Quadratic forms

a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x) = x^T A x = \sum_{i,j=1}^{n} A_{ij} x_i x_j$$

is called a quadratic form

in a quadratic form we may as well assume $A = A^T$ since

$$x^T A x = x^T ((A + A^T)/2) x$$

$((A + A^T)/2$ is called the symmetric part of $A$)

**uniqueness:** if $x^T A x = x^T B x$ for all $x \in \mathbb{R}^n$ and $A = A^T$, $B = B^T$, then $A = B$
Examples

- $\|Bx\|^2 = x^T B^T Bx$

- $\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$

- $\|Fx\|^2 - \|Gx\|^2$

sets defined by quadratic forms:

- $\{ x \mid f(x) = a \}$ is called a \textit{quadratic surface}

- $\{ x \mid f(x) \leq a \}$ is called a \textit{quadratic region}
Inequalities for quadratic forms

suppose $A = A^T$, $A = QΛQ^T$ with eigenvalues sorted so $λ_1 ≥ · · · ≥ λ_n$

\[
x^T A x = x^T QΛQ^T x = (Q^T x)^T Λ (Q^T x) = \sum_{i=1}^{n} λ_i (q_i^T x)^2 \leq λ_1 \sum_{i=1}^{n} (q_i^T x)^2 = λ_1 \|x\|^2
\]

i.e., we have $x^T A x ≤ λ_1 x^T x$
similar argument shows \( x^T A x \geq \lambda_n \|x\|^2 \), so we have

\[
\lambda_n x^T x \leq x^T A x \leq \lambda_1 x^T x
\]

sometimes \( \lambda_1 \) is called \( \lambda_{\text{max}} \), \( \lambda_n \) is called \( \lambda_{\text{min}} \)

note also that

\[
q_1^T A q_1 = \lambda_1 \|q_1\|^2, \quad q_n^T A q_n = \lambda_n \|q_n\|^2,
\]

so the inequalities are tight
Positive semidefinite and positive definite matrices

suppose $A = A^T \in \mathbb{R}^{n \times n}$

we say $A$ is positive semidefinite if $x^T A x \geq 0$ for all $x$

• denoted $A \geq 0$ (and sometimes $A \succeq 0$)
• $A \geq 0$ if and only if $\lambda_{\min}(A) \geq 0$, i.e., all eigenvalues are nonnegative
• not the same as $A_{ij} \geq 0$ for all $i, j$

we say $A$ is positive definite if $x^T A x > 0$ for all $x \neq 0$

• denoted $A > 0$
• $A > 0$ if and only if $\lambda_{\min}(A) > 0$, i.e., all eigenvalues are positive
Matrix inequalities

• we say $A$ is negative semidefinite if $-A \geq 0$

• we say $A$ is negative definite if $-A > 0$

• otherwise, we say $A$ is indefinite

matrix inequality: if $B = B^T \in \mathbb{R}^n$ we say $A \geq B$ if $A - B \geq 0$, $A < B$ if $B - A > 0$, etc.

for example:

• $A \geq 0$ means $A$ is positive semidefinite

• $A > B$ means $x^T Ax > x^T Bx$ for all $x \neq 0$
many properties that you’d guess hold actually do, \( e.g., \)

- if \( A \geq B \) and \( C \geq D \), then \( A + C \geq B + D \)
- if \( B \leq 0 \) then \( A + B \leq A \)
- if \( A \geq 0 \) and \( \alpha \geq 0 \), then \( \alpha A \geq 0 \)
- if \( A \geq 0 \), then \( A^2 > 0 \)
- if \( A > 0 \), then \( A^{-1} > 0 \)

matrix inequality is only a \textit{partial order}: we can have

\[
A \not\geq B, \quad B \not\geq A
\]

(such matrices are called \textit{incomparable})
Ellipsoids

if $A = A^T > 0$, the set

$$\mathcal{E} = \{ x \mid x^T A x \leq 1 \}$$

is an ellipsoid in $\mathbb{R}^n$, centered at 0
semi-axes are given by $s_i = \lambda_i^{-1/2} q_i$, i.e.:

- eigenvectors determine directions of semi-axes
- eigenvalues determine lengths of semi-axes

note:

- in direction $q_1$, $x^T Ax$ is large, hence ellipsoid is thin in direction $q_1$
- in direction $q_n$, $x^T Ax$ is small, hence ellipsoid is fat in direction $q_n$
- $\sqrt{\lambda_{\text{max}}/\lambda_{\text{min}}}$ gives maximum eccentricity

if $\tilde{\mathcal{E}} = \{ x \mid x^T B x \leq 1 \}$, where $B > 0$, then $\mathcal{E} \subseteq \tilde{\mathcal{E}} \iff A \geq B$
Gain of a matrix in a direction

suppose $A \in \mathbb{R}^{m \times n}$ (not necessarily square or symmetric)

for $x \in \mathbb{R}^n$, $\|Ax\|/\|x\|$ gives the amplification factor or gain of $A$ in the direction $x$

obviously, gain varies with direction of input $x$

questions:

• what is maximum gain of $A$
  (and corresponding maximum gain direction)?

• what is minimum gain of $A$
  (and corresponding minimum gain direction)?

• how does gain of $A$ vary with direction?
Matrix norm

the maximum gain

\[
\max_{x \neq 0} \frac{\|Ax\|}{\|x\|}
\]

is called the *matrix norm* or *spectral norm* of \( A \) and is denoted \( \|A\| \)

\[
\max_{x \neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \max_{x \neq 0} \frac{x^T A^T A x}{\|x\|^2} = \lambda_{\text{max}}(A^T A)
\]

so we have \( \|A\| = \sqrt{\lambda_{\text{max}}(A^T A)} \)

similarly the minimum gain is given by

\[
\min_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sqrt{\lambda_{\text{min}}(A^T A)}
\]
note that

- $A^T A \in \mathbb{R}^{n \times n}$ is symmetric and $A^T A \geq 0$ so $\lambda_{\text{min}}, \lambda_{\text{max}} \geq 0$

- ‘max gain’ input direction is $x = q_1$, eigenvector of $A^T A$ associated with $\lambda_{\text{max}}$

- ‘min gain’ input direction is $x = q_n$, eigenvector of $A^T A$ associated with $\lambda_{\text{min}}$
example: $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$

$A^T A = \begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix}$

$= \begin{bmatrix} 0.620 & 0.785 \\ 0.785 & -0.620 \end{bmatrix} \begin{bmatrix} 90.7 & 0 \\ 0 & 0.265 \end{bmatrix} \begin{bmatrix} 0.620 & 0.785 \\ 0.785 & -0.620 \end{bmatrix}^T$

then $\|A\| = \sqrt{\lambda_{\text{max}}(A^T A)} = 9.53$:

$\left\| \begin{bmatrix} 0.620 \\ 0.785 \end{bmatrix} \right\| = 1, \quad \left\| A \begin{bmatrix} 0.620 \\ 0.785 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 2.18 \\ 4.99 \\ 7.78 \end{bmatrix} \right\| = 9.53$
min gain is $\sqrt{\lambda_{\min}(A^T A)} = 0.514$:

$$\begin{align*}
\| \begin{bmatrix} 0.785 \\ -0.620 \end{bmatrix} \| &= 1, \\
\| A \begin{bmatrix} 0.785 \\ -0.620 \end{bmatrix} \| &= \begin{bmatrix} 0.46 \\ 0.14 \end{bmatrix} \| = 0.514
\end{align*}$$

for all $x \neq 0$, we have

$$0.514 \leq \frac{\|Ax\|}{\|x\|} \leq 9.53$$
Properties of matrix norm

• consistent with vector norm: matrix norm of $a \in \mathbb{R}^{n \times 1}$ is
  \[
  \sqrt{\lambda_{\text{max}}(a^T a)} = \sqrt{a^T a}
  \]

• for any $x$, $\|Ax\| \leq \|A\|\|x\|$

• scaling: $\|aA\| = |a|\|A\|$

• triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$

• definiteness: $\|A\| = 0 \iff A = 0$

• norm of product: $\|AB\| \leq \|A\|\|B\|$
Singular value decomposition

more complete picture of gain properties of $A$ given by *singular value decomposition* (SVD) of $A$:

$$A = U\Sigma V^T$$

where

- $A \in \mathbb{R}^{m \times n}$, $\text{Rank}(A) = r$
- $U \in \mathbb{R}^{m \times r}$, $U^TU = I$
- $V \in \mathbb{R}^{n \times r}$, $V^TV = I$
- $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$, where $\sigma_1 \geq \cdots \geq \sigma_r > 0$
with \( U = [u_1 \cdots u_r] \), \( V = [v_1 \cdots v_r] \),

\[
A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T
\]

- \( \sigma_i \) are the (nonzero) singular values of \( A \)
- \( v_i \) are the right or input singular vectors of \( A \)
- \( u_i \) are the left or output singular vectors of \( A \)
\[ A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^2 V^T \]

hence:

- \( v_i \) are eigenvectors of \( A^T A \) (corresponding to nonzero eigenvalues)
- \( \sigma_i = \sqrt{\lambda_i(A^T A)} \) (and \( \lambda_i(A^T A) = 0 \) for \( i > r \))
- \( \|A\| = \sigma_1 \)
similarly,

\[ AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma^2 U^T \]

hence:

- \( u_i \) are eigenvectors of \( AA^T \) (corresponding to nonzero eigenvalues)
- \( \sigma_i = \sqrt{\lambda_i(AA^T)} \) (and \( \lambda_i(AA^T) = 0 \) for \( i > r \))
- \( u_1, \ldots u_r \) are orthonormal basis for \( \text{range}(A) \)
- \( v_1, \ldots v_r \) are orthonormal basis for \( \mathcal{N}(A)^\perp \)
Interpretations

\[ A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \]

linear mapping \( y = Ax \) can be decomposed as

- compute coefficients of \( x \) along input directions \( v_1, \ldots, v_r \)
- scale coefficients by \( \sigma_i \)
- reconstitute along output directions \( u_1, \ldots, u_r \)

difference with eigenvalue decomposition for symmetric \( A \): input and output directions are *different*
• $v_1$ is most sensitive (highest gain) input direction

• $u_1$ is highest gain output direction

• $Av_1 = \sigma_1 u_1$
SVD gives clearer picture of gain as function of input/output directions

**example:** consider $A \in \mathbb{R}^{4 \times 4}$ with $\Sigma = \text{diag}(10, 7, 0.1, 0.05)$

- input components along directions $v_1$ and $v_2$ are amplified (by about 10) and come out mostly along plane spanned by $u_1, u_2$

- input components along directions $v_3$ and $v_4$ are attenuated (by about 10)

- $\|Ax\|/\|x\|$ can range between 10 and 0.05

- $A$ is nonsingular

- for some applications you might say $A$ is *effectively* rank 2
Lecture 16
SVD Applications

- general pseudo-inverse
- full SVD
- image of unit ball under linear transformation
- SVD in estimation/inversion
- sensitivity of linear equations to data error
- low rank approximation via SVD
Min Squared Error: Over-Constrained

- Given \( y \in \mathbb{R}^q \) and \( A \in \mathbb{R}^{q \times n} \) so that \( q > n \) (\( A \) is slim) and \( \text{rank}(A) = n \) we’d like to find \( x \in \mathbb{R}^n \) such that \( Ax \approx y \) in the minimum \( l_2 \) sense:

  \[
  \arg \min_x \| y - Ax \|^2
  \]

  where \( \|v\|^2 = \sum_i v_i^2 \)

- If \( A \) were invertible we would simply take \( x = A^{-1}y \)

- This is a quadratic expression in \( x \) so it has a single minimum where its gradient is 0.

  \[
  J = \| y - Ax \|^2 = (y - Ax)'(y - Ax) = y'y - 2y'Ax + x'A'Ax
  \]

  \[
  \frac{\partial J}{\partial x} = -2y'A + 2x'A'A = 0
  \]

  \[
  A'y = A'Ax
  \]

- \( (A'A)^{-1} \) exists (rank = \( n \)) so \( x = (A'A)^{-1}A'y \)

- Plug in SVD and get \( x = V\Sigma^{-2}V' \Sigma U'y = V\Sigma^{-1}U'y \)

- \( V\Sigma^{-1}U' \) is denoted by \( A^\dagger \) and is called \( A \)'s pseudo inverse since \( A^\dagger A = I \)
Under-Constrained

- Given $y \in \mathbb{R}^q$ and $A \in \mathbb{R}^{q \times n}$ so that $q < n$ (A is fat) and $\text{rank}(A) = n$ we’d like to find $x \in \mathbb{R}^n$ such that $Ax = y$ (easy).
Of all possible $x$s we want the smallest $x$, i.e.

$$\arg \min_x ||x||^2 \quad \text{s.t.} \quad Ax = y$$

- This is a constrained optimization problem, so we solve with Lagrange multipliers

$$J = x'x + \lambda'(Ax - y)$$
$$\frac{\partial J}{\partial x} = 2x' + \lambda'A = 0$$
$$x = \frac{1}{2}A'\lambda$$

- Plug into constraint $Ax = y$

$$A\left(\frac{1}{2}A'\lambda\right) = y$$

- $(AA)^{-1}$ exists, so

$$\lambda = 2(AA')^{-1}y \quad x = A'(AA')^{-1}y = A^+y$$

where, as before $A^+ = V\Sigma^{-1}U'$
Optimal Control*

*Of noiseless, open loop, discrete time, LTI system

- Given system \( x_{n+1} = Ax_n + Bu_n \) with \( x_0 = 0 \) bring the system to specified \( x_n \) (with a minimum energy control signal)

- We can expand the recursive definition and get
  \[
  x_n = \sum_{i=0}^{n-1} A^i Bu_i
  \]

- or, in matrix form
  \[
  x_n = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_{n-1} \end{bmatrix}
  \]

- This is an under constrained problem. If \( \tilde{A} \) is of rank \( n \) (i.e. system is controllable) then there are infinite possible solutions for \( \tilde{u} \)

- but there is only one solution that minimizes \( ||\tilde{u}||^2 \):
  \[
  \tilde{u} = \tilde{A}'(\tilde{A} \tilde{A}')^{-1}x_n
  \]
• Plugging in definition of $\tilde{A}$ to $\tilde{A}^* = \tilde{A}' (\tilde{A} \tilde{A}')^{-1}$ we see that

$$u_i = B'A^i \left( \sum_{j=0}^{n-1} A^iBB'(A^i)' \right)^{-1} x_n$$

$$\underline{w_c^{-1}(n-1)}$$

• The minimum energy (smallest $||\tilde{u}||^2$) control signal is the same signal used in the proof that the system is controllable iff the grammian is invertible (How did we assure that the grammian is invertible here?)
General pseudo-inverse

If \( A \) has SVD \( A = U \Sigma V^T \),

\[
A^\dagger = V \Sigma^{-1} U^T
\]

is the \textit{pseudo-inverse} or \textit{Moore-Penrose inverse} of \( A \)

If \( A \) is skinny and full rank,

\[
A^\dagger = (A^T A)^{-1} A^T
\]

gives the least-squares solution \( x_{ls} = A^\dagger y \)

If \( A \) is fat and full rank,

\[
A^\dagger = A^T (A A^T)^{-1}
\]

gives the least-norm solution \( x_{ln} = A^\dagger y \)
in general case:

\[ X_{ls} = \{ z \mid \|Az - y\| = \min_w \|Aw - y\| \} \]

is set of least-squares solutions

\[ x_{pinv} = A^\dagger y \in X_{ls} \]

has minimum norm on \( X_{ls} \), i.e., \( x_{pinv} \) is the minimum-norm, least-squares solution
Pseudo-inverse via regularization

for $\mu > 0$, let $x_\mu$ be (unique) minimizer of

$$\|Ax - y\|^2 + \mu\|x\|^2$$

i.e.,

$$x_\mu = \left( A^T A + \mu I \right)^{-1} A^T y$$

here, $A^T A + \mu I > 0$ and so is invertible

then we have $\lim_{\mu \to 0} x_\mu = A^\dagger y$

in fact, we have $\lim_{\mu \to 0} \left( A^T A + \mu I \right)^{-1} A^T = A^\dagger$

(check this!)
Full SVD

SVD of $A \in \mathbb{R}^{m \times n}$ with $\text{Rank}(A) = r$:

$$A = U_1 \Sigma_1 V_1^T = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

- find $U_2 \in \mathbb{R}^{m \times (m-r)}$, $V_2 \in \mathbb{R}^{n \times (n-r)}$ s.t. $U = [U_1 \ U_2] \in \mathbb{R}^{m \times m}$ and $V = [V_1 \ V_2] \in \mathbb{R}^{n \times n}$ are orthogonal

- add zero rows/cols to $\Sigma_1$ to form $\Sigma \in \mathbb{R}^{m \times n}$:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0_{r \times (n - r)} \\ 0_{(m - r) \times r} & 0_{(m - r) \times (n - r)} \end{bmatrix}$$
then we have

\[ A = U_1 \Sigma_1 V_1^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \]

\[ i.e.: \]

\[ A = U \Sigma V^T \]

called full SVD of \( A \)

(SVD with positive singular values only called compact SVD)
full SVD:

\[ A = U \Sigma V^T \]

gives interpretation of \( y = Ax \):

- rotate (by \( V^T \))
- stretch along axes by \( \sigma_i \) (\( \sigma_i = 0 \) for \( i > r \))
- zero-pad (if \( m > n \)) or truncate (if \( m < n \)) to get \( m \)-vector
- rotate (by \( U \))
Image of unit ball under $A$

$$\{Ax \mid \|x\| \leq 1\}$$ is ellipsoid with principal axes $\sigma_i u_i$.
SVD in estimation/inversion

suppose $y = Ax + v$, where

- $y \in \mathbb{R}^m$ is measurement
- $x \in \mathbb{R}^n$ is vector to be estimated
- $v$ is a measurement noise or error

‘norm-bound’ model of noise: we assume $\|v\| \leq \alpha$ but otherwise know nothing about $v$ ($\alpha$ gives max norm of noise)
• consider estimator \( \hat{x} = By \), with \( BA = I \) \((i.e.,\) unbiased\)

• estimation or inversion error is \( \tilde{x} = \hat{x} - x = Bv \)

• set of possible estimation errors is ellipsoid

\[
\tilde{x} \in \mathcal{E}_{\text{unc}} = \{ Bv \mid \|v\| \leq \alpha \}
\]

• \( x = \hat{x} - \tilde{x} \in \hat{x} - \mathcal{E}_{\text{unc}} = \hat{x} + \mathcal{E}_{\text{unc}}, \ i.e.: \)

true \( x \) lies in \textit{uncertainty ellipsoid} \( \mathcal{E}_{\text{unc}} \), centered at estimate \( \hat{x} \)

• ‘good’ estimator has ‘small’ \( \mathcal{E}_{\text{unc}} \) \((\text{with} \ BA = I, \ \text{of course})\)
semiaxes of $E_{\text{unc}}$ are $\alpha \sigma_i u_i$ (singular values & vectors of $B$)

e.g., maximum norm of error is $\alpha \|B\|$, i.e., $\|\hat{x} - x\| \leq \alpha \|B\|

**optimality of least-squares:** suppose $BA = I$ is any estimator, and $B_{ls} = A^\dagger$ is the least-squares estimator

then:

- $B_{ls}B_{ls}^T \leq BB^T$
- $\mathcal{E}_{ls} \subseteq \mathcal{E}$
- in particular $\|B_{ls}\| \leq \|B\|$

*i.e.*, the least-squares estimator gives the *smallest* uncertainty ellipsoid
Proof of optimality property

suppose $A \in \mathbb{R}^{m \times n}$, $m > n$, is full rank

SVD: $A = U\Sigma V^T$, with $V$ orthogonal

$B_{ls} = A^\dagger = V\Sigma^{-1}U^T$, and $B$ satisfies $BA = I$

define $Z = B - B_{ls}$, so $B = B_{ls} + Z$

then $ZA = ZU\Sigma V^T = 0$, so $ZU = 0$ (multiply by $V\Sigma^{-1}$ on right)

therefore

$$BB^T = (B_{ls} + Z)(B_{ls} + Z)^T$$
$$= B_{ls}B_{ls}^T + B_{ls}Z^T + ZB_{ls}^T + ZZ^T$$
$$= B_{ls}B_{ls}^T + ZZ^T$$
$$\geq B_{ls}B_{ls}^T$$

using $ZB_{ls}^T = (ZU)\Sigma^{-1}V^T = 0$
Sensitivity of linear equations to data error

consider $y = Ax$, $A \in \mathbb{R}^{n \times n}$ invertible; of course $x = A^{-1}y$

suppose we have an error or noise in $y$, *i.e.*, $y$ becomes $y + \delta y$

then $x$ becomes $x + \delta x$ with $\delta x = A^{-1}\delta y$

hence we have $\|\delta x\| = \|A^{-1}\delta y\| \leq \|A^{-1}\|\|\delta y\|$

if $\|A^{-1}\|$ is large,

- small errors in $y$ can lead to large errors in $x$
- can’t solve for $x$ given $y$ (with small errors)
- hence, $A$ can be considered singular in practice
a more refined analysis uses *relative* instead of *absolute* errors in $x$ and $y$

since $y = Ax$, we also have $\|y\| \leq \|A\|\|x\|$, hence

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\|\|A^{-1}\| \frac{\|\delta y\|}{\|y\|}$$

$$\kappa(A) = \|A\|\|A^{-1}\| = \sigma_{\text{max}}(A)/\sigma_{\text{min}}(A)$$

is called the *condition number* of $A$

we have:

relative error in solution $x \leq \text{condition number} \cdot \text{relative error in data } y$

or, in terms of # bits of guaranteed accuracy:

$$\# \text{ bits accuracy in solution} \approx \# \text{ bits accuracy in data} - \log_2 \kappa$$
we say

- $A$ is well conditioned if $\kappa$ is small
- $A$ is poorly conditioned if $\kappa$ is large

(definition of ‘small’ and ‘large’ depend on application)

same analysis holds for least-squares solutions with $A$ nonsquare,

$$\kappa = \frac{\sigma_{\text{max}}(A)}{\sigma_{\text{min}}(A)}$$
another interpretation of $\sigma_i$:

$$\sigma_i = \min\{ \|A - B\| \mid \text{Rank}(B) \leq i - 1 \}$$

i.e., the distance (measured by matrix norm) to the nearest rank $i - 1$ matrix

for example, if $A \in \mathbb{R}^{n \times n}$, $\sigma_n = \sigma_{\min}$ is distance to nearest singular matrix

hence, small $\sigma_{\min}$ means $A$ is near to a singular matrix
**application:** model simplification

suppose \( y = Ax + v \), where

- \( A \in \mathbb{R}^{100 \times 30} \) has SVs
  
  \[10, 7, 2, 0.5, 0.01, \ldots, 0.0001\]

- \( \|x\| \) is on the order of 1

- unknown error or noise \( v \) has norm on the order of 0.1

then the terms \( \sigma_i u_i v_i^T x \), for \( i = 5, \ldots, 30 \), are substantially smaller than the noise term \( v \)

simplified model:

\[
y = \sum_{i=1}^{4} \sigma_i u_i v_i^T x + v
\]