$m_j = \sum_{k=1}^{n} L_{kj}$
the total number of webpages that j links to



Page Rank of Web pages

$i = 2$

Page Rank of Page *i* :

$$p_i = \sum_{j \to i} \frac{p_j}{m_j} \qquad p_i = \sum_{j=1}^{n} \frac{L_{ij}}{m_j} p_j$$

$$p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} \qquad Ap = p$$

Importance Matrix

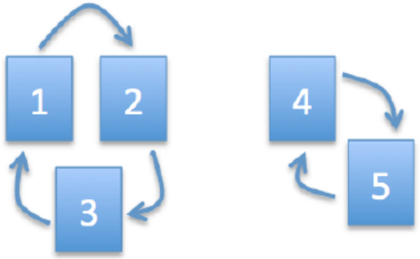
$$Ap = \begin{bmatrix} \frac{L_{11}}{m_1} & \frac{L_{12}}{m_2} & \dots & \frac{L_{1n}}{m_n} \\ & \vdots & & \\ \frac{L_{n1}}{m_1} & \frac{L_{n2}}{m_2} & \dots & \frac{L_{nn}}{m_n} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$$

A is a column stochastic Matrix with maximum eigen value $\lambda_1 = 1$

Stochastic - square matrix whose columns are probability vectors

Dominant Eigen Vector Application : Page Rank Algorithm (Google)

Importance Matrix (5×5)



$$A = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2 & 3 & & \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \\ \\ \end{array} & \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Let p be the dominant eigen vector

p is calculated using power iteration

Google Matrix:

$$A = \frac{0.15}{5} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} + 0.85 \cdot \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.03 & 0.03 & 0.88 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.88 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.03 \end{pmatrix}$$

A is a column stochastic Matrix with maximum eigen value **1**

A has millions and millions of rows and columns.

$$p^{(1)} = Ap^{(0)}$$
$$p^{(2)} = Ap^{(1)}$$
$$\vdots$$
$$p^{(t)} = Ap^{(t-1)}$$

Analysis of Rectangular Matrices

Four Fundamental Subspaces of $m \times n$ Matrix A

$$\begin{array}{cc} & n \\ m & \left[\begin{array}{c} A \end{array}\right] \end{array} \Big\} r$$

r

$Ax = 0$

$\mathbb{R}^n$

Row Space

dim(Row Space) = r
= #pivots
= no if independent rows

Null Space
N(A)

dim(Null Space) = n- r

= #free variables
in Ax

$\mathbb{R}^m$

Column Space
C(A)

dim(C(A)) = r
= #pivots
= no if independent columns

Left Null Space
$N(A^T)$

dim(Left Null Space) = m- r

= #free variables in A^Tx

Orthogonal Sub-spaces

S , T are 2 sub-spaces of $\mathbb{R}^n$

We say $S \perp T$ when every vector $s \in S$ & $t \in T$, s ⊥ t

$$\boldsymbol{RowSpace}\ (\boldsymbol{A}) \perp \boldsymbol{NullSpace}(\boldsymbol{A})$$

$$\boldsymbol{ColumnSpace}\ (A) \perp \boldsymbol{NullSpace}(A^T)$$

$$Ax = 0 \Rightarrow \begin{bmatrix} \leftarrow & \tilde{a}_1^T & \rightarrow \\ \leftarrow & \tilde{a}_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & \tilde{a}_m^T & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ x \\ \downarrow \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{a}_1^T x \\ \vdots \\ \tilde{a}_m^T x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow (c_1 \tilde{a}_1^T + c_2 \tilde{a}_2^T + \cdots c_m \tilde{a}_m^T) x = 0$$

Singular Value Decomposition of $m \times n$ Matrix A

Eigenvalue Decomposition

Problems with general Eigendecomposition $A = S\,\Lambda\,S^{-1}$:

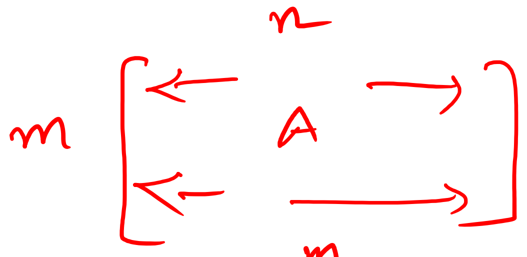
- doesn't work with rectangular matrices
- S is usually not orthonormal (unless A is symmetric)

Finding SVD

Goal:

- find orthonormal bases in the row space of A as well as in the column space of A
- s.t. A maps from row space basis to the column space basis

Orthogonalization of Rowspace of $m \times n$ Matrix A



let r be the rank of A
select orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in $\mathbb{R}^n$ s.t. it spans the Row Space of A
e.g. using the Gram-Schmidt Process
continue the process to find $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$ s.t it spans the Nullspace of A

$u_i \in \mathbb{R}^m$

for $i = 1..r$ define $\mathbf{u}_i$ as $A\mathbf{v}_i$

$\underset{m\times n}{A}\ \underset{n\times 1}{v_i} = \underset{m\times 1}{u_i}$

Here $\{\ \mathbf{v}_i\ \}$ are orthogonal by construction

- but $\{\ \mathbf{u}_i\ \}$ aren't necessarily orthogonal
- we want to find such $\{\ \mathbf{v}_i\ \}$ that $\{\ \mathbf{u}_i\ \}$ are also orthogonal

Objective:

$u_1 = Av_1$
$u_2 = Av_2$
$\vdots$
$u_r = Av_r$

$u_1, \ldots, u_r$ is an orthonormal basis for the **column space**
$u_{r+1}, \ldots, u_m$ is an orthonormal basis for the **left nullspace** $N(A^T)$
$v_1, \ldots, v_r$ is an orthonormal basis for the **row space**
$v_{r+1}, \ldots, v_n$ is an orthonormal basis for the **nullspace** $N(A)$.

Show $A = U\Sigma V^T$

$U = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \quad \Sigma = \begin{bmatrix} \ddots \end{bmatrix} \quad V = \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$

Orthogonal basis of Row Space using Eigen Vectors of A^{T}A

$$\underset{m}{\Big[\quad A \quad\Big]}_{n}$$

$\underset{(n\times m)}{A^T}\underset{(m\times n)}{A} = {A^T A}_{n\times n}$

Let $\{\ \mathbf{v}_i\ \}$ be eigenvectors of A^TA with λ_i being corresponding eigenvalues

$\{\ \mathbf{v}_i\ \}$ are orthogonal

$A^TA\mathbf{v}_i = \lambda_i\mathbf{v}_i$ and $\underbrace{A^TA = V\Lambda V^T}$ (with $\mathbf{v}_i$ being the columns of V)

Orthogonal basis for Column Space of $m \times n$ Matrix A

$(Av_i)^T(Av_j)$

$v_i^T \underbrace{A^T A} v_j$

Inner Product $\langle \overbrace{A\mathbf{v}_i}^{u_i}, \overbrace{A\mathbf{v}_j}^{u_j} \rangle$:

$$(A\mathbf{v}_i)^T(A\mathbf{v}_j) = \mathbf{v}_i^T A^T A\mathbf{v}_j = \mathbf{v}_i^T(\underbrace{A^T A\mathbf{v}_j}_{})$$

$\lambda_j \cdot v_j$

$$= \mathbf{v}_i^T \lambda_j \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j$$

if $i \neq j$, then $\mathbf{v}_i^T \mathbf{v}_j = 0$ $\Longrightarrow$ the image $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ is also orthogonal

Finding the orthonormal $\{ \mathbf{u}_i \}$:

vectors $A\mathbf{v}_i$ are orthogonal, but not orthonormal

$$\|A\mathbf{v}_i\|^2 = \langle A\mathbf{v}_i, A\mathbf{v}_i \rangle = \mathbf{v}_i^T \underbrace{A^T A\mathbf{v}_i}_{} = \mathbf{v}_i^T \underbrace{\lambda_i \mathbf{v}_i}_{} = \lambda_i$$

let $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i$ for $i = 1..r$

if $r < m$, we extend this basis for $\mathbb{R}^m$

SVD Construction for $m \times n$ Matrix A

A is $m \times n$ real matrix

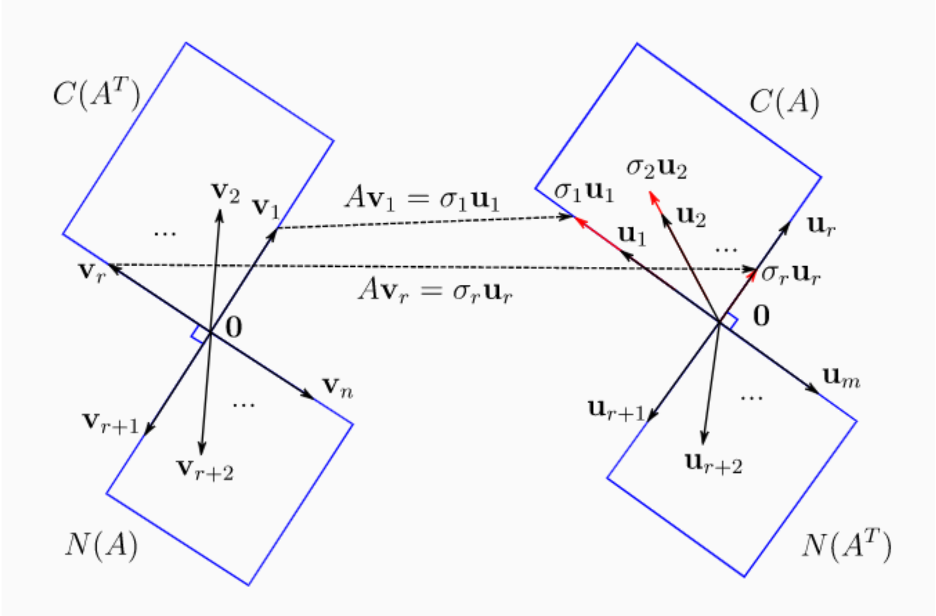
V is obtained from diagonal factorization $A^T A = V \Lambda V^T$

Let $\sigma_i = \sqrt{\lambda_i}$. Then $\mathbf{u}_i = \dfrac{1}{\sigma_i} A\mathbf{v}_i$ or $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$

Put $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ in columns of V and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ in columns of U

so we'll have $AV = U\Sigma$

thus, SVD is $A = U\Sigma V^T$



$\sigma_i \quad A v_i = u_i$

$A v_i = \sigma_i u_i$

$$\underbrace{A \begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{\Sigma}$$

$AV = U\Sigma$

SVD Construction for $m \times n$ Matrix A

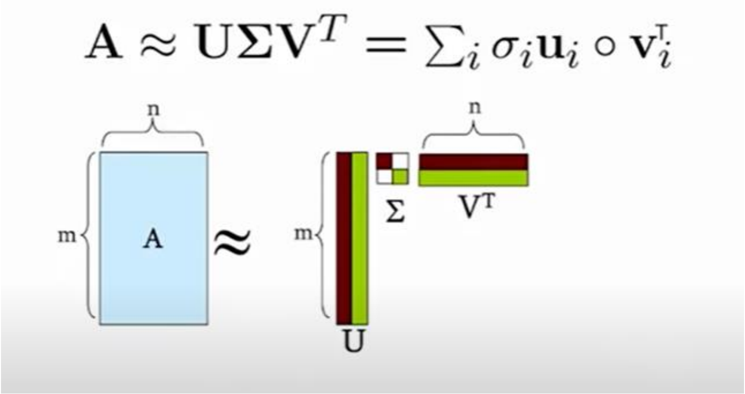
$$A = \left[\begin{array}{cccc|ccc} | & | & & | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ | & | & & | & | & & | \end{array}\right] \left[\begin{array}{cccc|ccc} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ \hline & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{array}\right] \left[\begin{array}{ccc} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_r^T & - \\ \hline - & \mathbf{v}_{r+1}^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{array}\right]$$

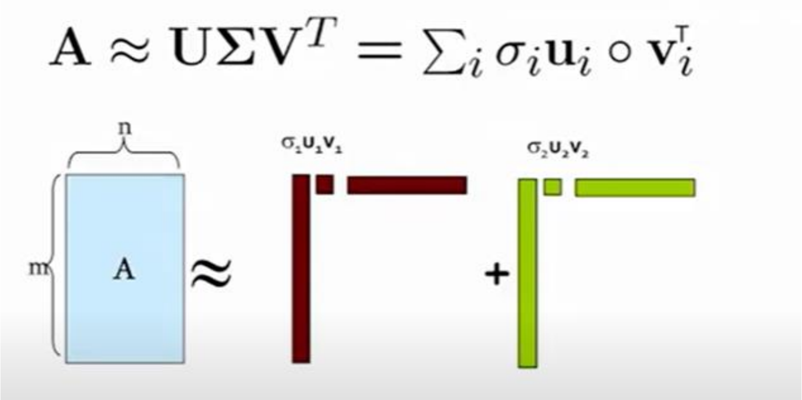
$$A = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_r^T & - \end{bmatrix}$$

$$+$$

$$\begin{bmatrix} | & & | \\ \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} 0 & & \\ & \ddots & \\ & & 0 \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_{r+1}^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}$$

SVD : Sum of Linear Transformations





Linear Transformation of x via Ax can be considered as :

1. Linear transformation using individual matrices $\sigma_i u_i v_i^T$.

2. Summation of linear transformation using rank-1 matrices $\sigma_i u_i v_i^T$.

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$$

$$Ax = (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)x =$$

$$= \underbrace{\sigma_1 \underbrace{u_1 v_1^T} x}_{y_1} + \underbrace{\sigma_2 \underbrace{u_2 v_2^T} x}_{y_2} = y$$