Machine Learning

CMPT 726

Mo Chen
SFU School of Computing Science
2021-09-13

Linear Algebra and Calculus Review (cont’d)

p-Norms

Also known as ℓ_p norms.

These are norms of **vectors**. In general, the p-norm of a vector $\vec{x}$ is

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

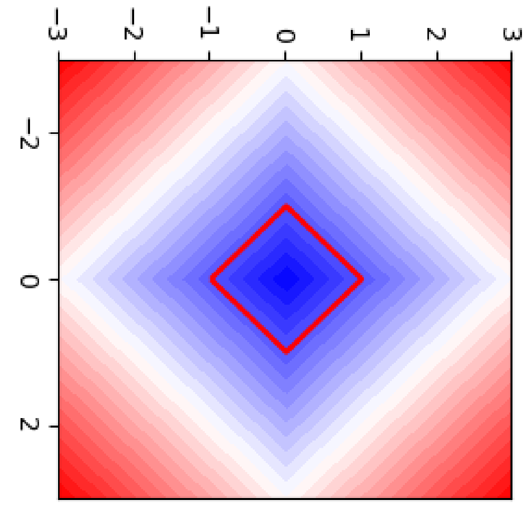
Common norms:

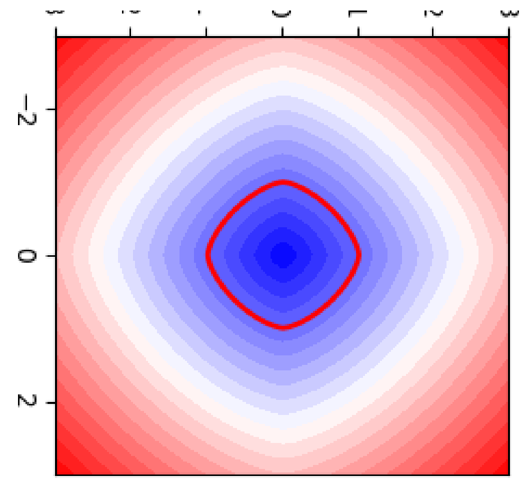
ℓ_1 norm ("Manhattan norm"): $\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$

ℓ_2 norm ("Euclidean norm"): $\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

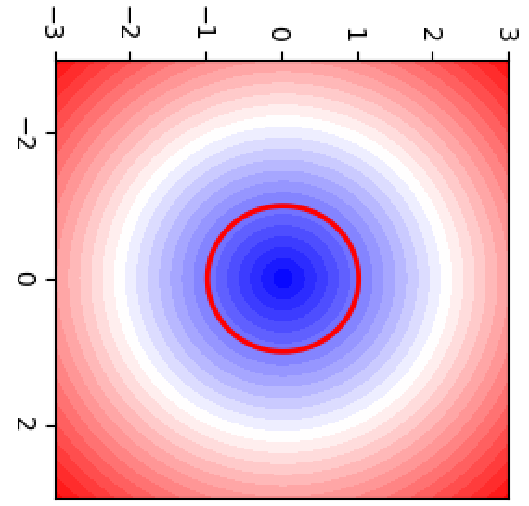
ℓ_∞ norm ("Max norm"): $\|\vec{x}\|_\infty = \max\{|x_1|, \dots, |x_n|\}$

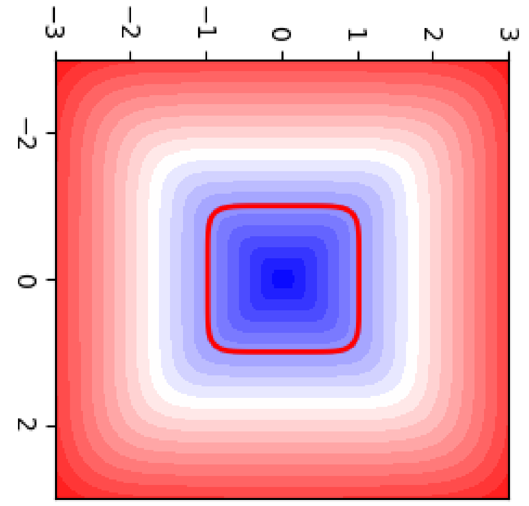
$p = 1$



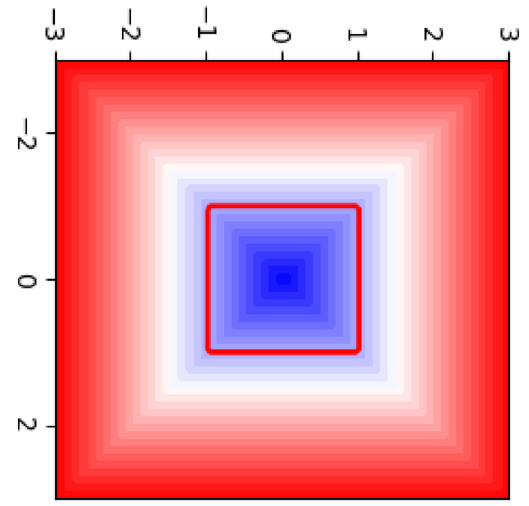


$p = 2$





$p = \infty$



$\|\vec{x}\|_1 \geq \|\vec{x}\|_2 \geq \|\vec{x}\|_\infty$

Matrix Norms

Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

Induced/operator norms:

$$\|A\|_p = \sup_{\|\vec{x}\|_p = 1} \left\{ \|A\vec{x}\|_p \right\}$$

Special case ($p = 2$): known as “spectral norm”:

$$\|A\|_2 = \sup_{\|\vec{x}\|_2 = 1} \{ \|A\vec{x}\|_2 \} = \sigma_{1,1}(A)$$

- $\sigma_{1,1}(A)$ denotes the largest singular value of A

Positive/Negative (Semi-)Definite Matrices

- A symmetric matrix A is
  - positive definite if all of its eigenvalues are positive
  - negative definite if all of its eigenvalues are negative
  - positive semi-definite if all of its eigenvalues are non-negative (≥ 0)
  - negative semi-definite if all of its eigenvalues are non-positive (≤ 0)
  - indefinite if some of its eigenvalues are positive and others are negative

Positive/Negative (Semi-)Definite Matrices

- A symmetric matrix A is
  - positive definite if all of its eigenvalues are positive $A \succ 0$
  - negative definite if all of its eigenvalues are negative $A \prec 0$
  - positive semi-definite if all of its eigenvalues are non-negative $A \succcurlyeq 0$
  - negative semi-definite if all of its eigenvalues are non-positive $A \preccurlyeq 0$
  - indefinite if some of its eigenvalues are positive and others are negative

Taylor Expansion

Polynomial: $g(x) = \sum_{i=1}^{d} \alpha_i x^i$, where d, the highest power, is known as the **degree**

How to approximate an arbitrary function $f: \mathbb{R} \to \mathbb{R}$ with a polynomial g?

We can try to match the function value at a certain point, the first derivative, the second derivative, etc.

$$f(x_0) = g(x_0)$$
$$f'(x_0) = g'(x_0)$$
$$f''(x_0) = g''(x_0)$$
$$\vdots$$

A polynomial g that satisfies these conditions is known as a **Taylor polynomial**

Taylor Expansion

Consider a function $f\colon \mathbb{R} \to \mathbb{R}$ and its approximations with Taylor polynomials of various degrees.

The 0th order Taylor expansion at x_0 is:

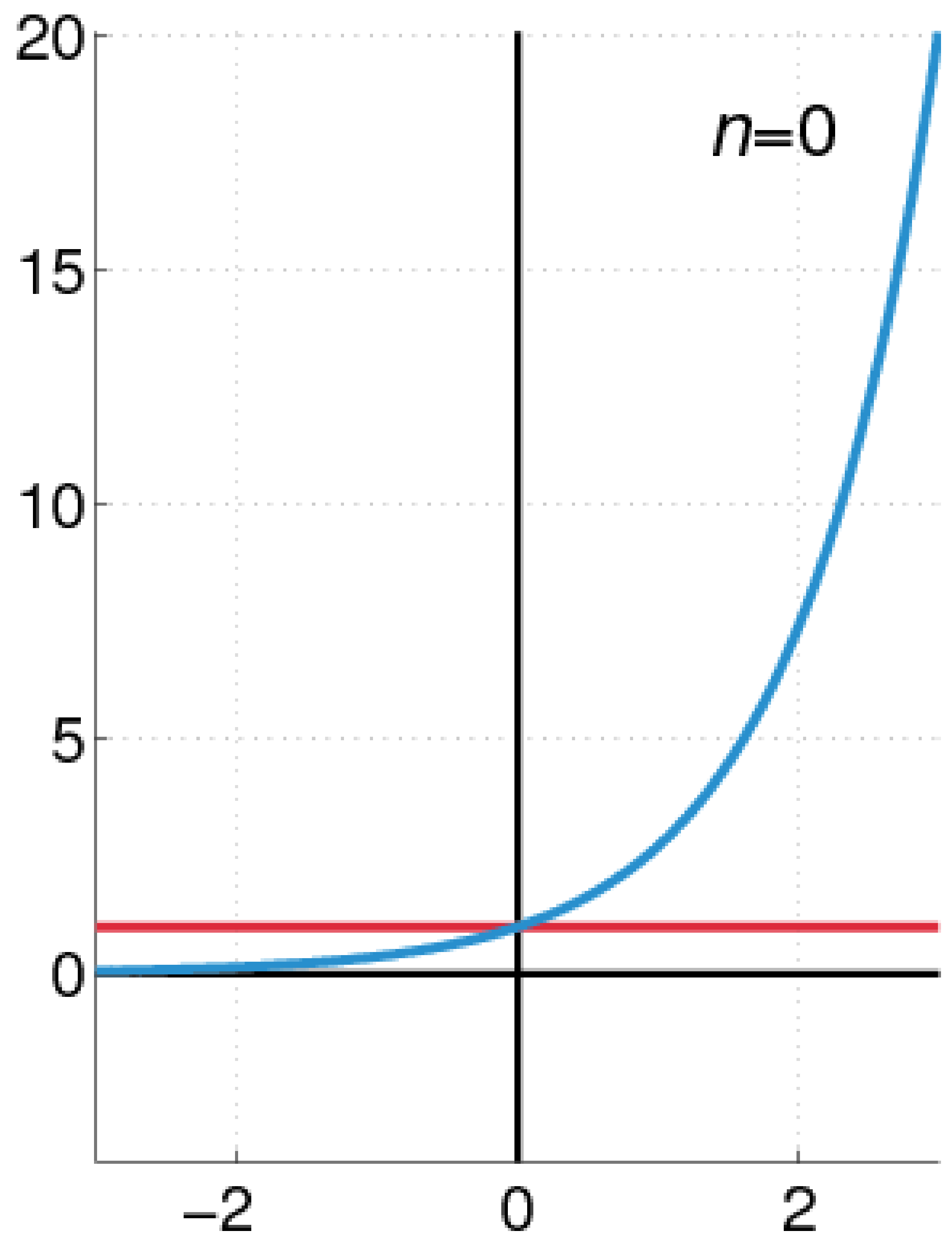
$$g(x) = f(x_0)$$

The 1st order Taylor expansion at x_0 is:

$$g(x) = f(x_0) + \frac{1}{1!}(x - x_0)f'(x_0)$$

The 2nd order Taylor expansion at x_0 is:

$$g(x) = f(x_0) + \frac{1}{1!}(x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0)$$



Taylor Expansion

Polynomials in multiple variables:

$$g(x_1, x_2) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \gamma_{11} x_1^2 + 2\gamma_{12} x_1 x_2 + \gamma_{22} x_2^2$$ (degree 2 polynomial)

In matrix notation:

Let $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$g(\vec{x}) = \alpha + \vec{x}^\top \vec{\beta} + \vec{x}^\top \Gamma \vec{x}$, where $\vec{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{pmatrix}$

Note that Γ is symmetric

Taylor Expansion

Consider a function $f: \mathbb{R}^n \to \mathbb{R}$.

The 0th order Taylor expansion at $\vec{x}_0$ is

$$f(\vec{x}_0)$$

The 1st order Taylor expansion at $\vec{x}_0$ is

$$f(\vec{x}_0) + (\vec{x} - \vec{x}_0)^\top \frac{\partial f}{\partial \vec{x}}(\vec{x}_0)$$

The 2nd order Taylor expansion at $\vec{x}_0$ is

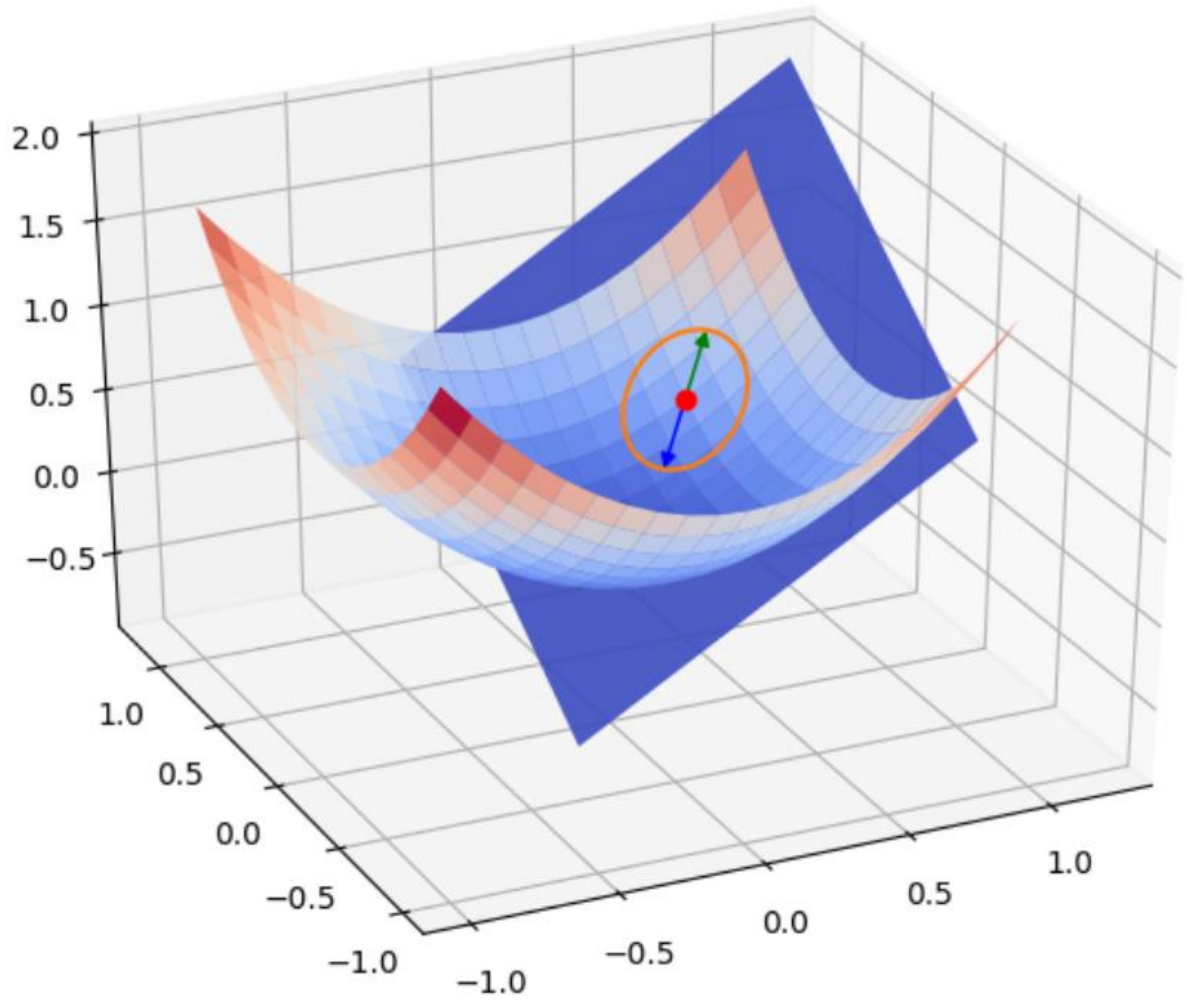
$$f(\vec{x}_0) + (\vec{x} - \vec{x}_0)^\top \frac{\partial f}{\partial \vec{x}}\left(\vec{x}_0\right) + \frac{1}{2}(\vec{x} - \vec{x}_0)^\top \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top}(\vec{x} - \vec{x}_0)$$

Taylor Expansion

The 2nd order Taylor expansion at $\vec{x}_0$ is:

$$f(\vec{x}_0) + (\vec{x} - \vec{x}_0)^\top \frac{\partial f}{\partial \vec{x}}(\vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^\top \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^\top}(\vec{x} - \vec{x}_0)$$

$$\frac{\partial f}{\partial \vec{x}}(\vec{x}_0) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}_0) \\ \frac{\partial f}{\partial x_2}(\vec{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}_0) \end{pmatrix}$$



Gradient, Direction of steepest ascent

Taylor Expansion

The 2nd order Taylor expansion at $\vec{x}_0$ is:

$$f(\vec{x}_0) + (\vec{x} - \vec{x}_0)^{\mathsf{T}} \frac{\partial f}{\partial \vec{x}}(\vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^{\mathsf{T}} \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^{\mathsf{T}}}(\vec{x} - \vec{x}_0)$$

$$\frac{\partial f}{\partial \vec{x}}(\vec{x}_0) \coloneqq \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}_0) \\ \frac{\partial f}{\partial x_2}(\vec{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}_0) \end{pmatrix}, \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^{\mathsf{T}}} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\vec{x}_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\vec{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{x}_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\vec{x}_0) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\vec{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\vec{x}_0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\vec{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\vec{x}_0) \end{pmatrix}$$

Gradient, Direction of steepest ascent, and Hessian

Order of differentiation doesn’t matter, so the Hessian is symmetric.

Quadratic Forms

A function $g(\vec{x}) = \vec{x}^\top A\vec{x}$ is known as a quadratic form.

Alternative definition of positive/negative (semi-)definiteness of A:

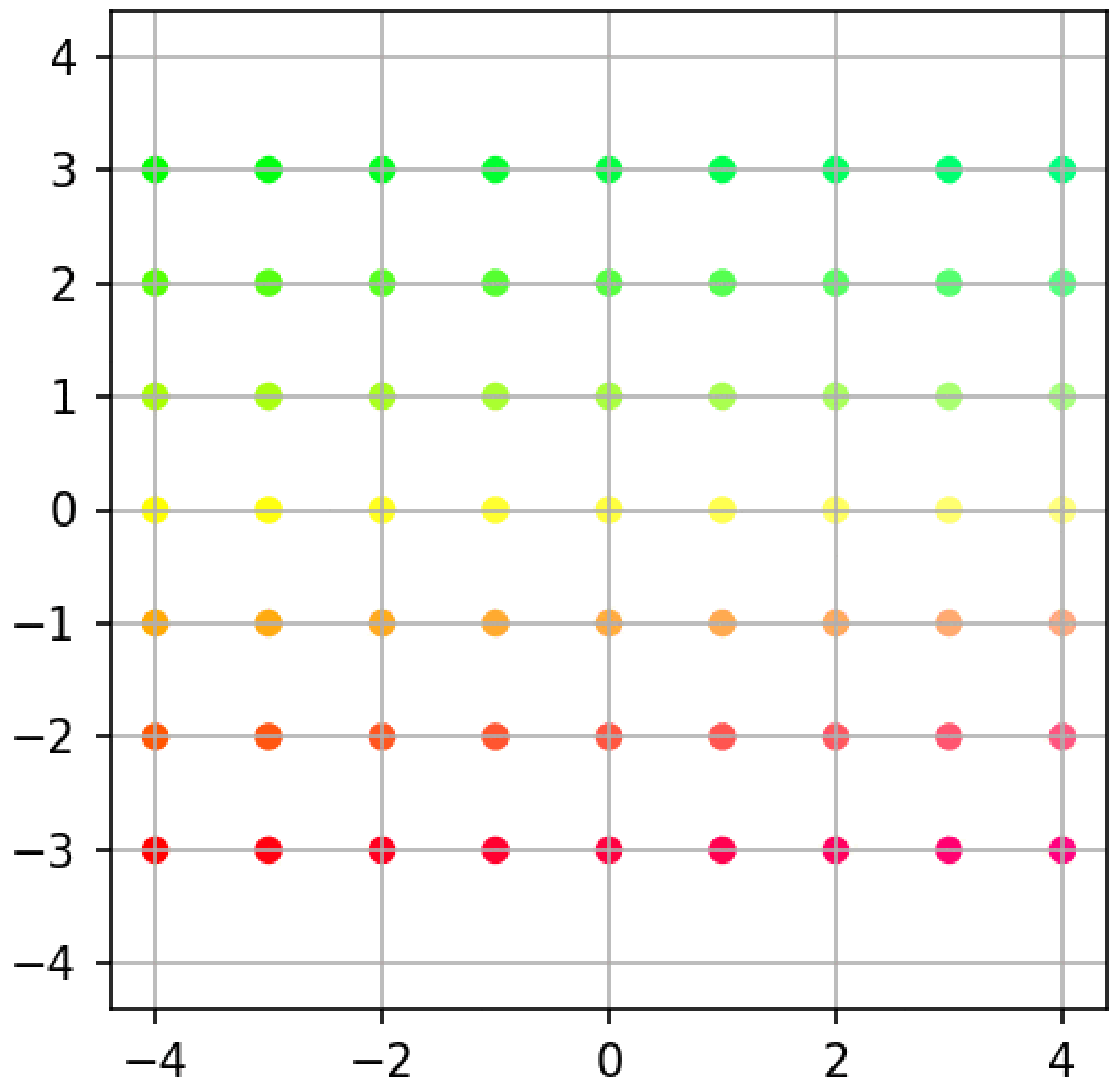
- Positive definite: $\vec{x}^\top A\vec{x} > 0 \ \forall \vec{x} \neq \vec{0}$
- Negative definite: $\vec{x}^\top A\vec{x} < 0 \ \forall \vec{x} \neq \vec{0}$
- Positive semi-definite: $\vec{x}^\top A\vec{x} \geq 0 \ \forall \vec{x}$
- Negative semi-definite: $\vec{x}^\top A\vec{x} \leq 0 \ \forall \vec{x}$
- Indefinite: $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} > 0$ and $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} < 0$

Quadratic Forms

Let's check if the two definitions agree.

- Indefinite: $\exists\vec{x}$ such that $\vec{x}^\top A\vec{x} > 0$ and $\exists\vec{x}$ such that $\vec{x}^\top A\vec{x} < 0$

$$A = \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} = I \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} I^\top$$



0.9 0.0

0.0 -0.5

Eigenvalues are 0.9 and -0.5, according to earlier definition, matrix is indefinite.

$$\vec{x}^\top A\vec{x} = \langle\vec{x}, A\vec{x}\rangle$$

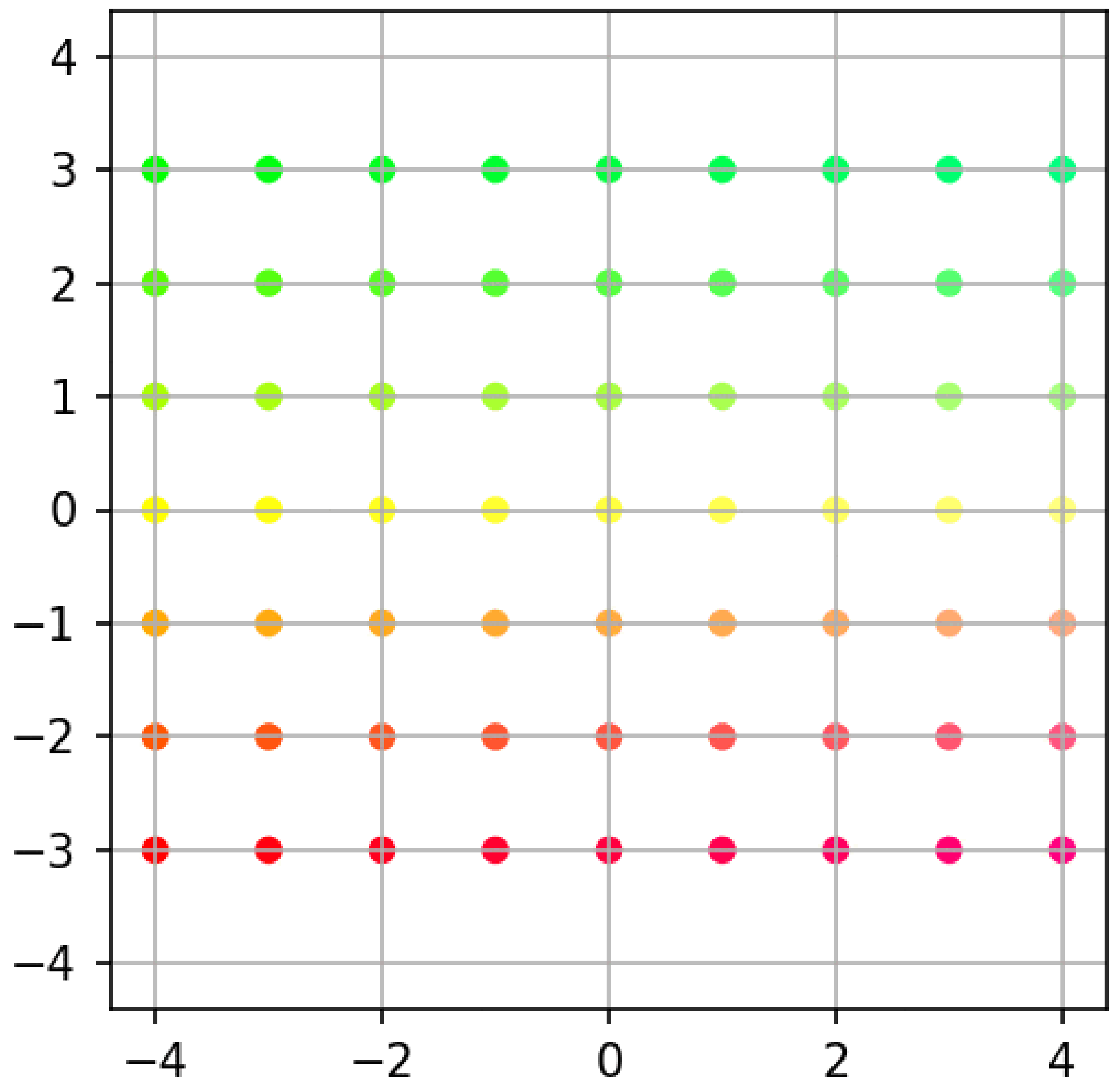
Recall: Positive if original and transformed vectors are less than 90 degrees apart, and negative otherwise.

Quadratic Forms

Let's check if the two definitions agree.

- Indefinite: $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} > 0$ and $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} < 0$

$$A = \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} = I \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} I^\top$$



Eigenvalues are 0.9 and -0.5, according to earlier definition, matrix is indefinite.

$$\vec{x}^\top A\vec{x} = \langle \vec{x}, A\vec{x} \rangle$$

Recall: Positive if original and transformed vectors are less than 90 degrees apart, and negative otherwise. Consider the two eigenvectors

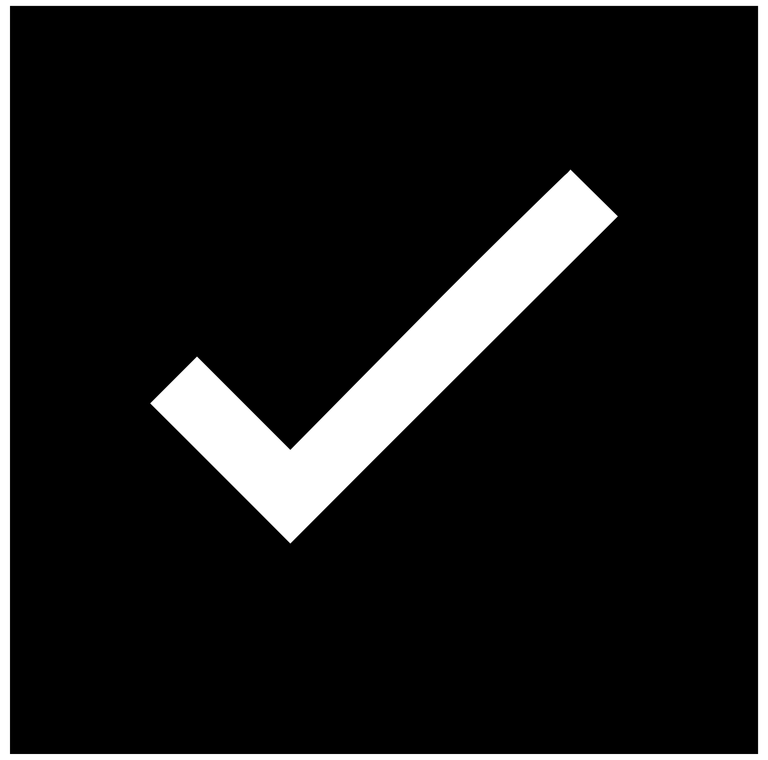
Quadratic Forms

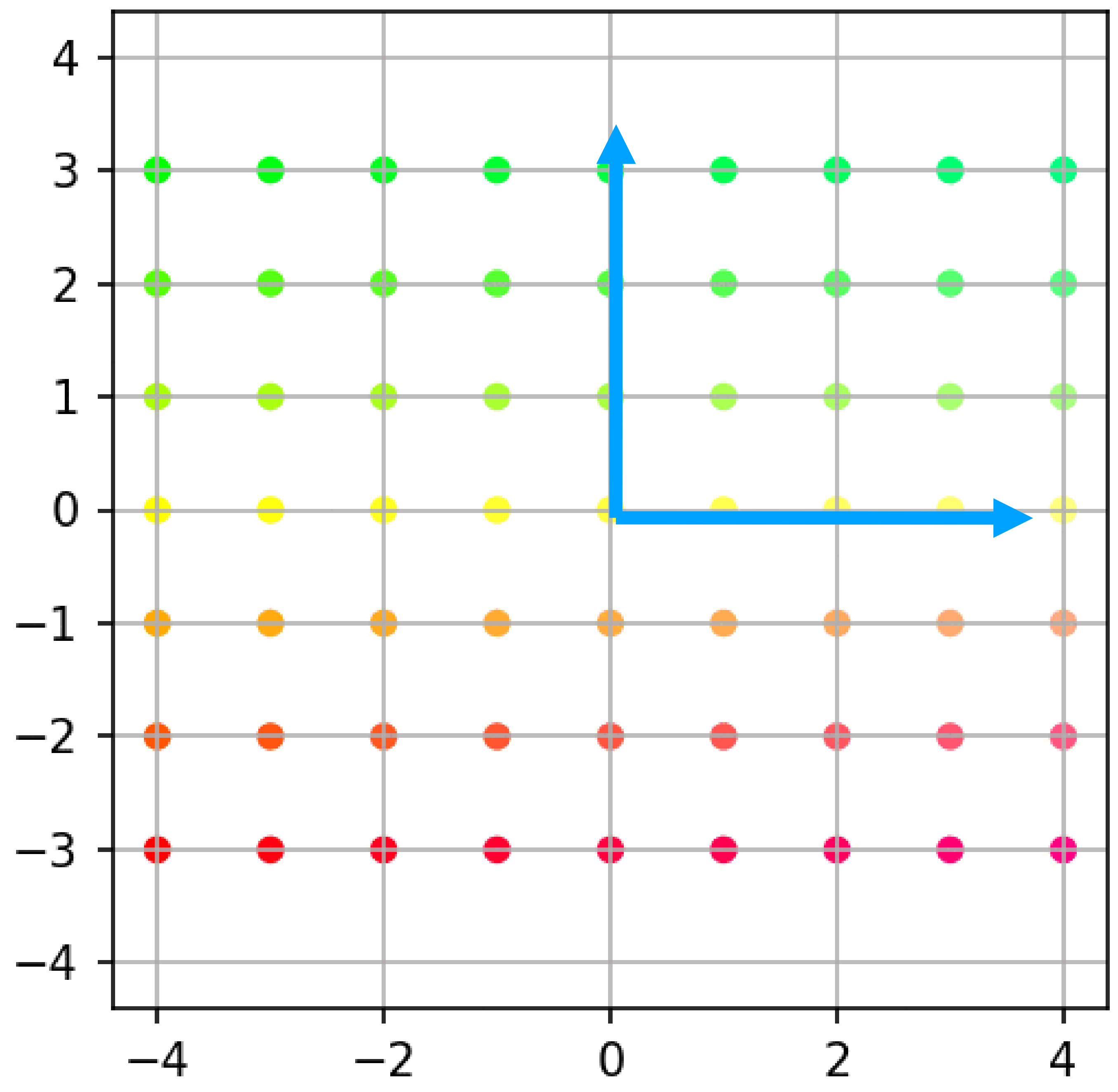
Let’s check if the two definitions agree.

- Indefinite: $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} > 0$ and $\exists \vec{x}$ such that $\vec{x}^\top A\vec{x} < 0$

$$\vec{e}_1^\top A\vec{e}_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} = 0.9$$

$$\vec{e}_2^\top A\vec{e}_2 = \begin{pmatrix} 0.9 & 0 \\ 0 & -0.5 \end{pmatrix} = -0.5$$





0.9	0.0
0.0	-0.5

Quadratic Forms

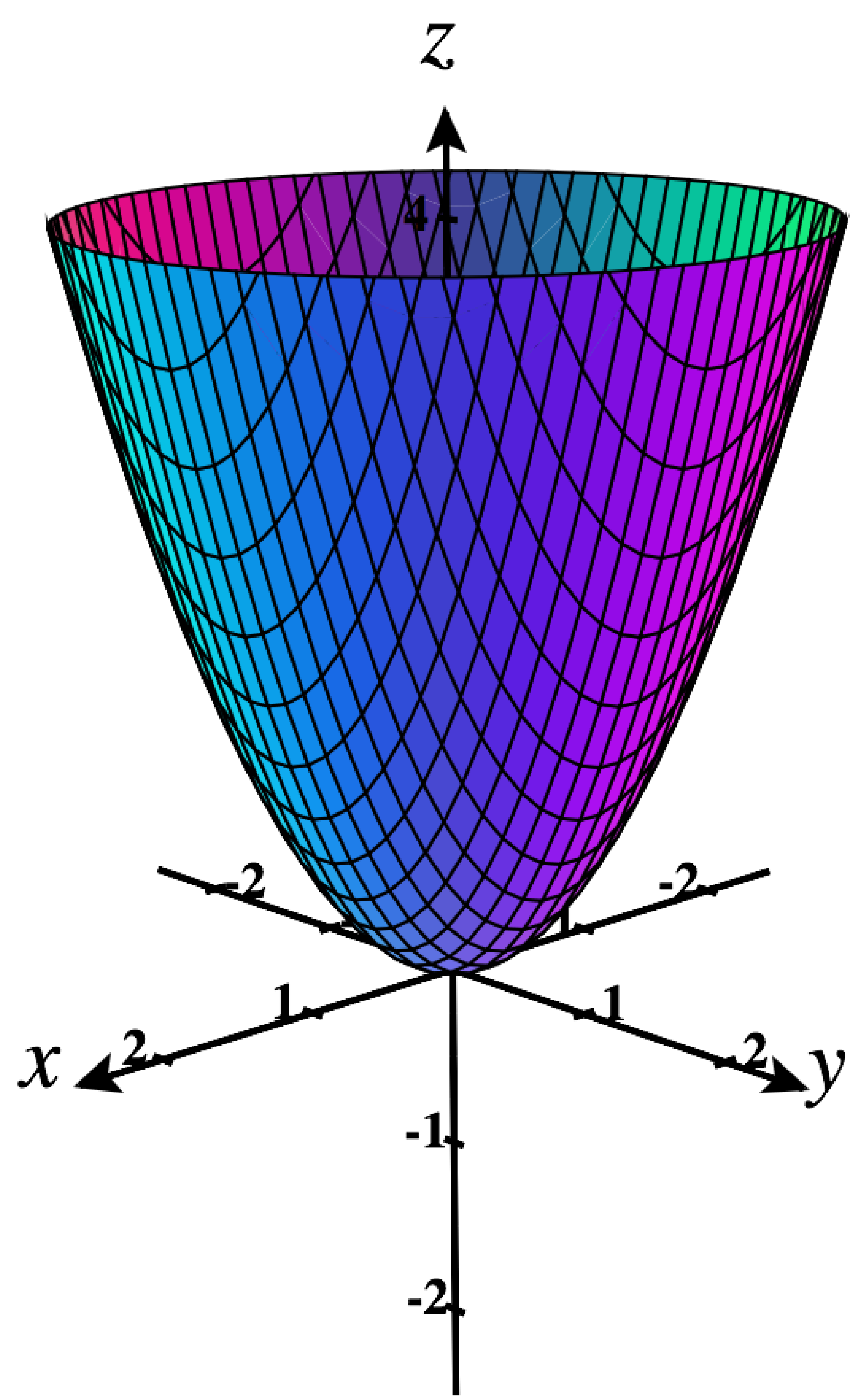
What does $\vec{x}^\top A\vec{x}$ look like when:

- A is positive definite?
- A is negative definite?
- A is positive semi-definite?
- A is negative semi-definite?
- A is indefinite?

Quadratic Forms

What does $\vec{x}^\top A\vec{x}$ look like when:

- A **is positive definite?**
- A is negative definite?
- A is positive semi-definite?
- A is negative semi-definite?
- A is indefinite?



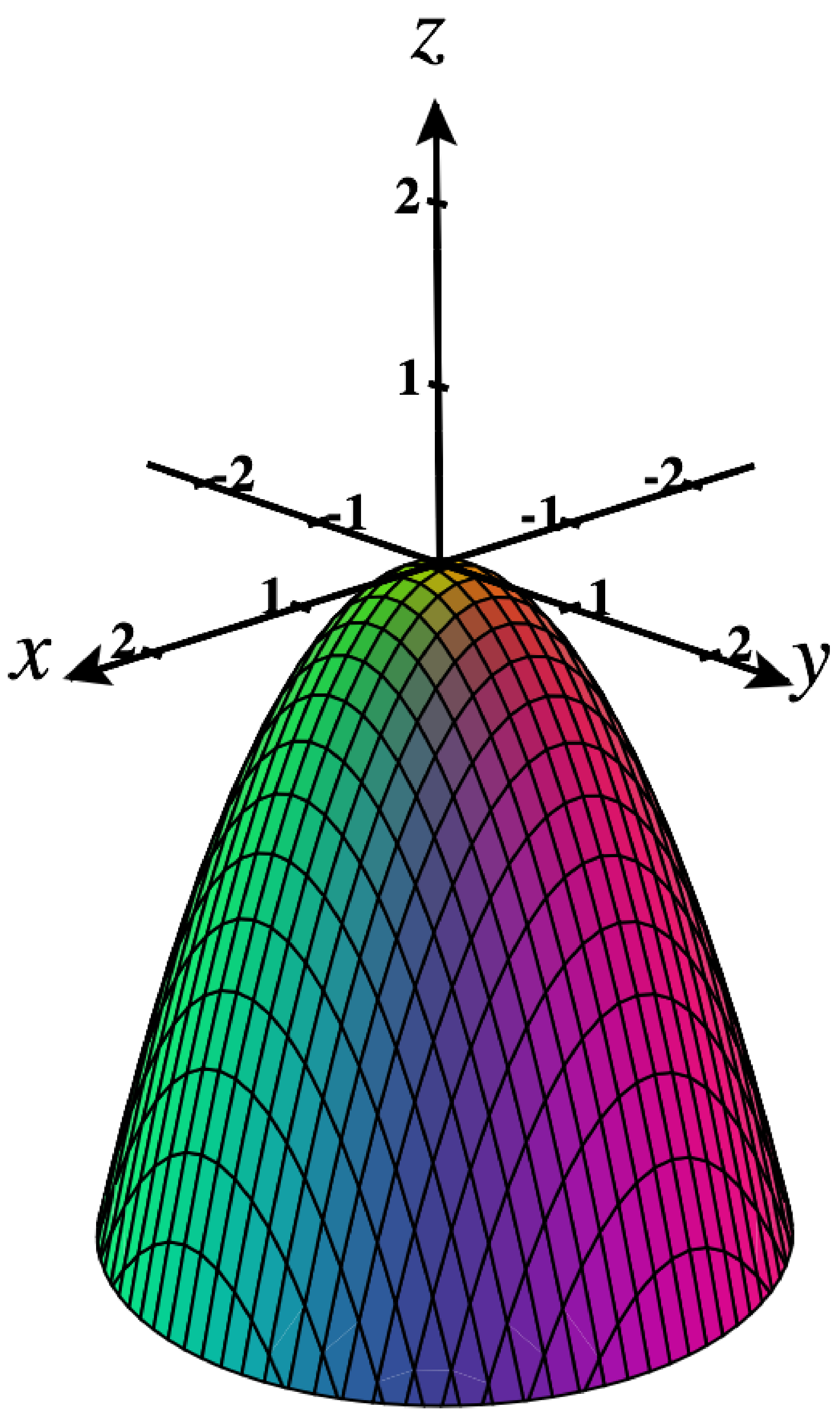
$$\vec{x}^\top A\vec{x} > 0 \;\forall \vec{x} \neq \vec{0}$$

Quadratic Forms

What does $\vec{x}^\top A\vec{x}$ look like when:

- A is positive definite?
- A **is negative definite?**
- A is positive semi-definite?
- A is negative semi-definite?
- A is indefinite?

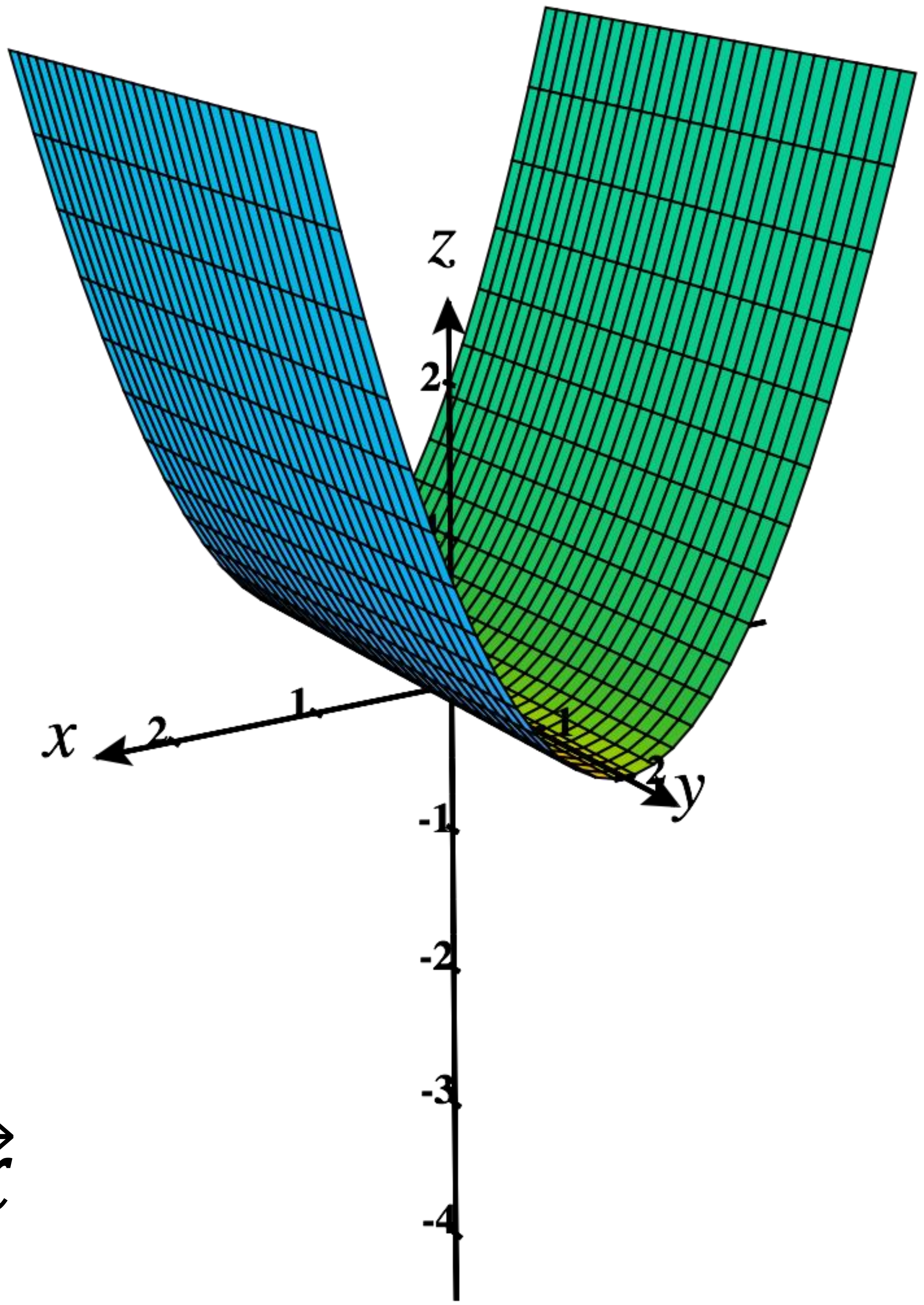
$$\vec{x}^\top A\vec{x} < 0 \;\forall \vec{x} \neq \vec{0}$$



Quadratic Forms

What does $\vec{x}^\top A\vec{x}$ look like when:

- A is positive definite?
- A is negative definite?
- A **is positive semi-definite?**
- A is negative semi-definite?
- A is indefinite?



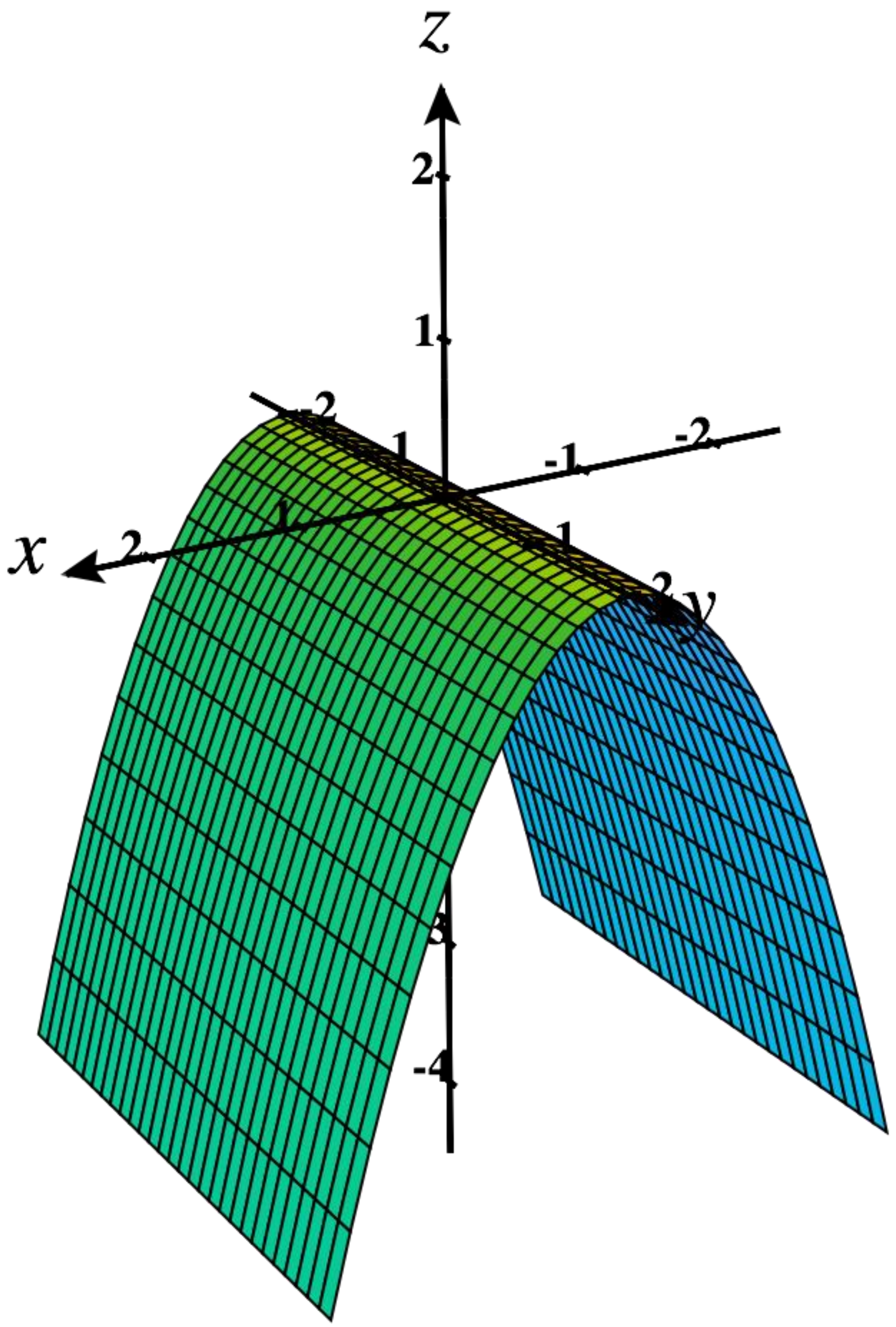
$$\vec{x}^\top A\vec{x} \geq 0 \ \forall \vec{x}$$

Quadratic Forms

What does $\vec{x}^\top A\vec{x}$ look like when:

- A is positive definite?
- A is negative definite?
- A is positive semi-definite?
- A **is negative semi-definite?**
- A is indefinite?

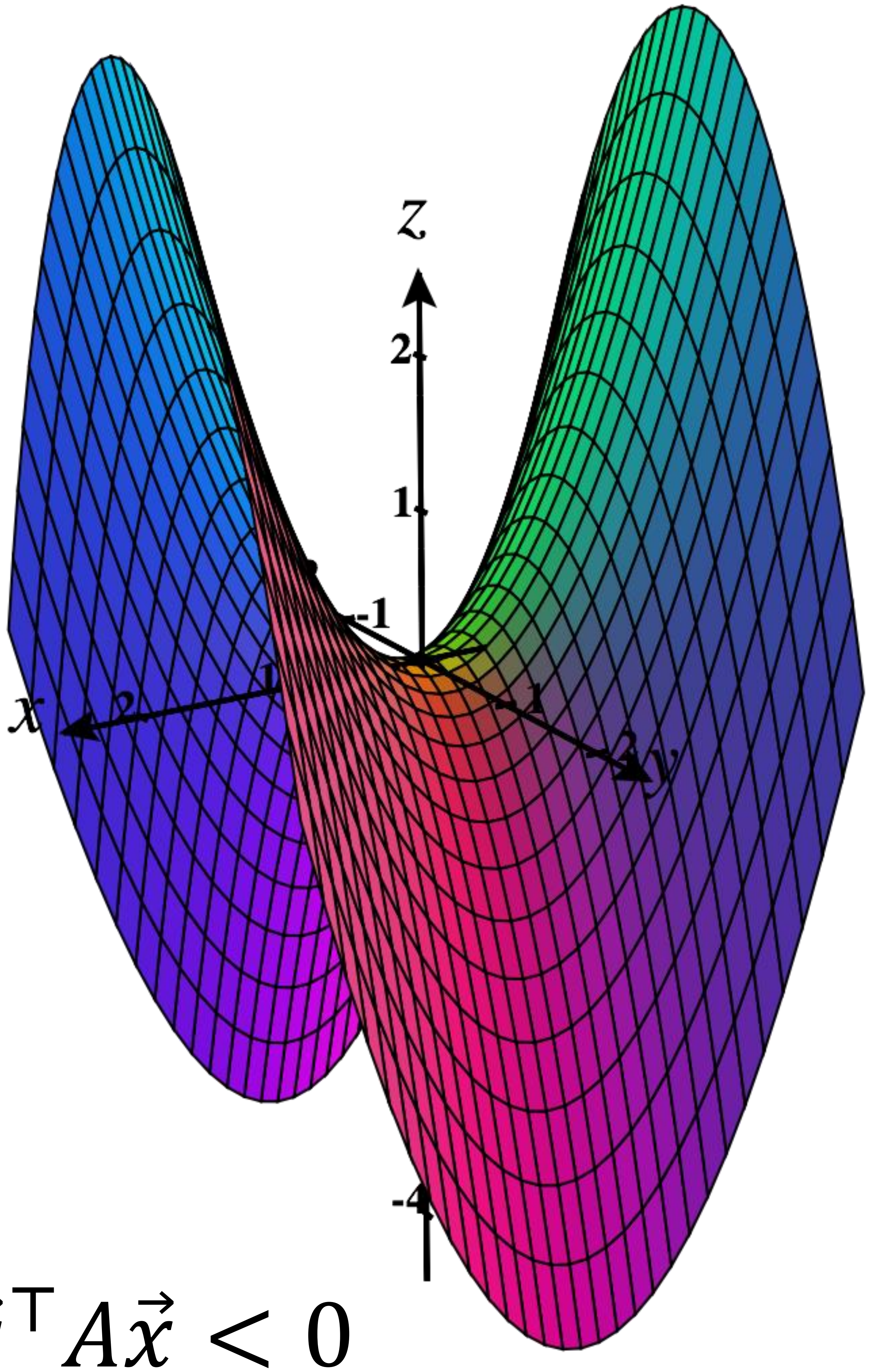
$$\vec{x}^\top A\vec{x} \leq 0 \forall \vec{x}$$



Quadratic Forms

What does $\vec{x}^\top A\vec{x}$ look like when:

- A is positive definite?
- A is negative definite?
- A is positive semi-definite?
- A is negative semi-definite?
- A **is indefinite?**
  $\exists\vec{x}$ such that $\vec{x}^\top A\vec{x} > 0$ and $\exists\vec{x}$ such that $\vec{x}^\top A\vec{x} < 0$



Quadratic Forms

What if A is non-symmetric?

Recall that the eigenvectors are not necessarily orthogonal - would weird things happen?

Quadratic Forms

What if A is non-symmetric?

Recall that the eigenvectors are not necessarily orthogonal - would weird things happen? No.

$$A = \frac{A + A^\top}{2} + \frac{A - A^\top}{2}$$

$$\vec{x}^\top A \vec{x} = \vec{x}^\top \left(\frac{A + A^\top}{2} + \frac{A - A^\top}{2} \right) \vec{x}$$

$$= \vec{x}^\top \left(\frac{A + A^\top}{2} \right) \vec{x} + \vec{x}^\top \left(\frac{A - A^\top}{2} \right) \vec{x}$$

$$\begin{aligned}
\vec{x}^\top \left(\frac{A - A^\top}{2} \right) \vec{x} &= \frac{1}{2} \vec{x}^\top A \vec{x} - \frac{1}{2} \vec{x}^\top A^\top \vec{x} \\
&= \frac{1}{2} \vec{x}^\top A \vec{x} - \frac{1}{2} (A\vec{x})^\top \vec{x} \\
&= \frac{1}{2} \langle \vec{x}, A\vec{x} \rangle - \frac{1}{2} \langle A\vec{x}, \vec{x} \rangle \\
&= \frac{1}{2} \langle \vec{x}, A\vec{x} \rangle - \frac{1}{2} \langle \vec{x}, A\vec{x} \rangle = 0
\end{aligned}$$

Quadratic Forms

What if A is non-symmetric?

Recall that the eigenvectors are not necessarily orthogonal - would weird things happen? No.

$$A = \frac{A + A^\top}{2} + \frac{A - A^\top}{2}$$

$$\begin{aligned}\vec{x}^\top A\vec{x} &= \vec{x}^\top \left(\frac{A + A^\top}{2} + \frac{A - A^\top}{2}\right)\vec{x} \\ &= \vec{x}^\top \left(\frac{A + A^\top}{2}\right)\vec{x} + \vec{x}^\top \left(\frac{A - A^\top}{2}\right)\vec{x}\end{aligned}$$

$$\begin{aligned}\vec{x}^\top \left(\frac{A - A^\top}{2}\right)\vec{x} &= \frac{1}{2}\vec{x}^\top A\vec{x} - \frac{1}{2}\vec{x}^\top A^\top\vec{x} \\ &= \frac{1}{2}\vec{x}^\top A\vec{x} - \frac{1}{2}(A\vec{x})^\top\vec{x} \\ &= \frac{1}{2}\langle\vec{x}, A\vec{x}\rangle - \frac{1}{2}\langle A\vec{x}, \vec{x}\rangle \\ &= \frac{1}{2}\langle\vec{x}, A\vec{x}\rangle - \frac{1}{2}\langle\vec{x}, A\vec{x}\rangle = 0\end{aligned}$$

Hence $\vec{x}^\top A\vec{x} = \vec{x}^\top \left(\frac{A+A^\top}{2}\right)\vec{x}$

Quadratic Forms

What if A is non-symmetric?

Recall that the eigenvectors are not necessarily orthogonal - would weird things happen? No.

$$A = \frac{A + A^\top}{2} + \frac{A - A^\top}{2}$$

$$\vec{x}^\top A\vec{x} = \vec{x}^\top \left(\frac{A + A^\top}{2} + \frac{A - A^\top}{2}\right)\vec{x}$$

$$= \vec{x}^\top \left(\frac{A + A^\top}{2}\right)\vec{x} + \vec{x}^\top \left(\frac{A - A^\top}{2}\right)\vec{x}$$

$$\begin{aligned}
\vec{x}^\top \left(\frac{A - A^\top}{2}\right)\vec{x} &= \frac{1}{2}\vec{x}^\top A\vec{x} - \frac{1}{2}\vec{x}^\top A^\top \vec{x} \\
&= \frac{1}{2}\vec{x}^\top A\vec{x} - \frac{1}{2}(A\vec{x})^\top \vec{x} \\
&= \frac{1}{2}\langle \vec{x}, A\vec{x}\rangle - \frac{1}{2}\langle A\vec{x}, \vec{x}\rangle \\
&= \frac{1}{2}\langle \vec{x}, A\vec{x}\rangle - \frac{1}{2}\langle \vec{x}, A\vec{x}\rangle = 0
\end{aligned}$$

Hence $\vec{x}^\top A\vec{x} = \vec{x}^\top \left(\frac{A+A^\top}{2}\right)\vec{x}$

$\left(\frac{A + A^\top}{2}\right)$ is always a symmetric matrix

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$ (i.e.: $A^\top A$ is positive semi-definite)

Show this.

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$ (i.e.: $A^\top A$ is positive semi-definite)

Show this.

$\vec{x}^\top (A^\top A) \vec{x}$

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$

(i.e.: $A^\top A$ is positive semi-definite)

Show this.

$\vec{x}^\top (A^\top A)\vec{x} = (\vec{x}^\top A^\top)(A\vec{x})$

$(AB)C = A(BC)$, but $AB \neq BA$

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$

(i.e.: $A^\top A$ is positive semi-definite)

Show this.

$$\begin{aligned}\vec{x}^\top(A^\top A)\vec{x} &= (\vec{x}^\top A^\top)(A\vec{x}) \\ &= (A\vec{x})^\top(A\vec{x})\end{aligned}$$

$(AB)^\top = B^\top A^\top$

Quadratic Forms

For **any matrix** A:

$$A^\top A \succcurlyeq 0$$

(i.e.: $A^\top A$ is positive semi-definite)

Show this.

$$\begin{aligned}\vec{x}^\top (A^\top A)\vec{x} &= (\vec{x}^\top A^\top)(A\vec{x}) \\ &= (A\vec{x})^\top (A\vec{x}) \\ &= \langle A\vec{x}, A\vec{x} \rangle\end{aligned}$$

Alternative inner product notation:

$$\vec{x}^\top \vec{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} x_i y_i = \langle \vec{x}, \vec{y} \rangle$$

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$

(i.e.: $A^\top A$ is positive semi-definite)

Show this.

$$\begin{aligned}\vec{x}^\top(A^\top A)\vec{x} &= (\vec{x}^\top A^\top)(A\vec{x})\\ &= (A\vec{x})^\top(A\vec{x})\\ &= \langle A\vec{x}, A\vec{x}\rangle\\ &= \|A\vec{x}\|_2^2\end{aligned}$$

Euclidean norm:

$$\|\vec{x}\|_2 = \sqrt{\langle\vec{x},\vec{x}\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\|\vec{x}\|_2 \geq 0 \ \forall\vec{x}$$

Quadratic Forms

For **any matrix** A:

$A^\top A \succcurlyeq 0$

(i.e.: $A^\top A$ is positive semi-definite)

Show this.

$$\begin{aligned}
\vec{x}^\top (A^\top A)\vec{x} &= (\vec{x}^\top A^\top)(A\vec{x}) \\
&= (A\vec{x})^\top (A\vec{x}) \\
&= \langle A\vec{x}, A\vec{x} \rangle \\
&= \|A\vec{x}\|_2^2 \\
&\geq 0 \; \forall A, \vec{x}
\end{aligned}$$

Euclidean norm:

$$\|\vec{x}\|_2 = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\|\vec{x}\|_2 \geq 0 \; \forall \vec{x}$$