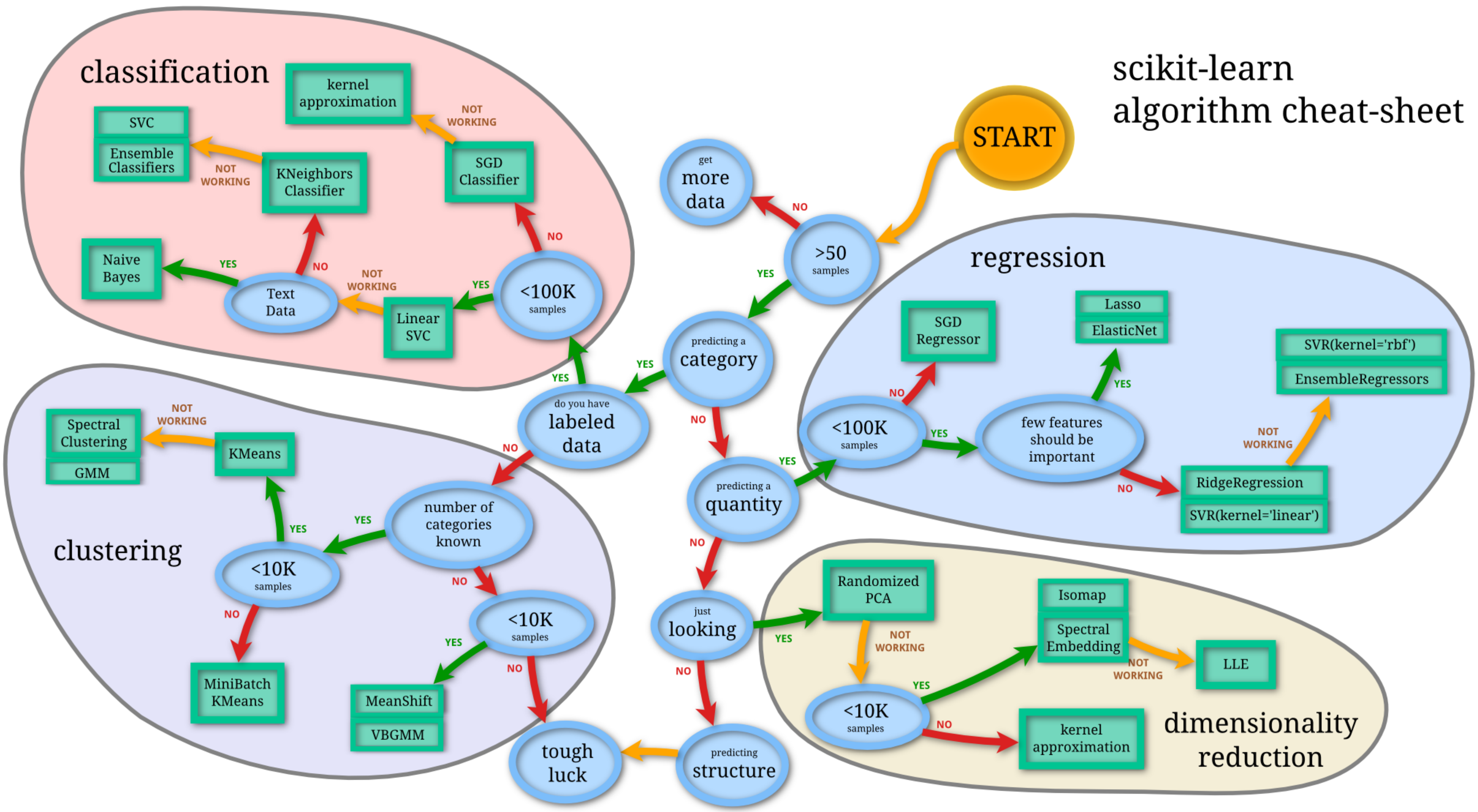
LINEAR ALGEBRA

Sourav Sen Gupta


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
<100K samples
predicting a quantity
just looking
predicting structure
tough luck
classification
kernel approximation
SVC
Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
clustering
Spectral Clustering
GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
MeanShift
VBGMM
regression
SGD Regressor
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
few features should be important
RidgeRegression
SVR(kernel='linear')
dimensionality reduction
Randomized PCA
Isomap
Spectral Embedding
LLE
kernel approximation
YES
NO
NOT WORKING

PLAN OF ACTION

Part I — Geometry of Multivariate Data

1. Vector space modelling of multivariate data
2. Understanding Matrices as linear operators
3. Subspaces and Singular Value Decomposition

Part II — Applications in Multivariate Analysis

1. SVD and Principal Component Analysis
2. Eigenvalues and Eigenvectors in PageRank
3. Factor Analysis and SVD in Recommenders

MULTIVARIATE DATA

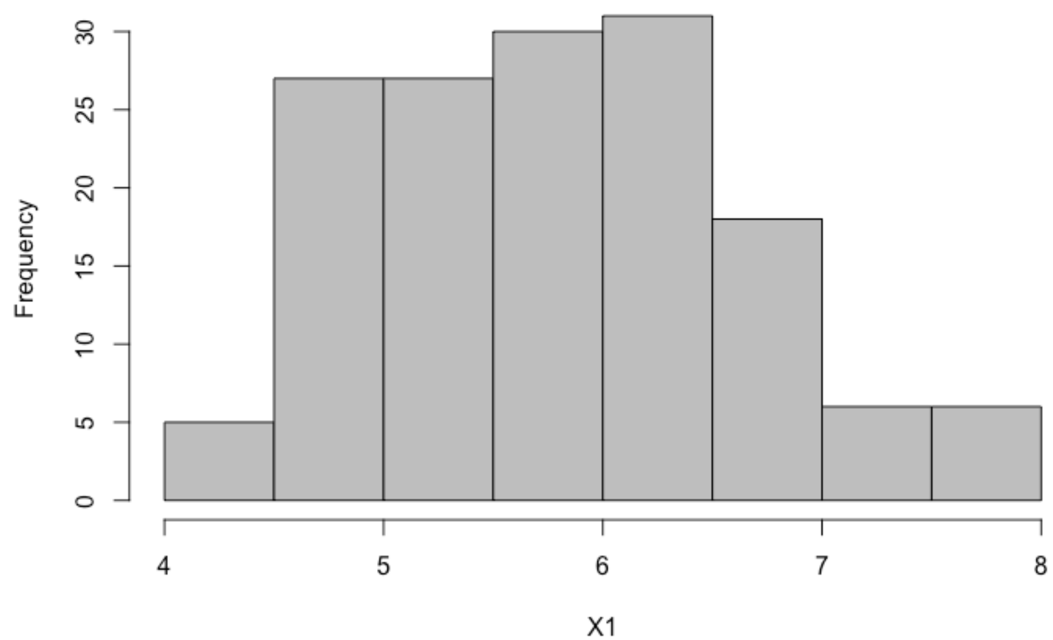
MULTIVARIATE DATA

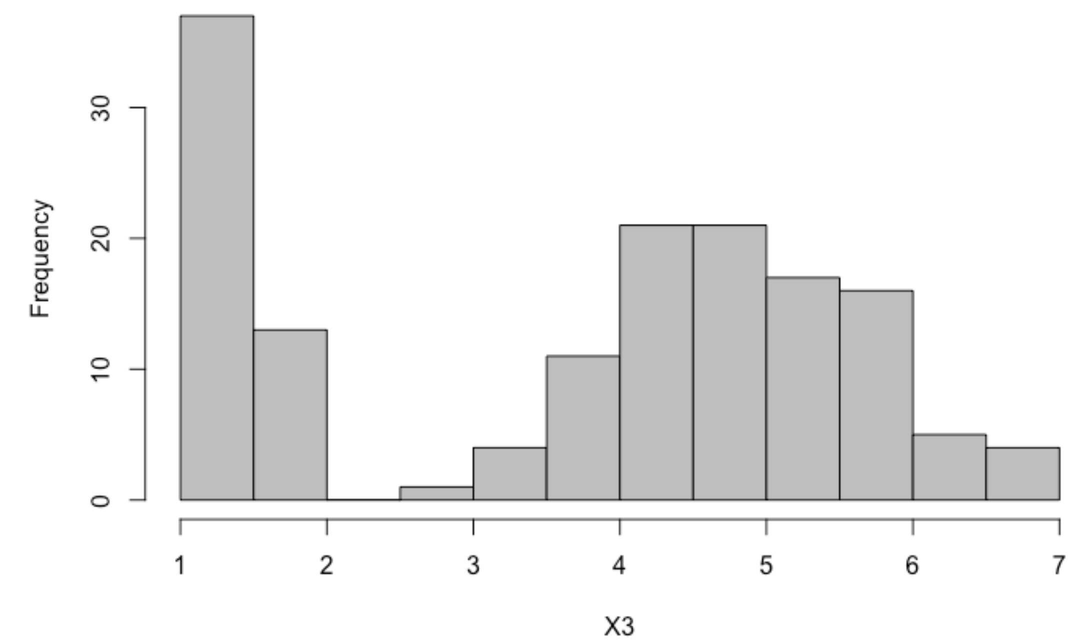
p variables/features

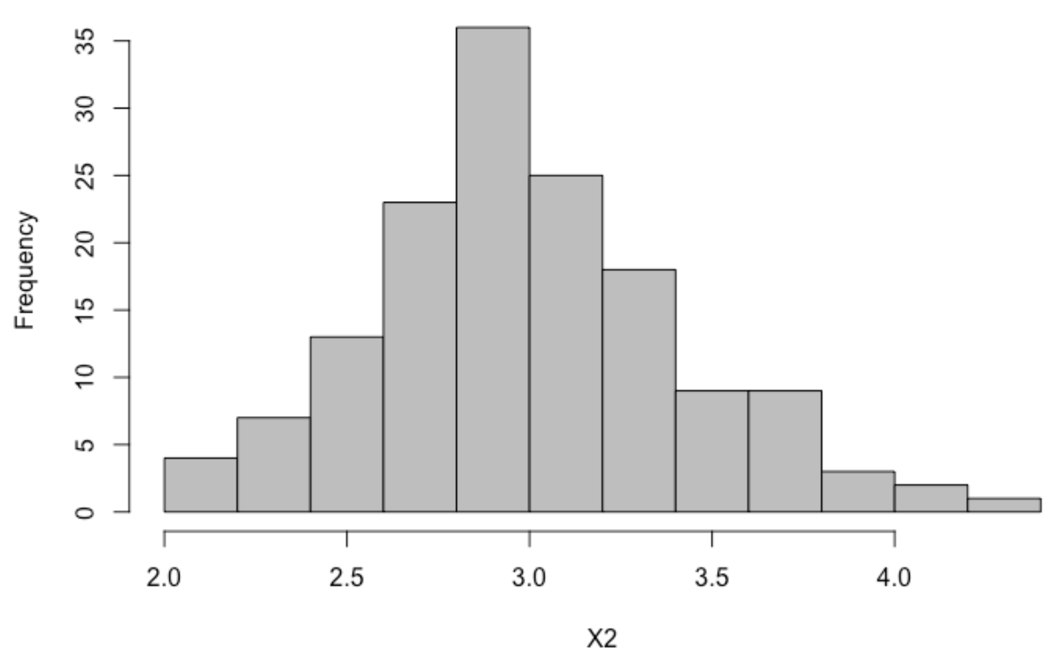
n samples/observations

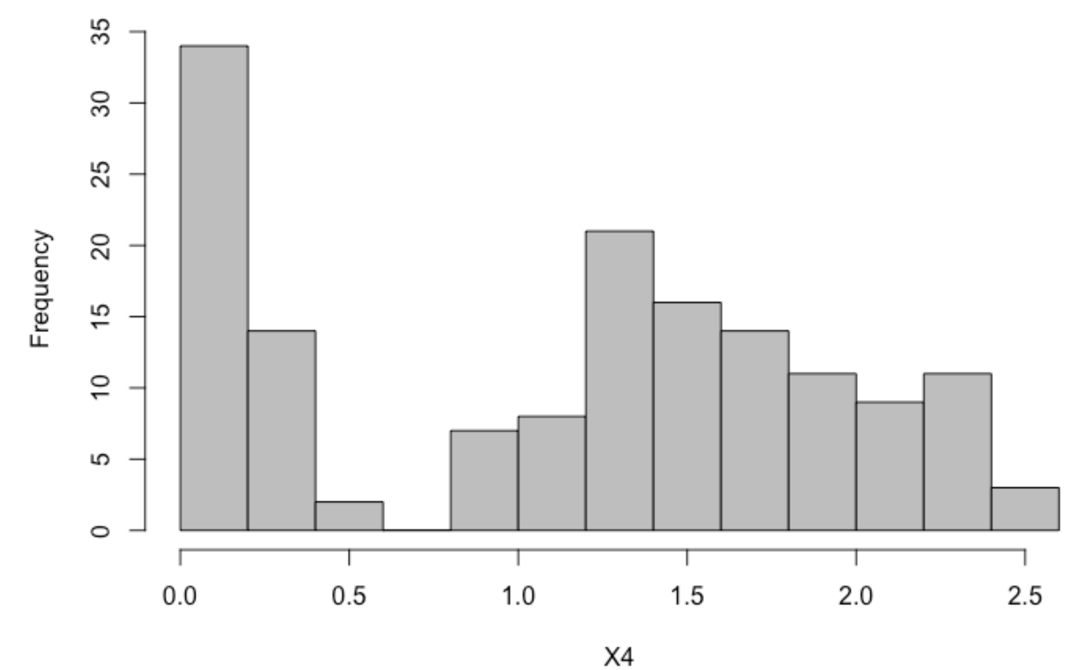
	X1	X2	X3	X4
S1	5.1	3.5	1.4	0.2
S2	4.9	3	1.4	0.2
S3	4.7	3.2	1.3	0.2
S4	4.6	3.1	1.5	0.2
S5	5	3.6	1.4	0.2
⋮	⋮	⋮	⋮	⋮
S146	6.7	3	5.2	2.3
S147	6.3	2.5	5	1.9
S148	6.5	3	5.2	2
S149	6.2	3.4	5.4	2.3
S150	5.9	3	5.1	1.8

MULTIVARIATE DATA — HISTOGRAMS

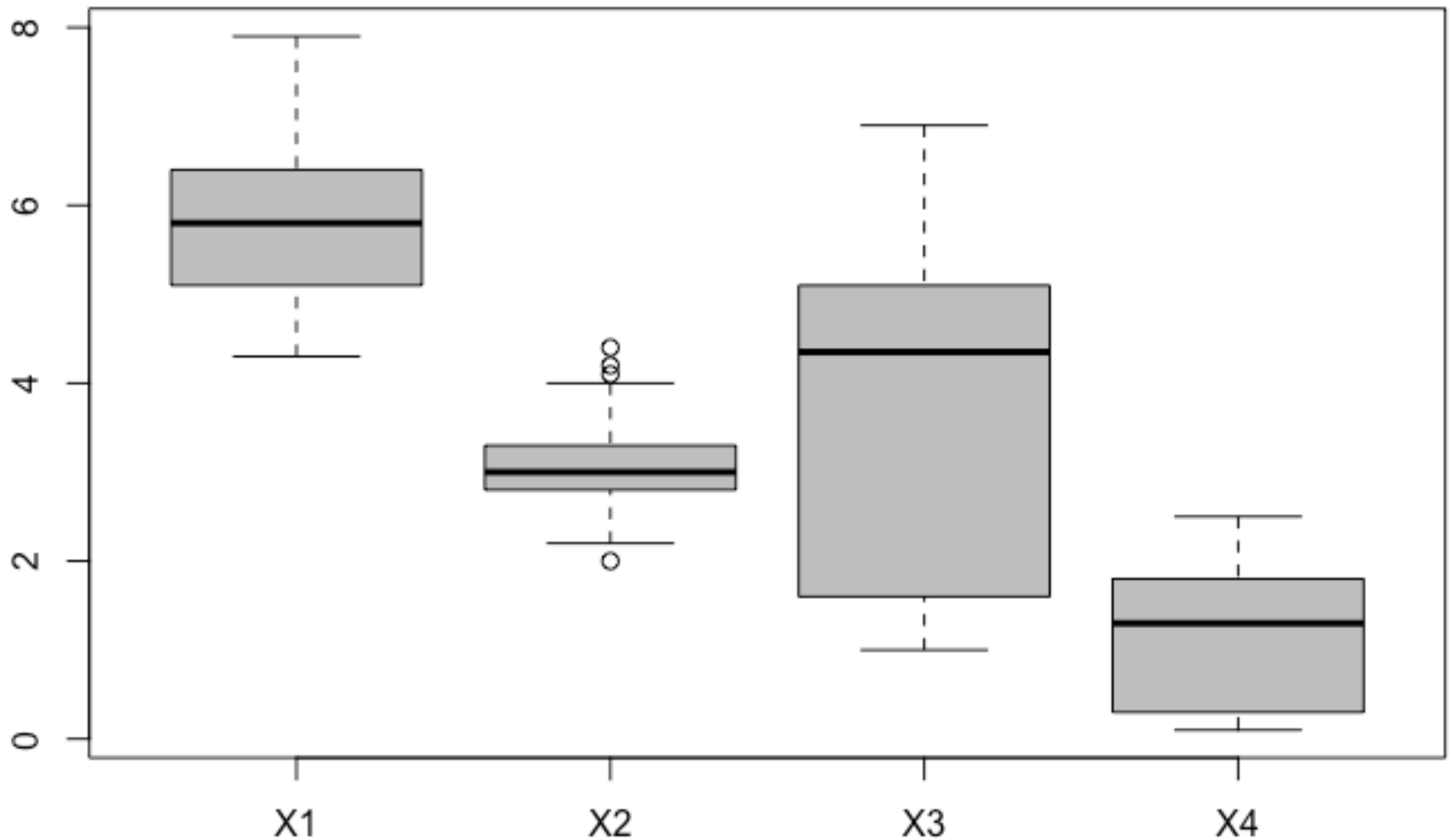




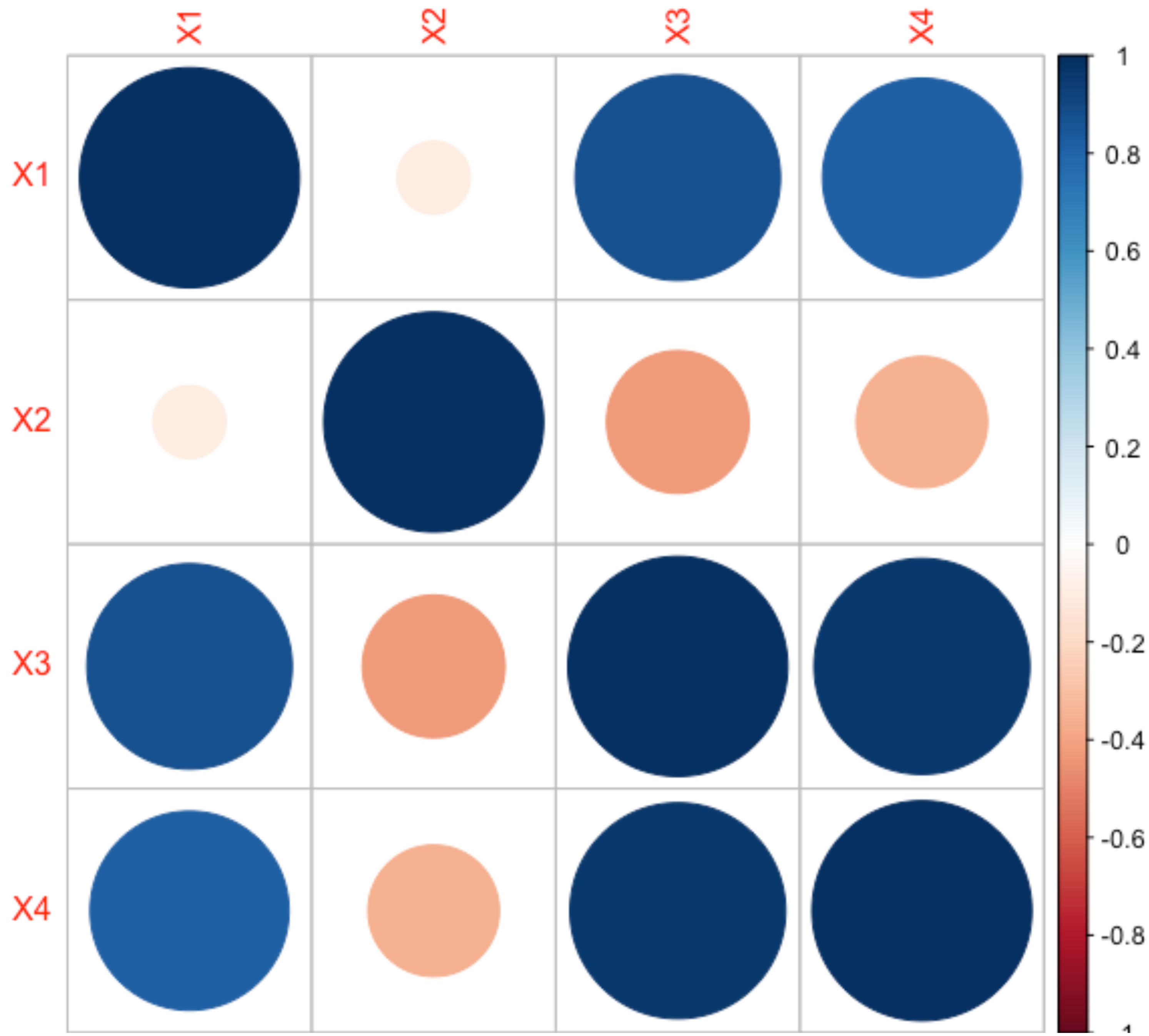




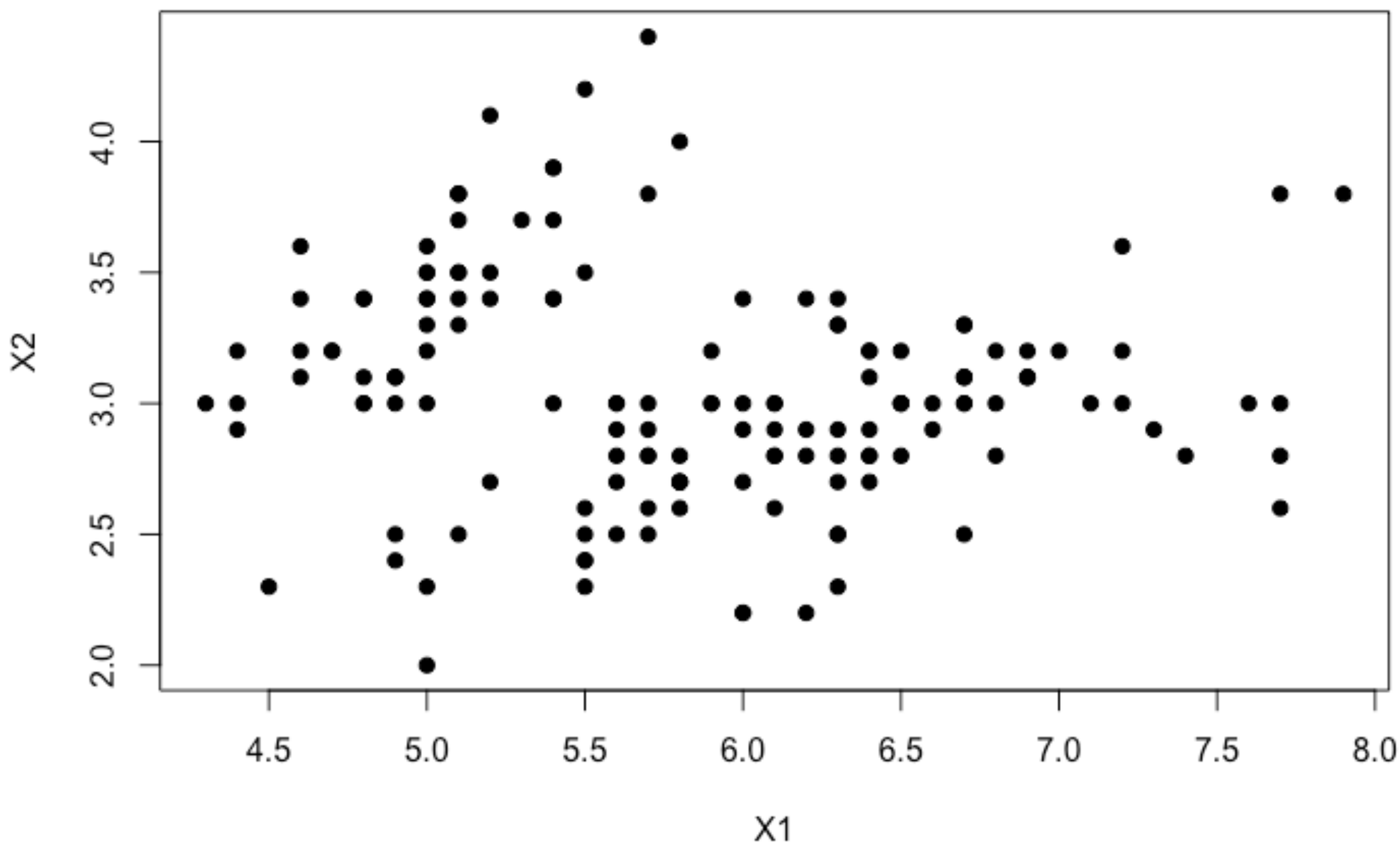
MULTIVARIATE DATA — BOXPLOTS



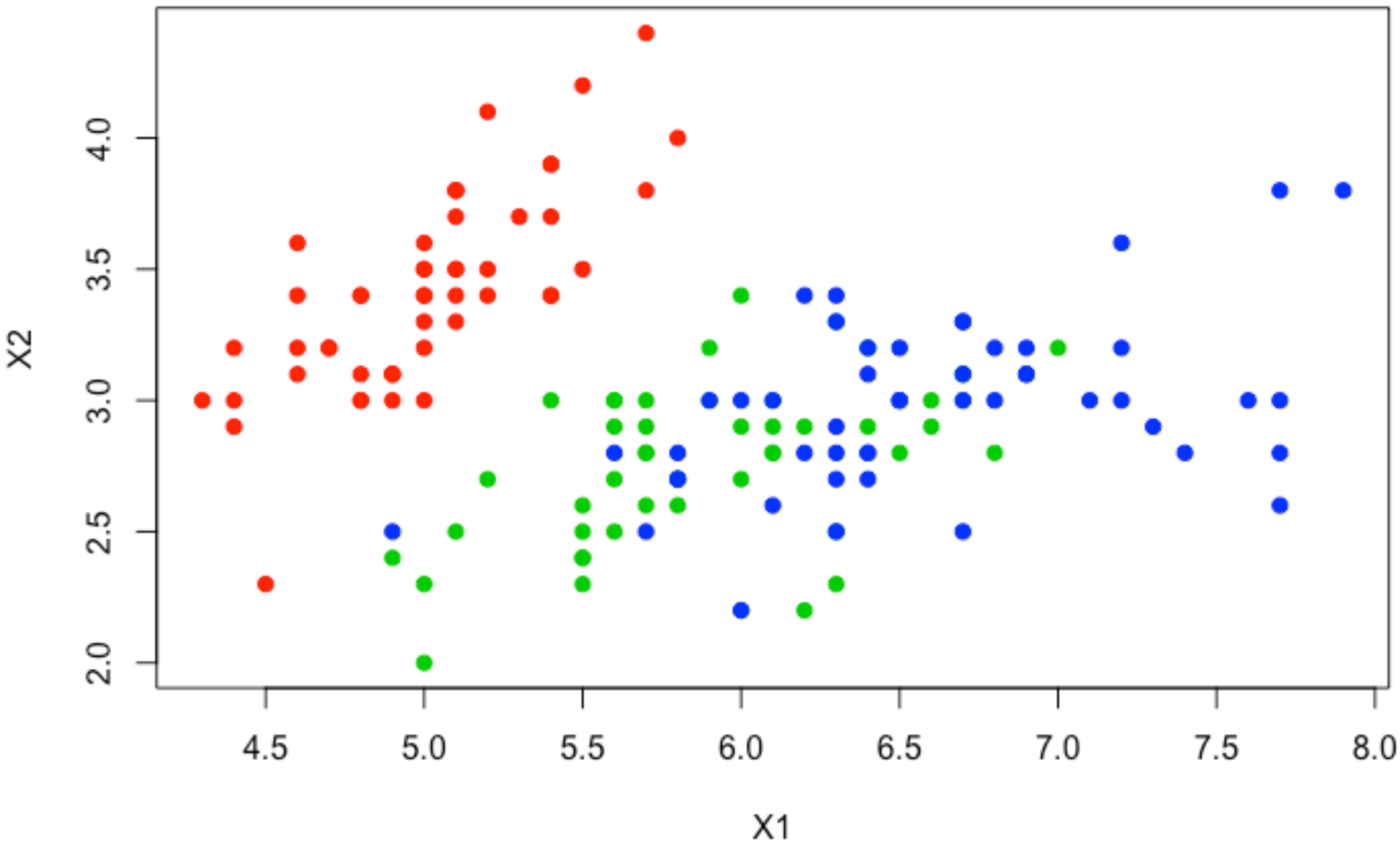
MULTIVARIATE DATA — CORRPLOT



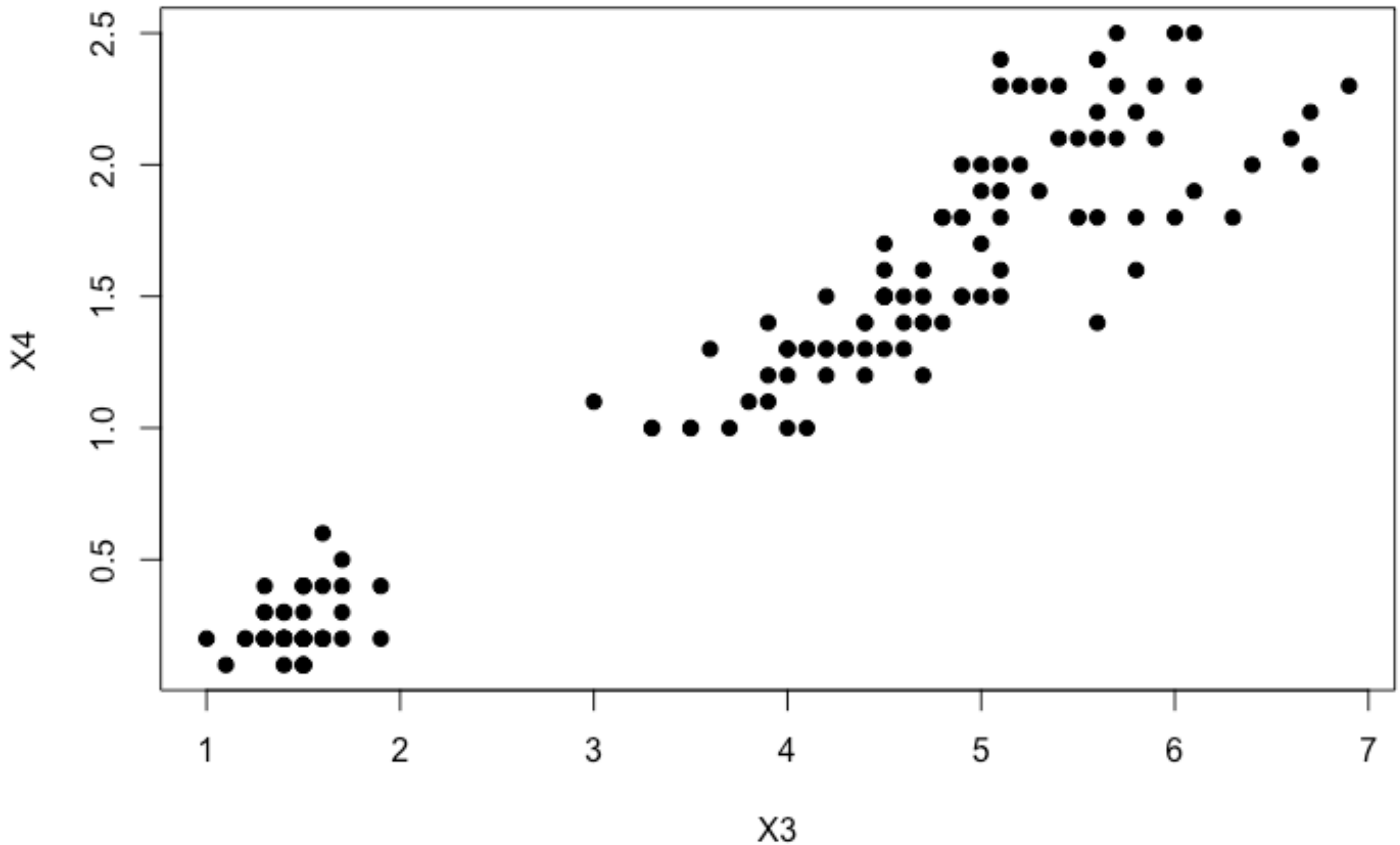
MULTIVARIATE DATA — SCATTERPLOT



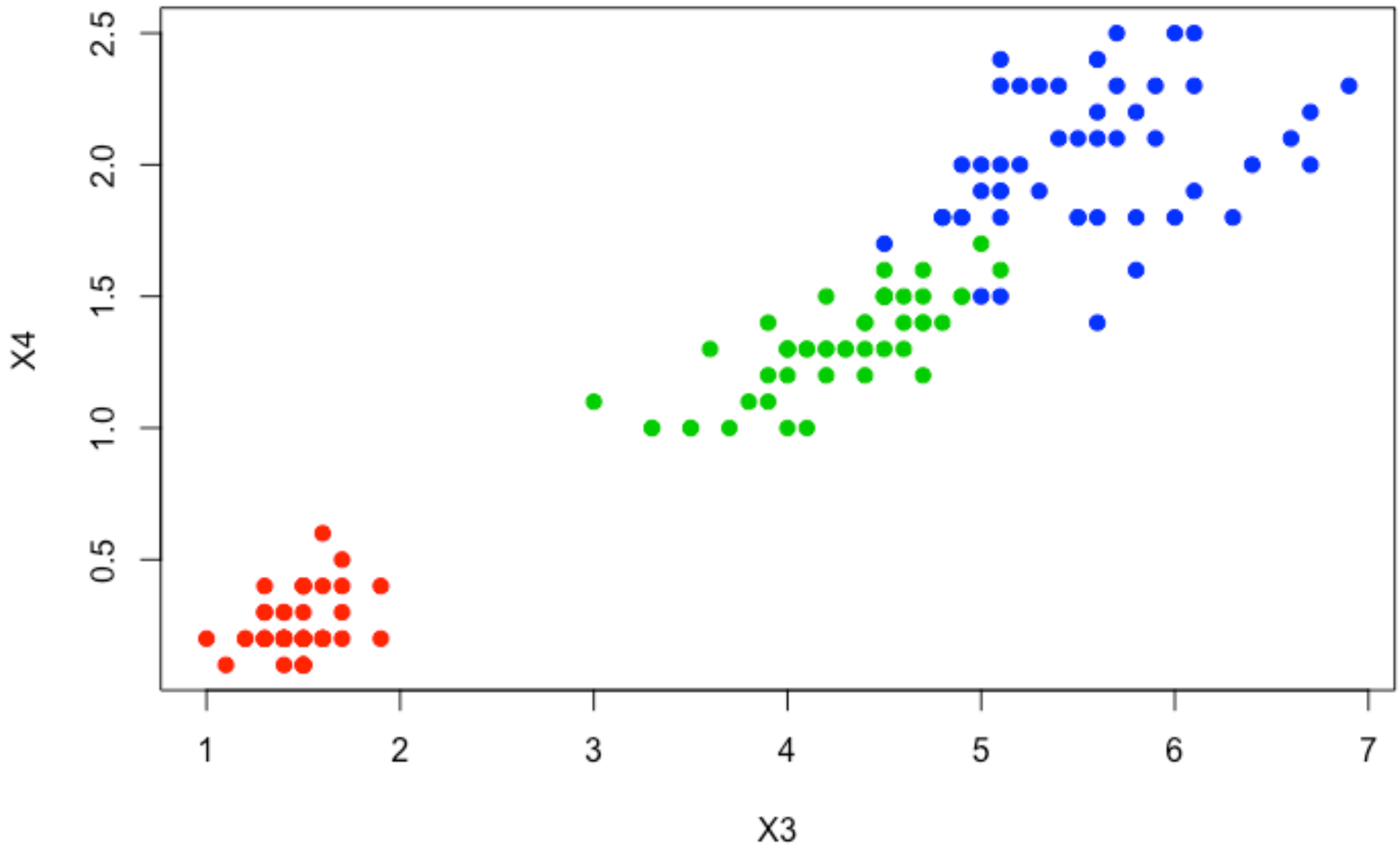
MULTIVARIATE DATA — SCATTERPLOT



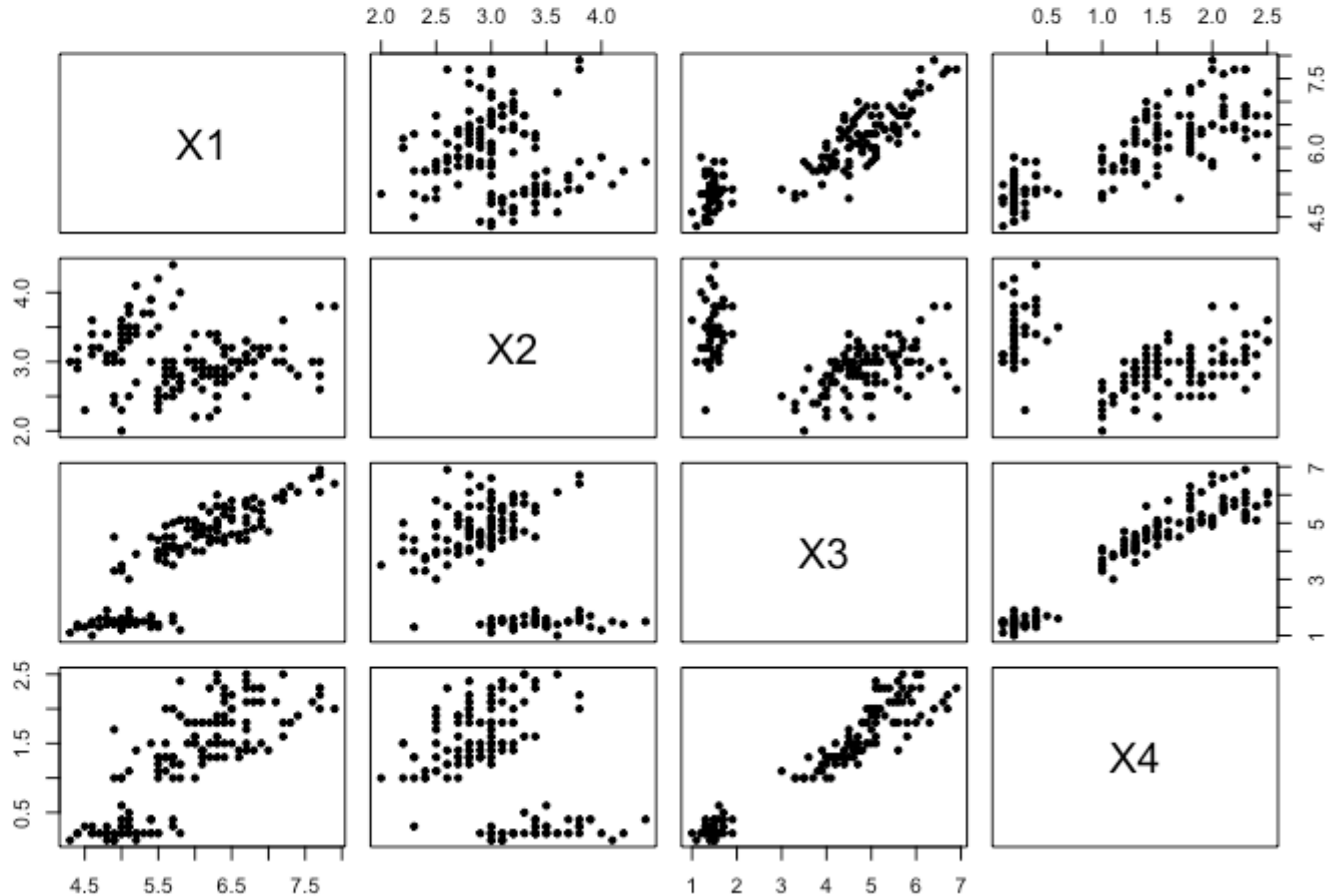
MULTIVARIATE DATA — SCATTERPLOT



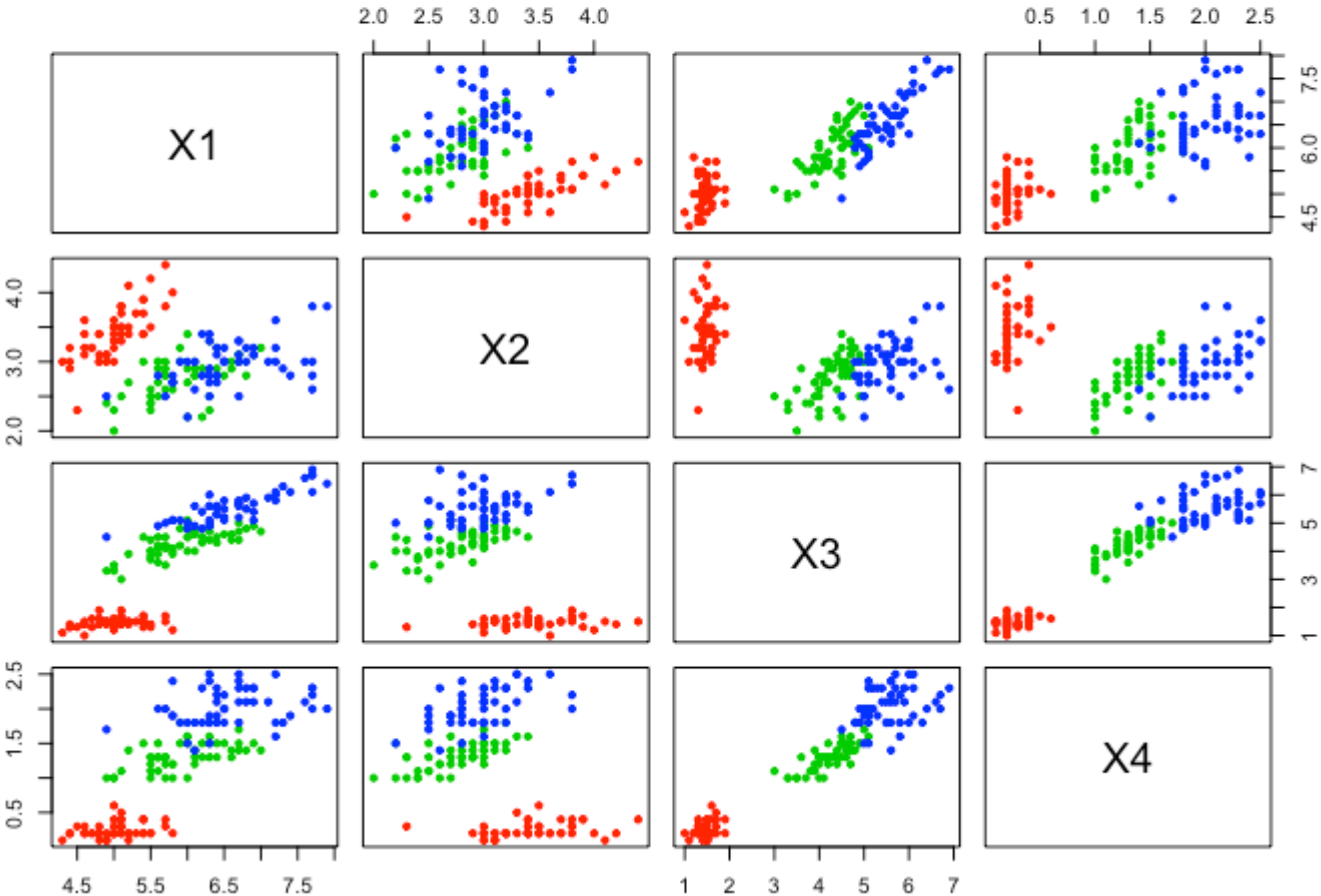
MULTIVARIATE DATA — SCATTERPLOT



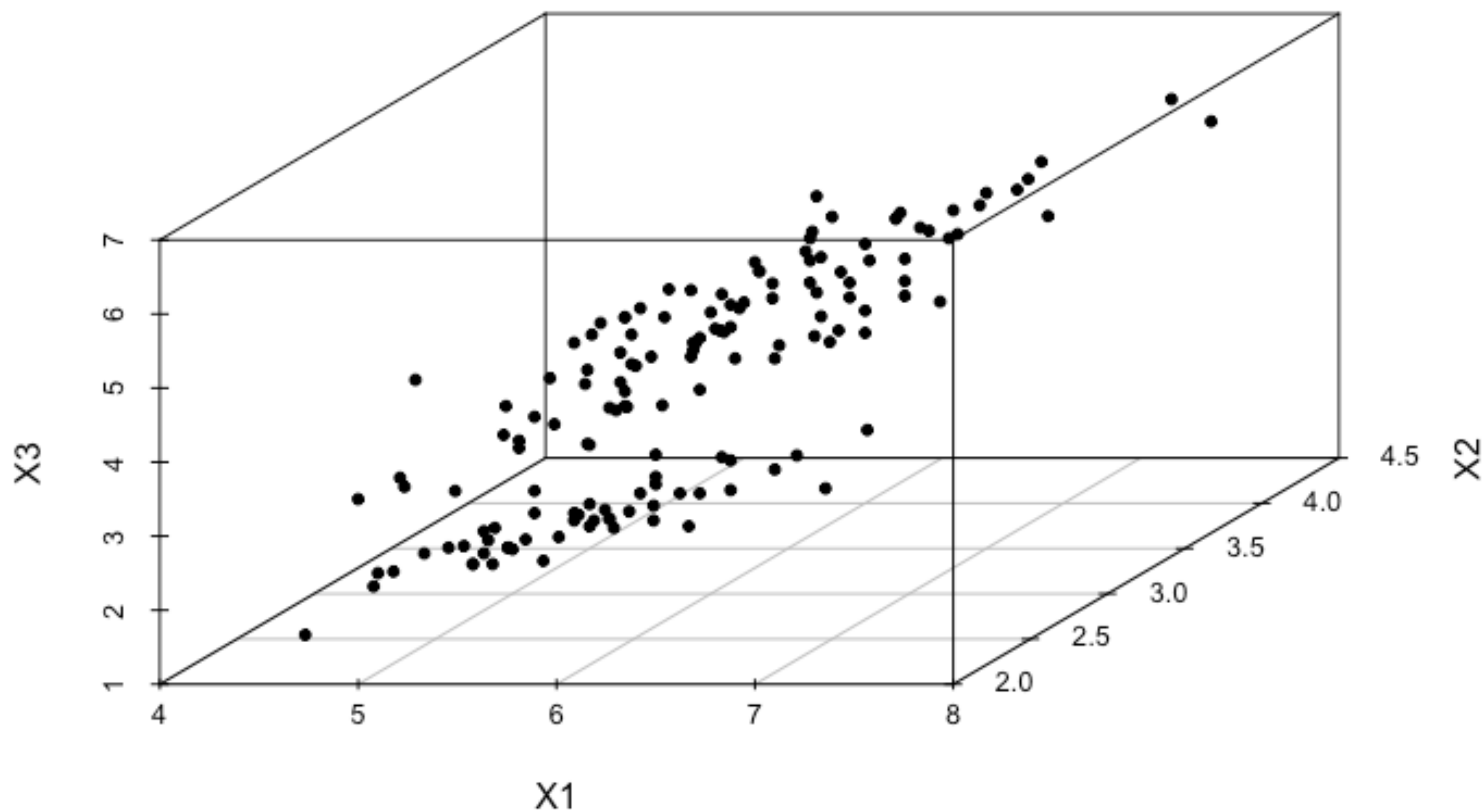
MULTIVARIATE DATA —— SCATTERPLOTS



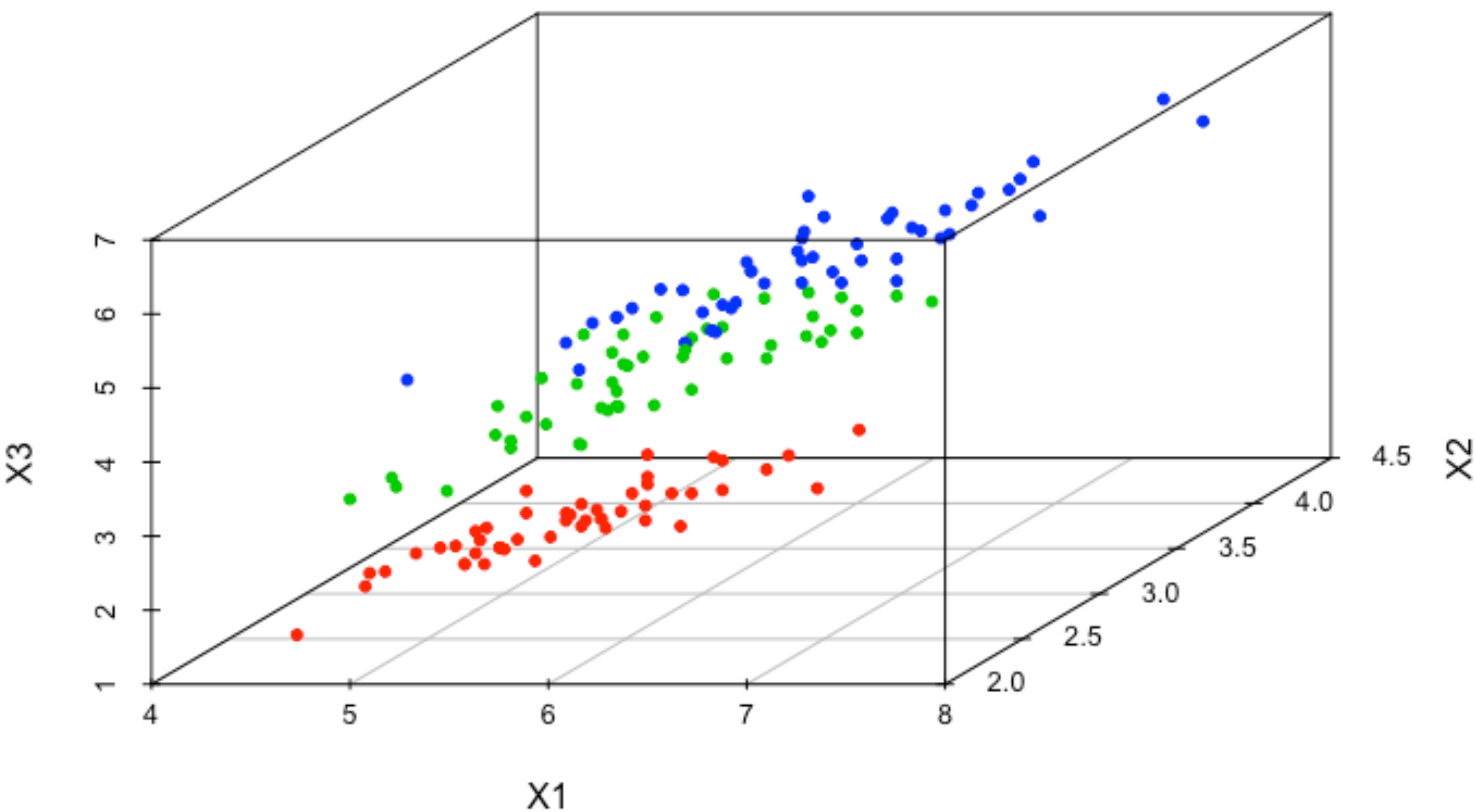
MULTIVARIATE DATA — SCATTERPLOTS



MULTIVARIATE DATA — 3D SCATTERPLOT

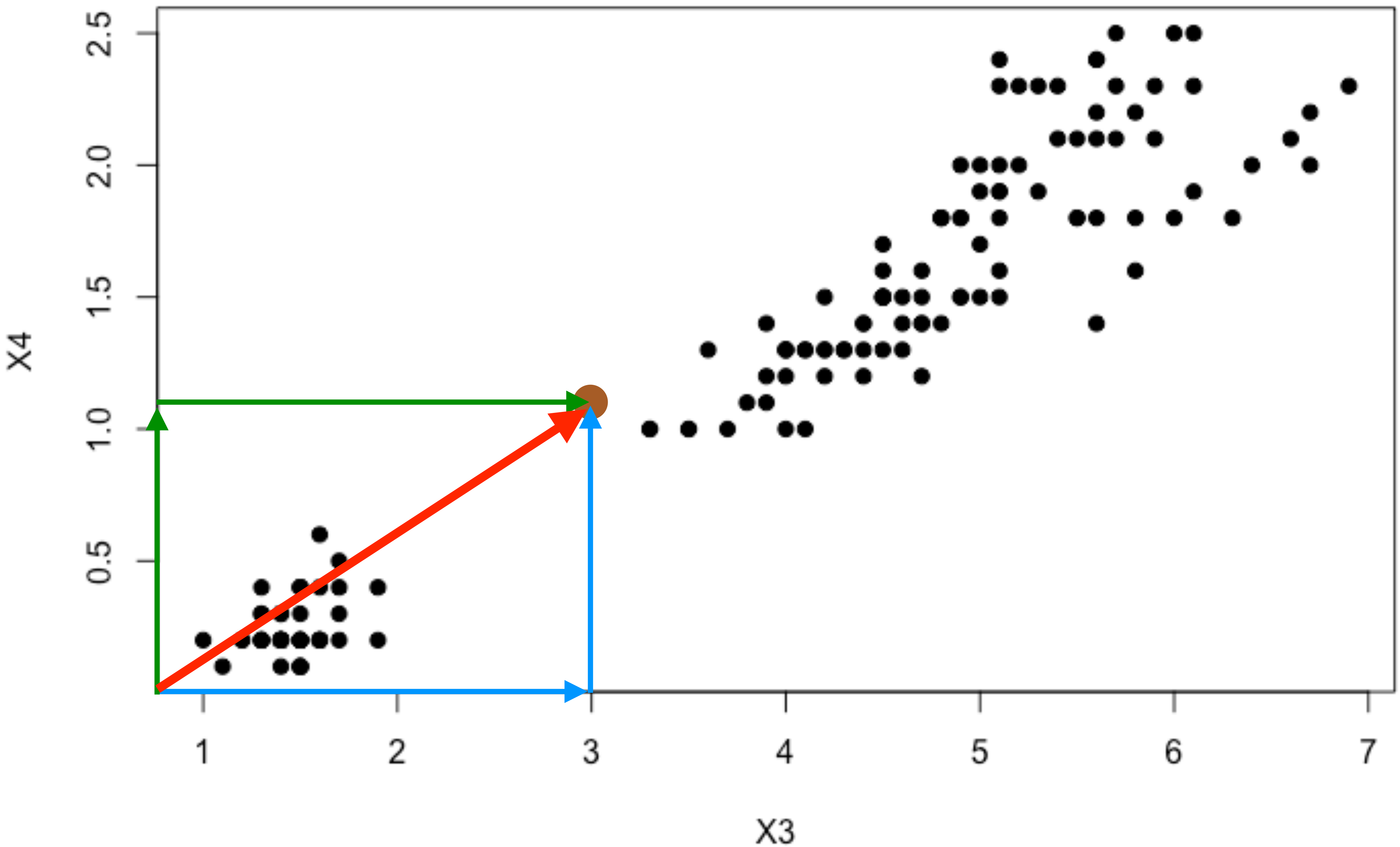


MULTIVARIATE DATA — 3D SCATTERPLOT

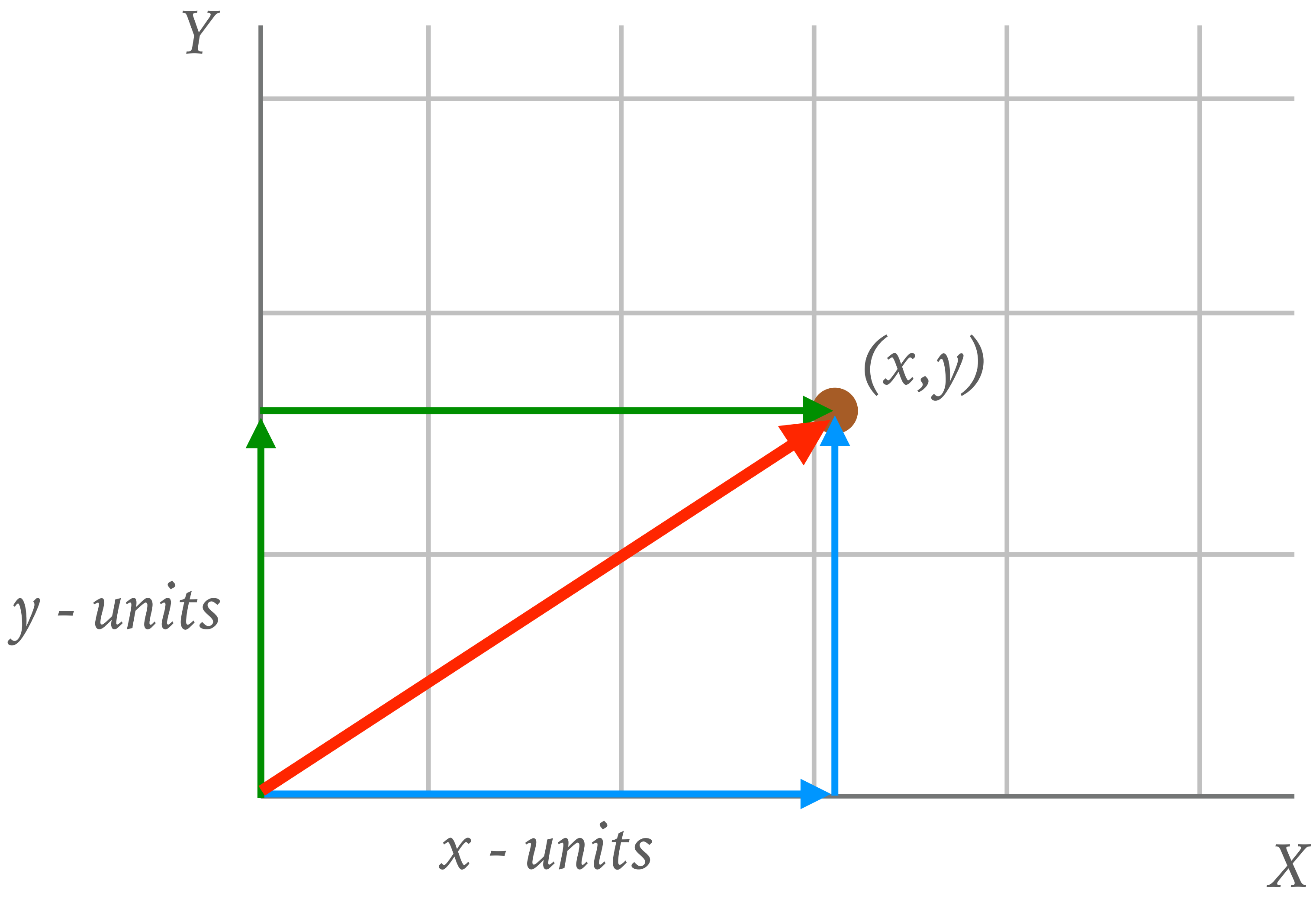


LET'S START WITH 2D

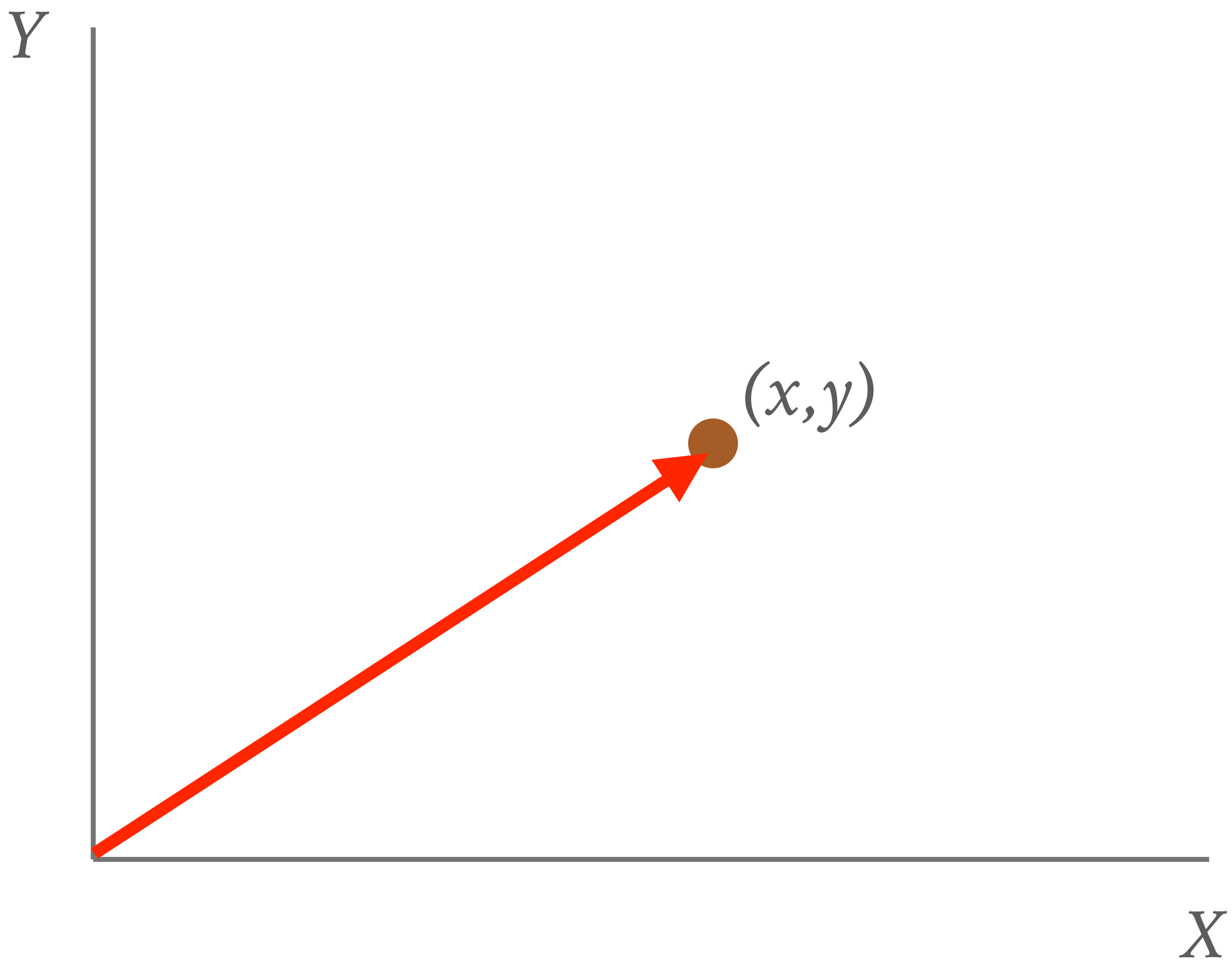
2D GEOMETRY



2D GEOMETRY



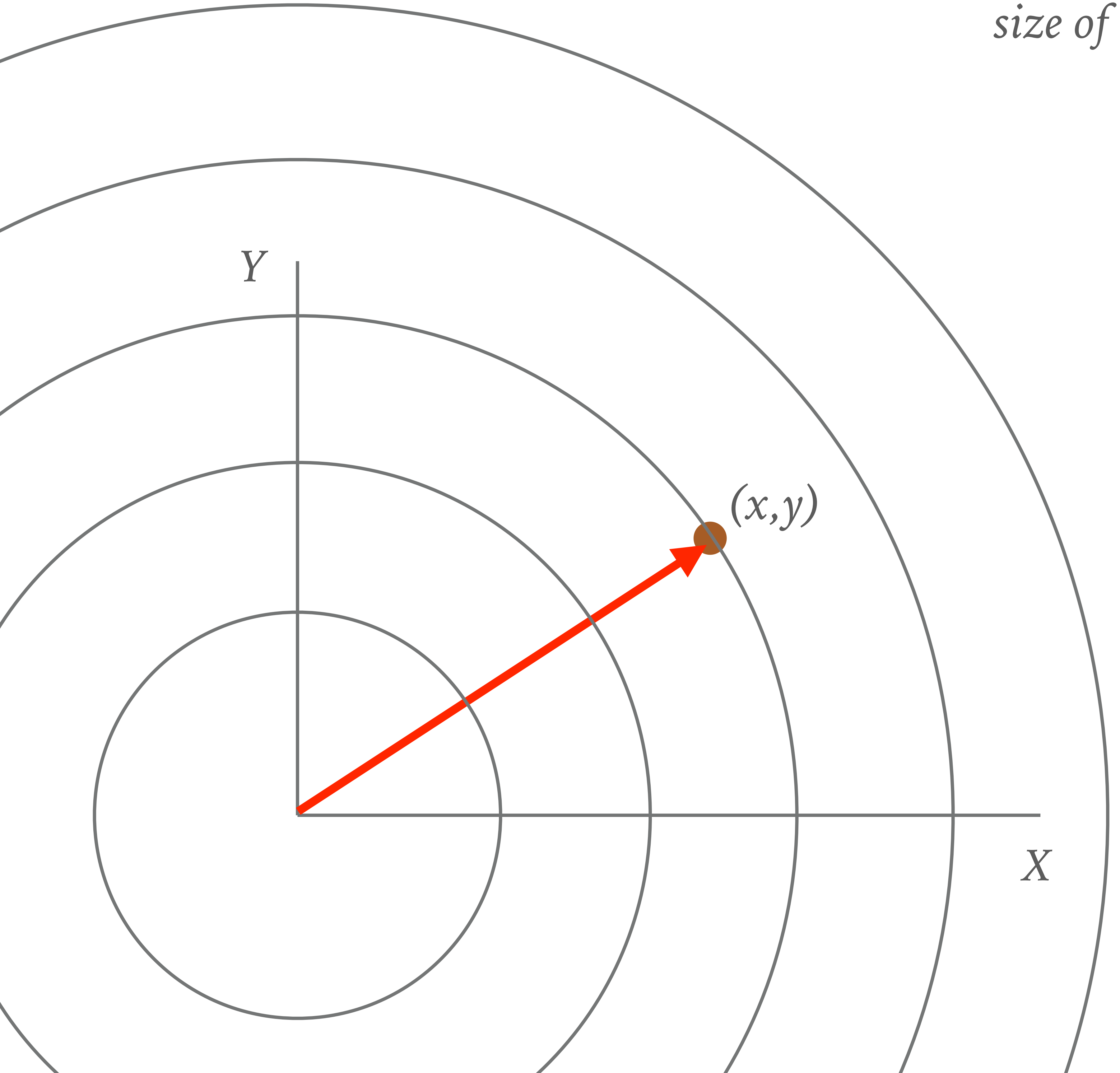
SIZE



SIZE

L2 Norm

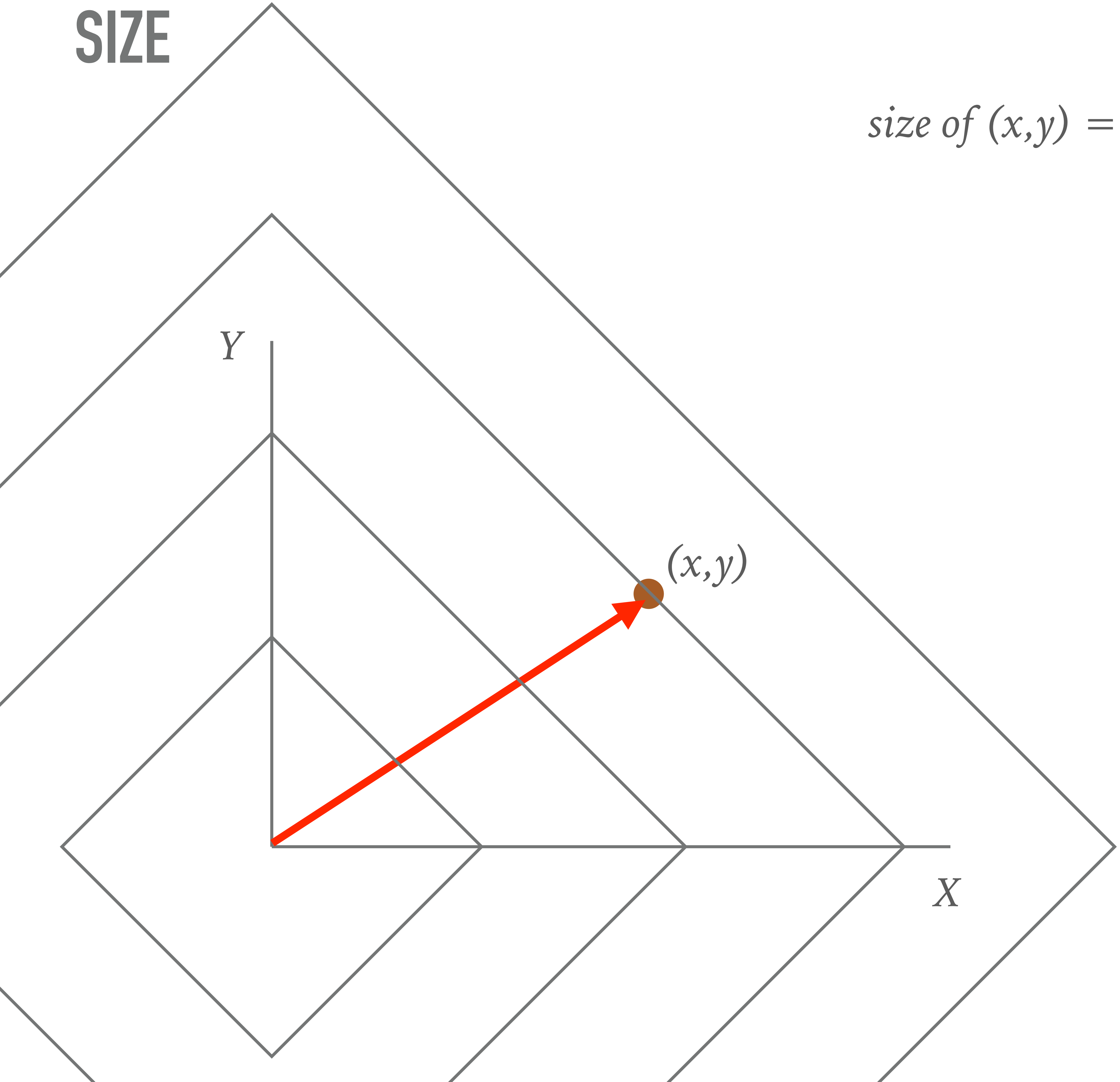
size of $(x,y) = sqrt(x^2 + y^2)$



SIZE

L1 Norm

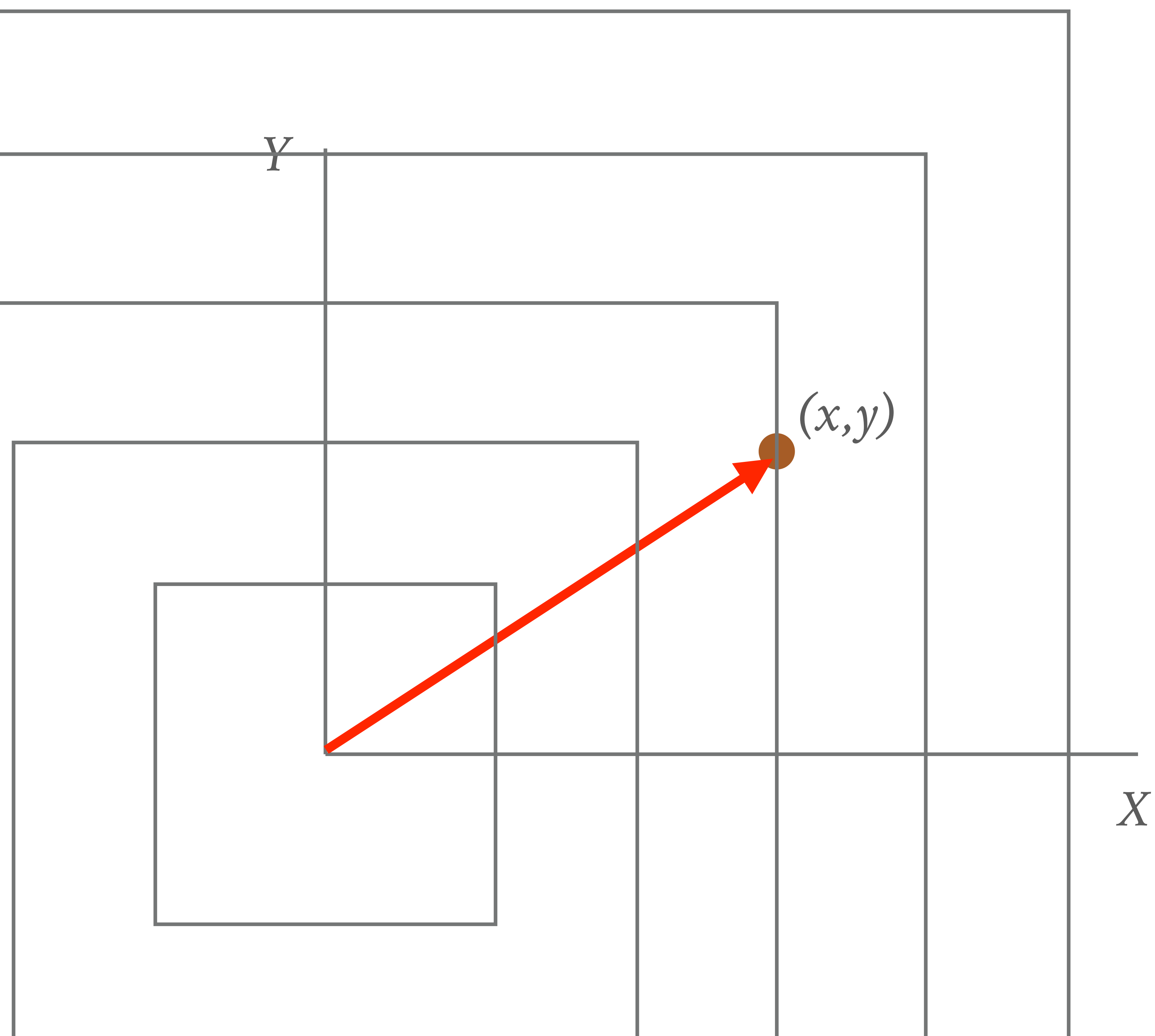
$$size\ of\ (x,y) = abs(x) + abs(y)$$



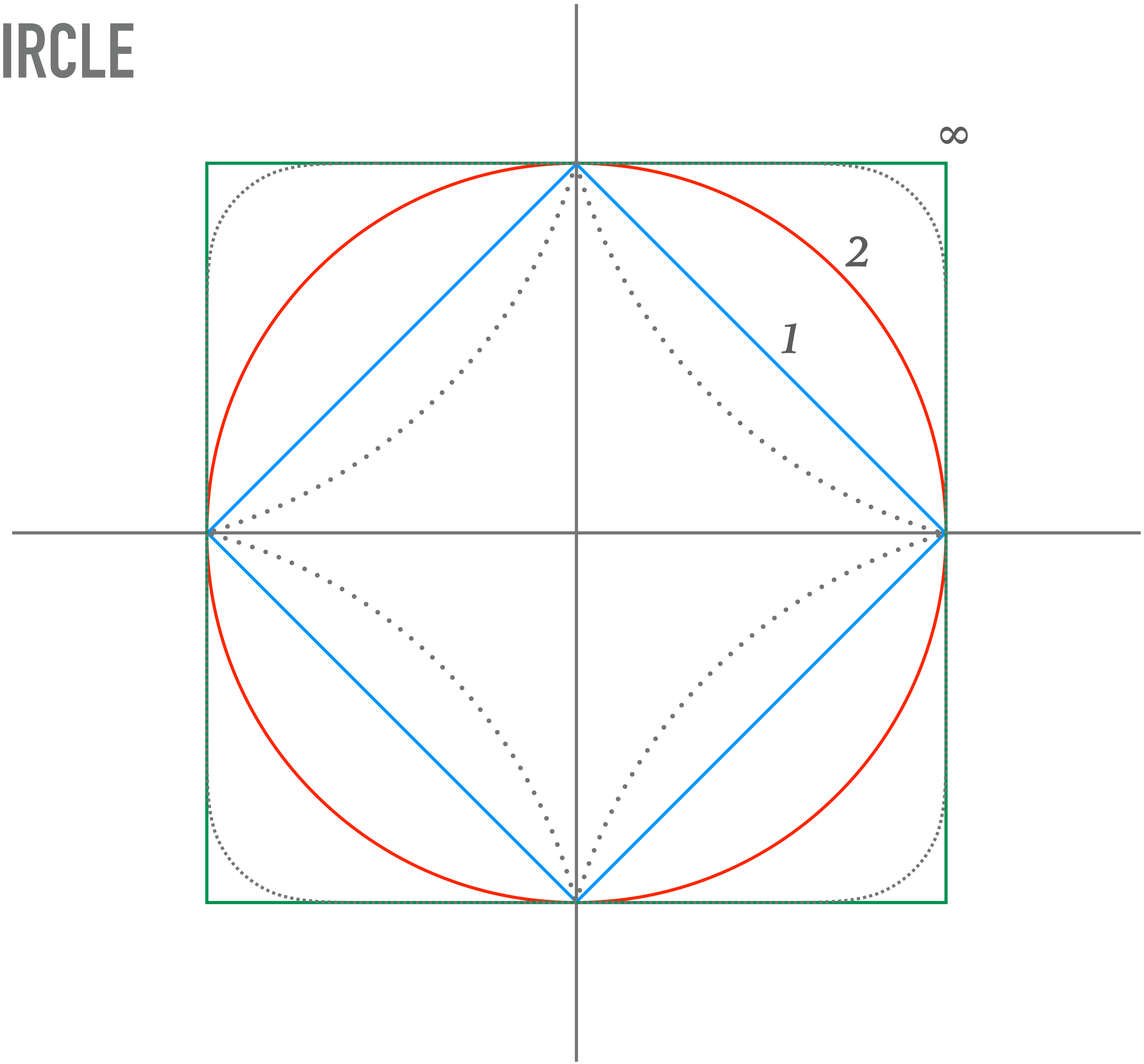
SIZE

L∞ Norm

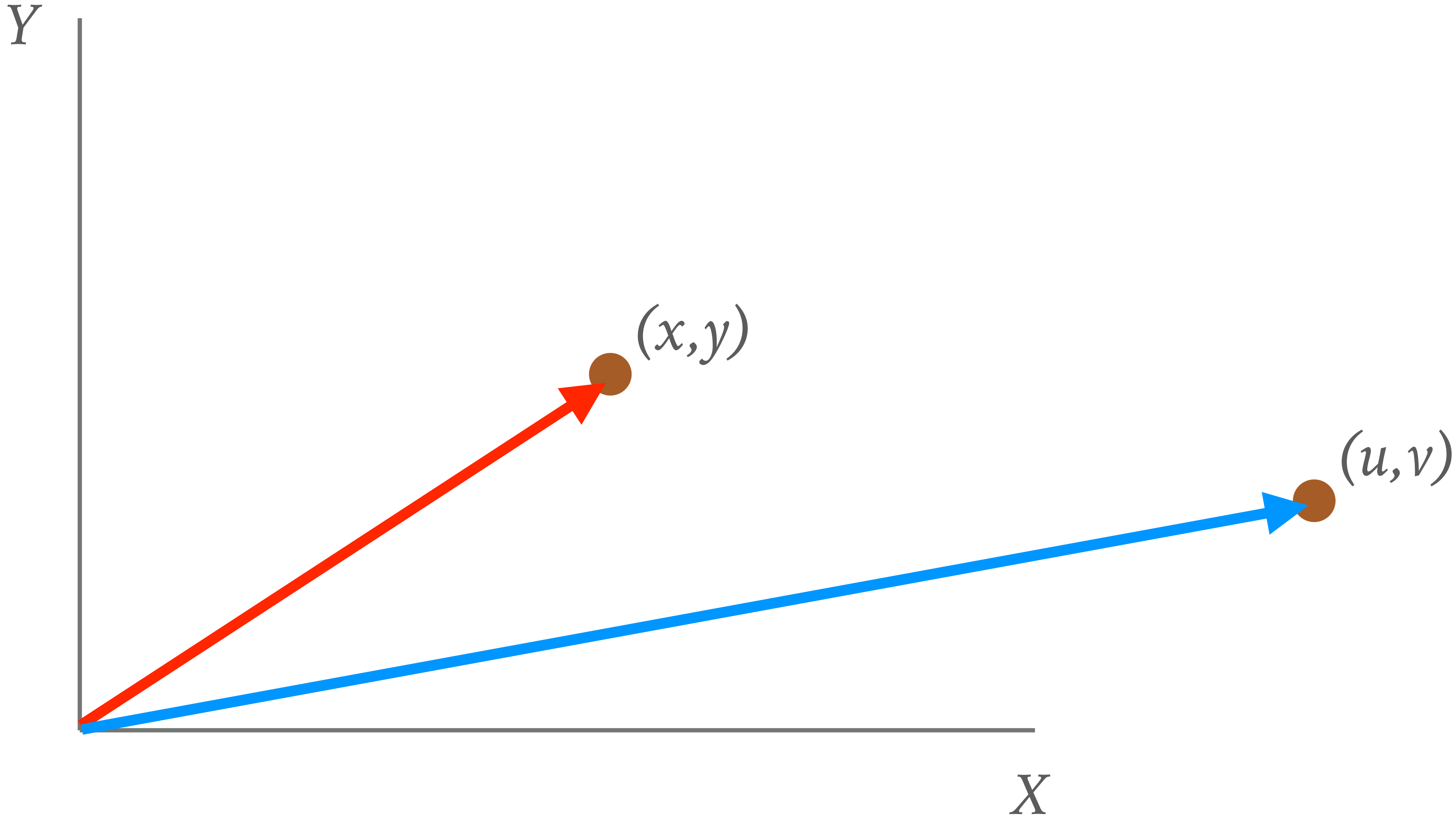
$$\textit{size of } (x,y) = \max(\mathrm{abs}(x), \mathrm{abs}(y))$$



UNIT CIRCLE



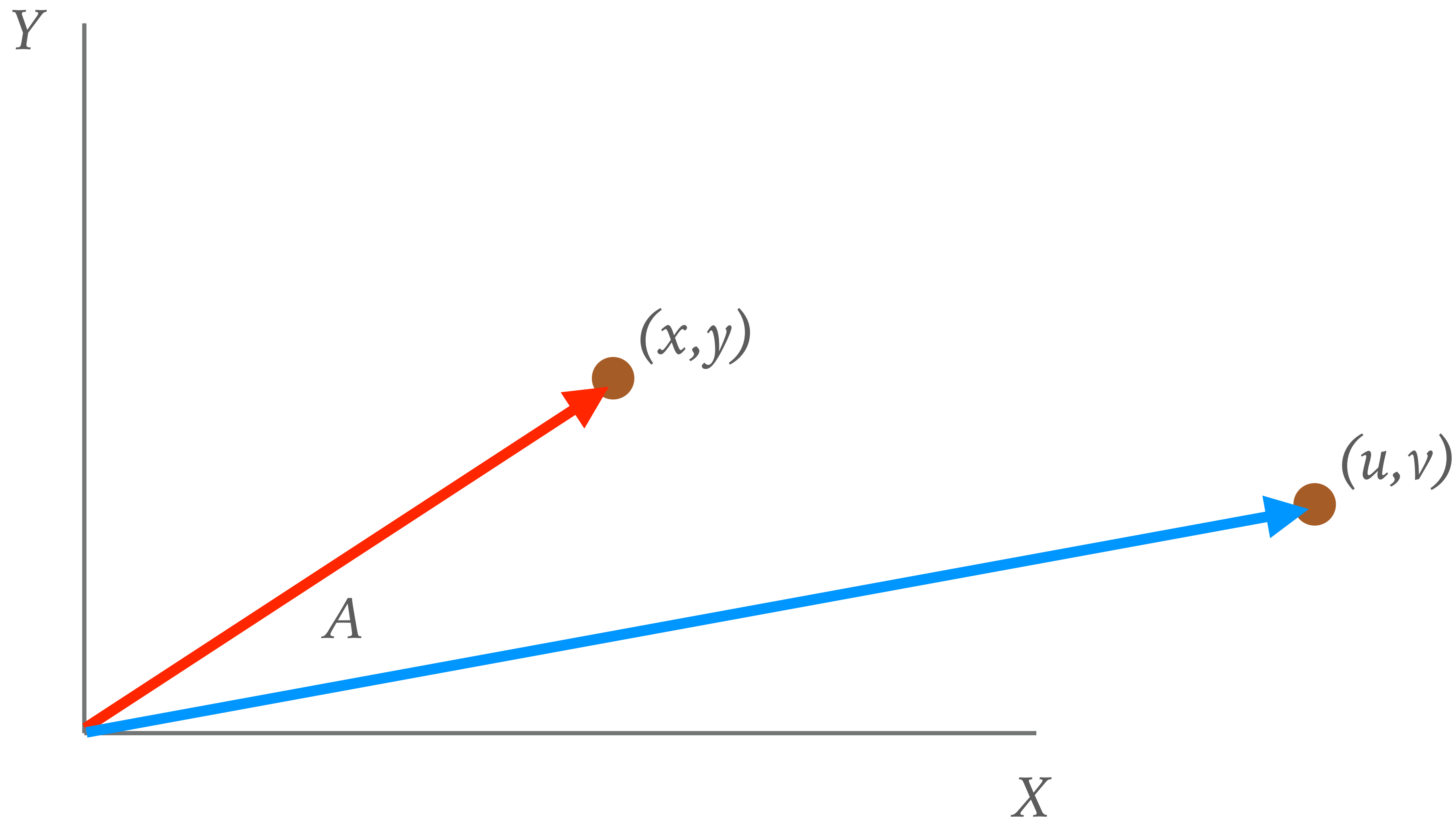
ANGLE



ANGLE

Dot Product

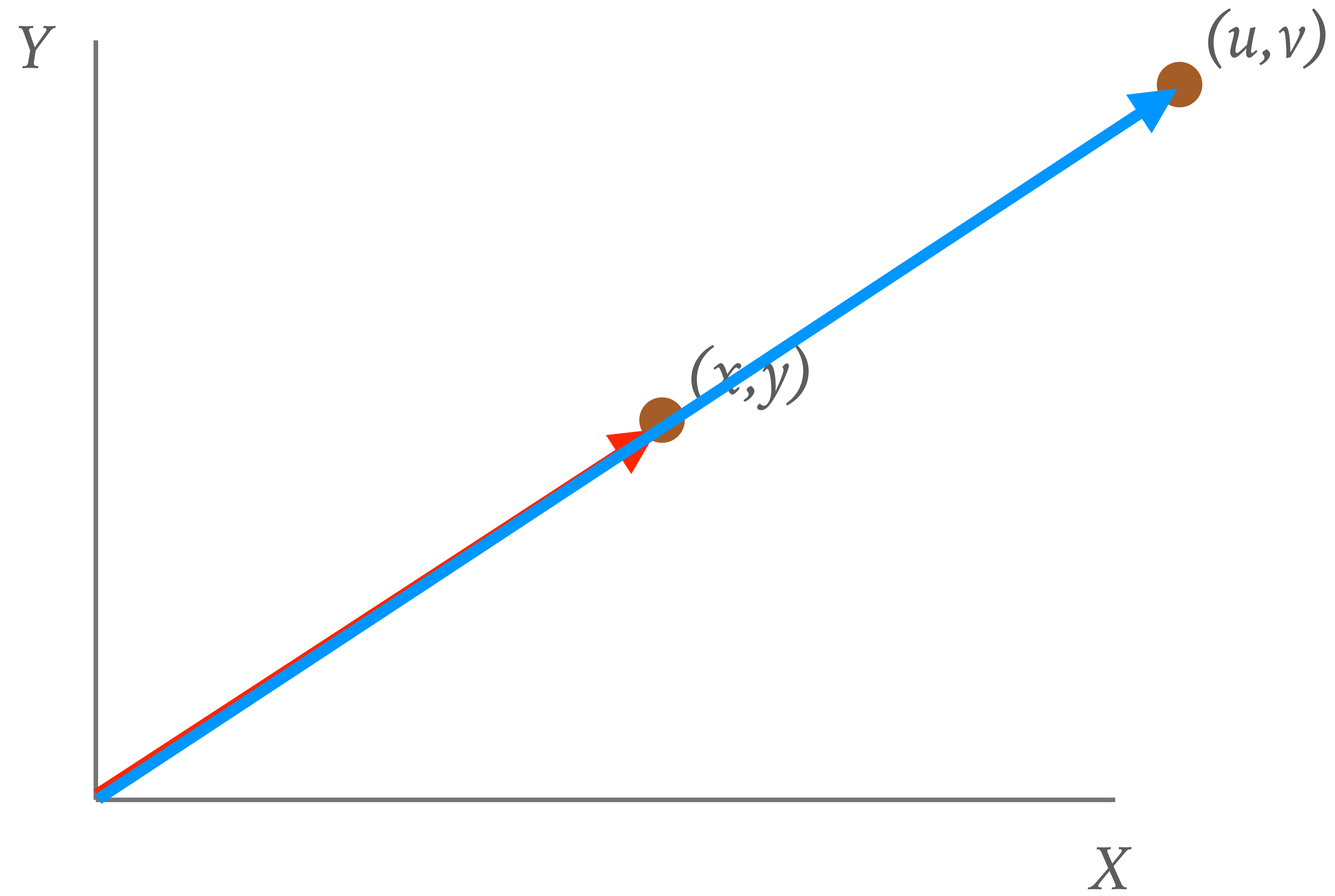
$$(x,y).(u,v) = |(x,y)|\ |(u,v)| \cos A$$



ANGLE

Dot Product

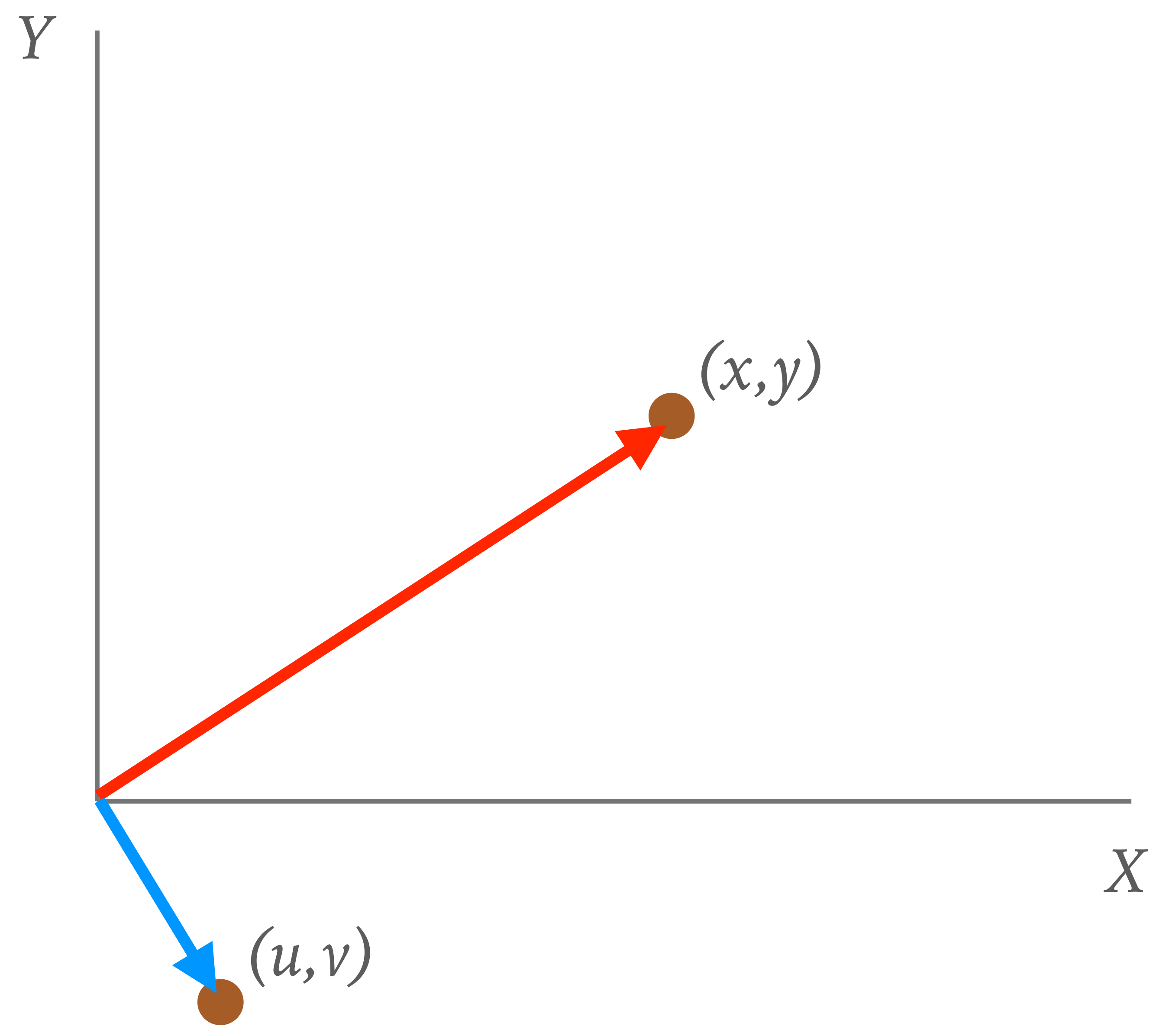
$$(x,y).(u,v) = |(x,y)|\ |(u,v)|$$



ANGLE

Dot Product

$(x,y).(u,v) = 0$



SIZE AND ANGLE CHEATSHEET

Notation for dot product : $v \cdot w$ or $v^T w = \sum_{i=1}^{n} v_i w_i$

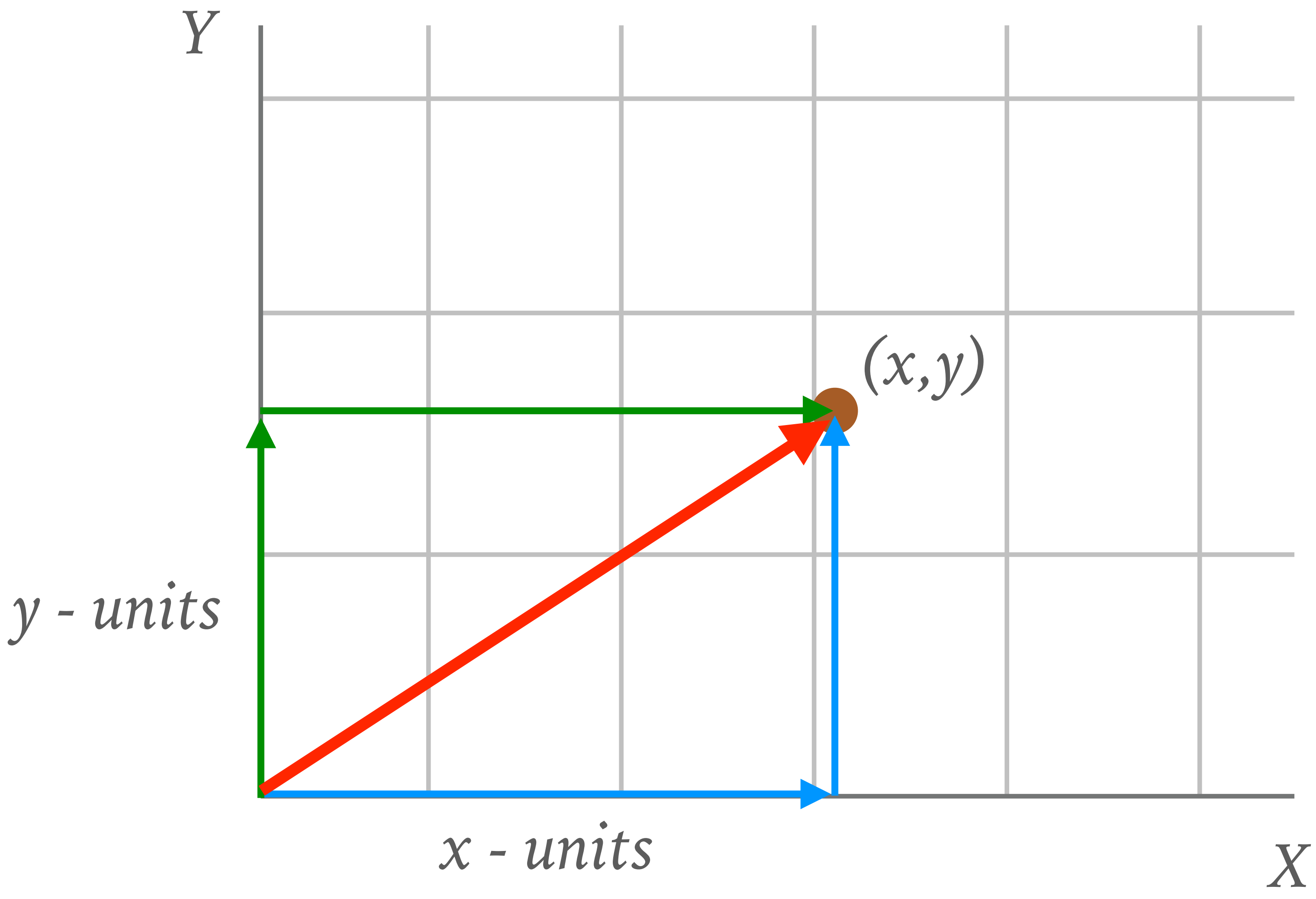
L2 norm of a vector : $v^T v = \sum_{i=1}^{n} v_i^2 = ||v||_2^2$

Distance between two vectors : $||v - w||_2 = \sqrt{\sum_{i=1}^{n} (v_i - w_i)^2}$

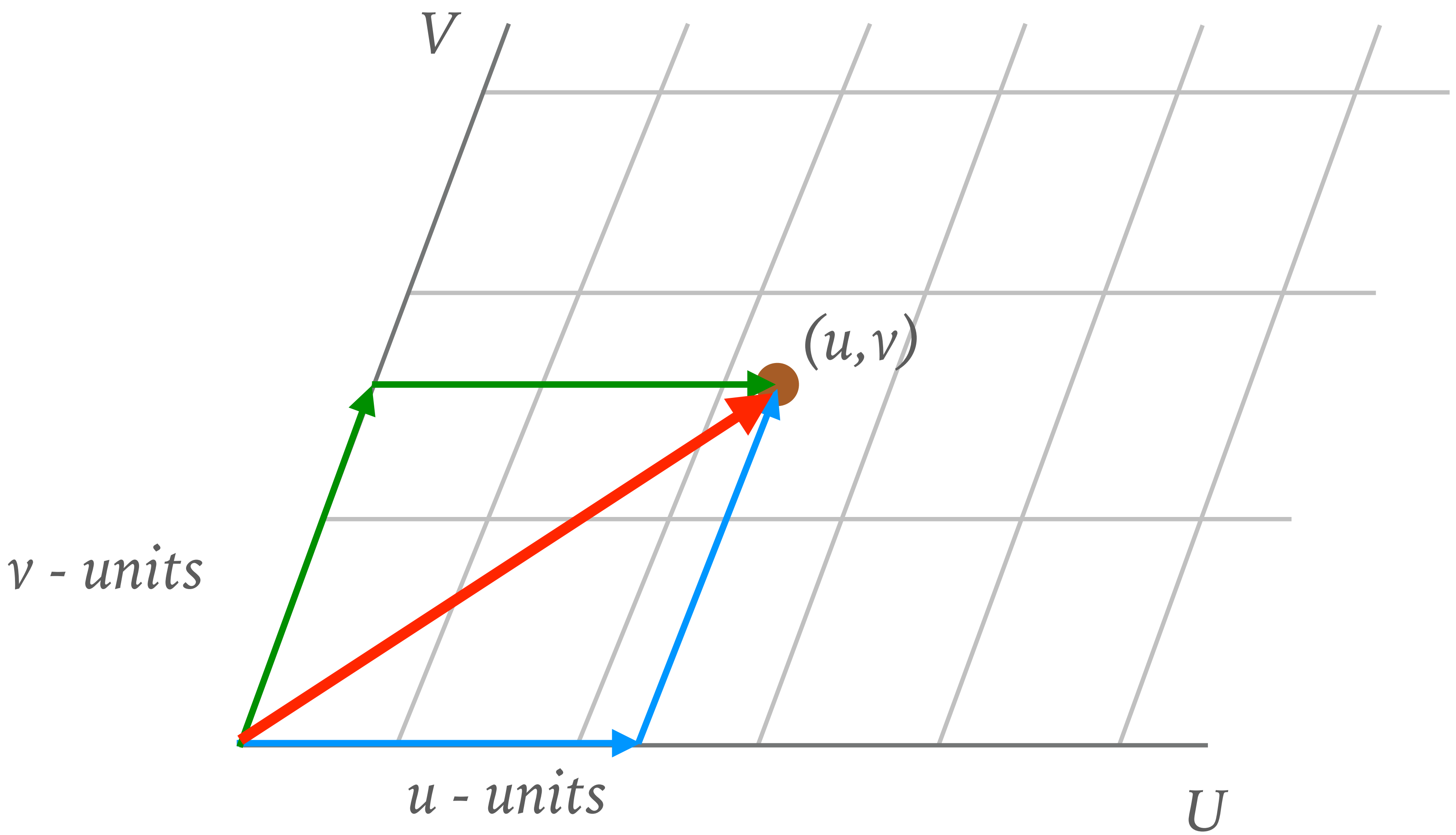
Do you see any connection with Statistics?

ANOTHER LOOK AT 2D

2D GEOMETRY



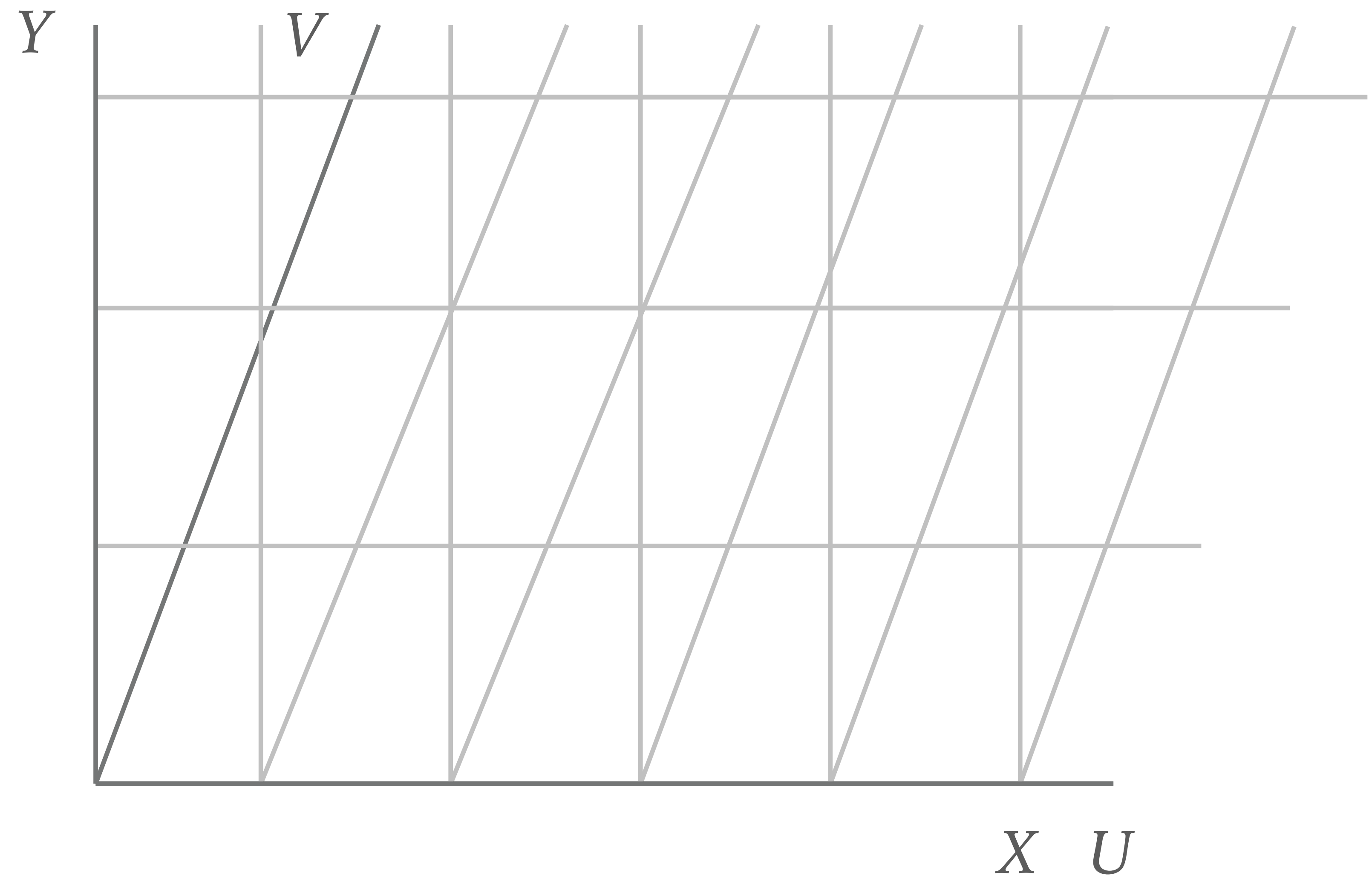
2D GEOMETRY



2D GEOMETRY

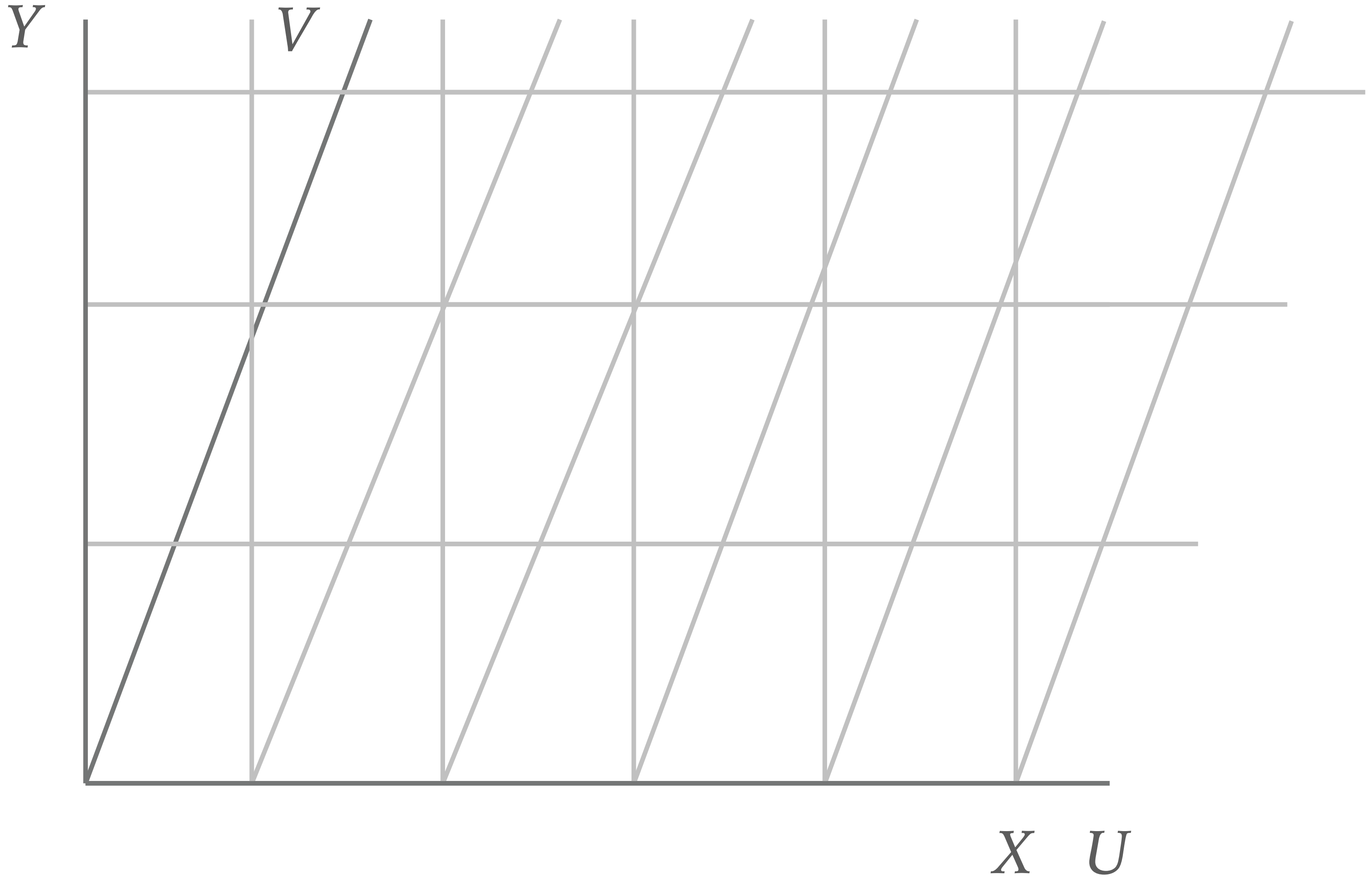
$$U = aX + bY$$

$$V = cX + dY$$



2D GEOMETRY

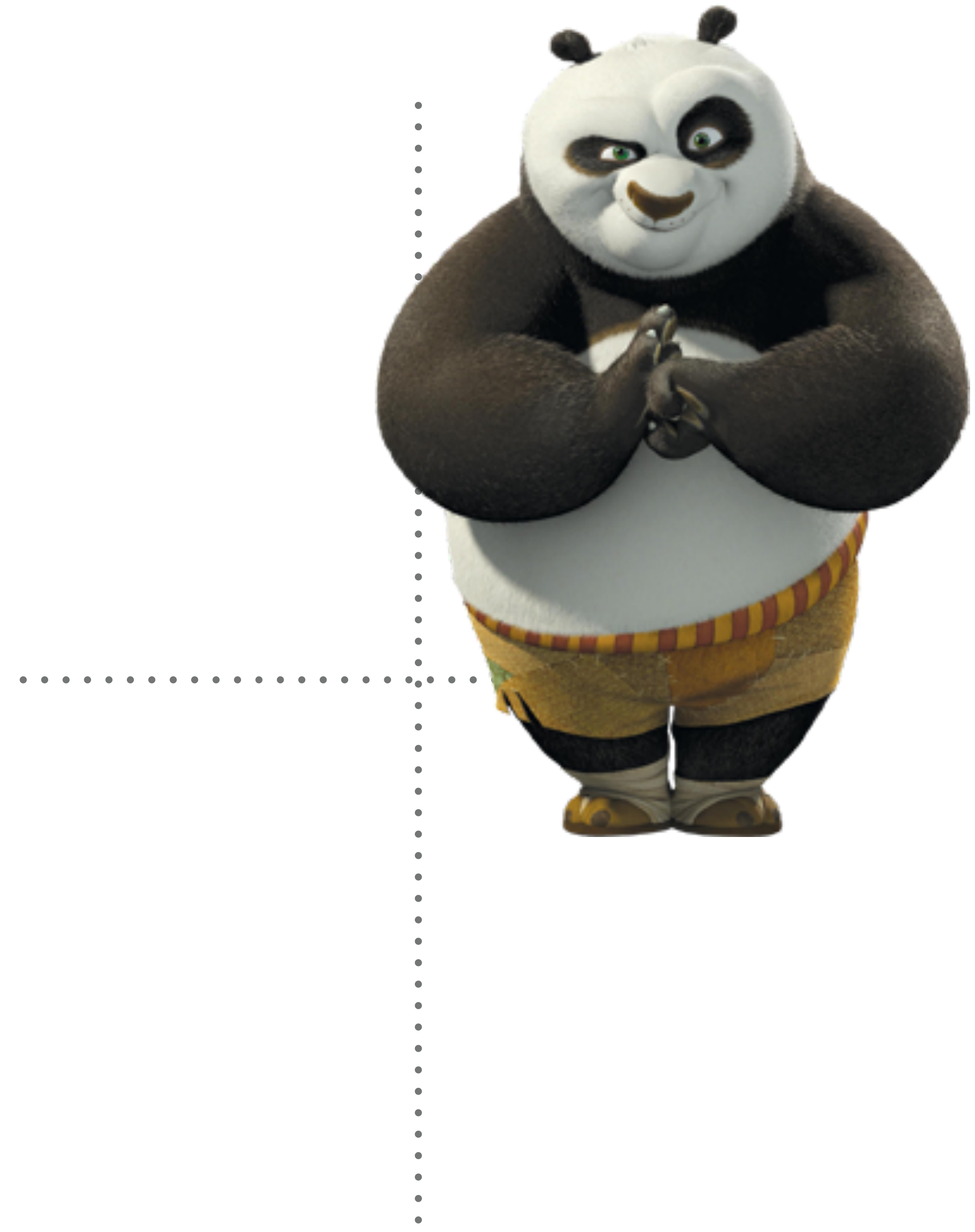
$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$



TRANSFORMATIONS

TRANSLATION

$$\begin{bmatrix} p \\ q \end{bmatrix}$$



ROTATION

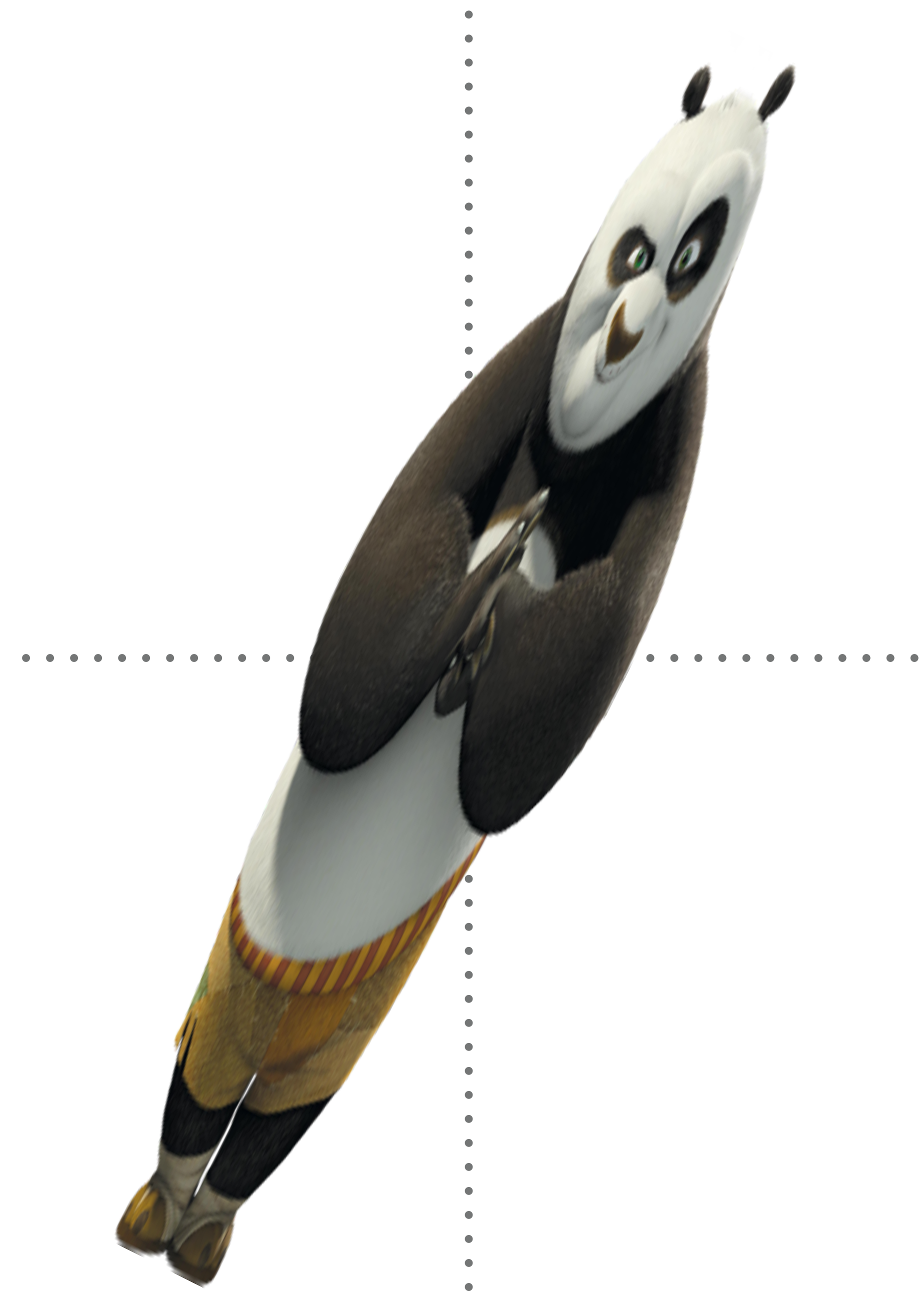
$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

SCALING

$$\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$$

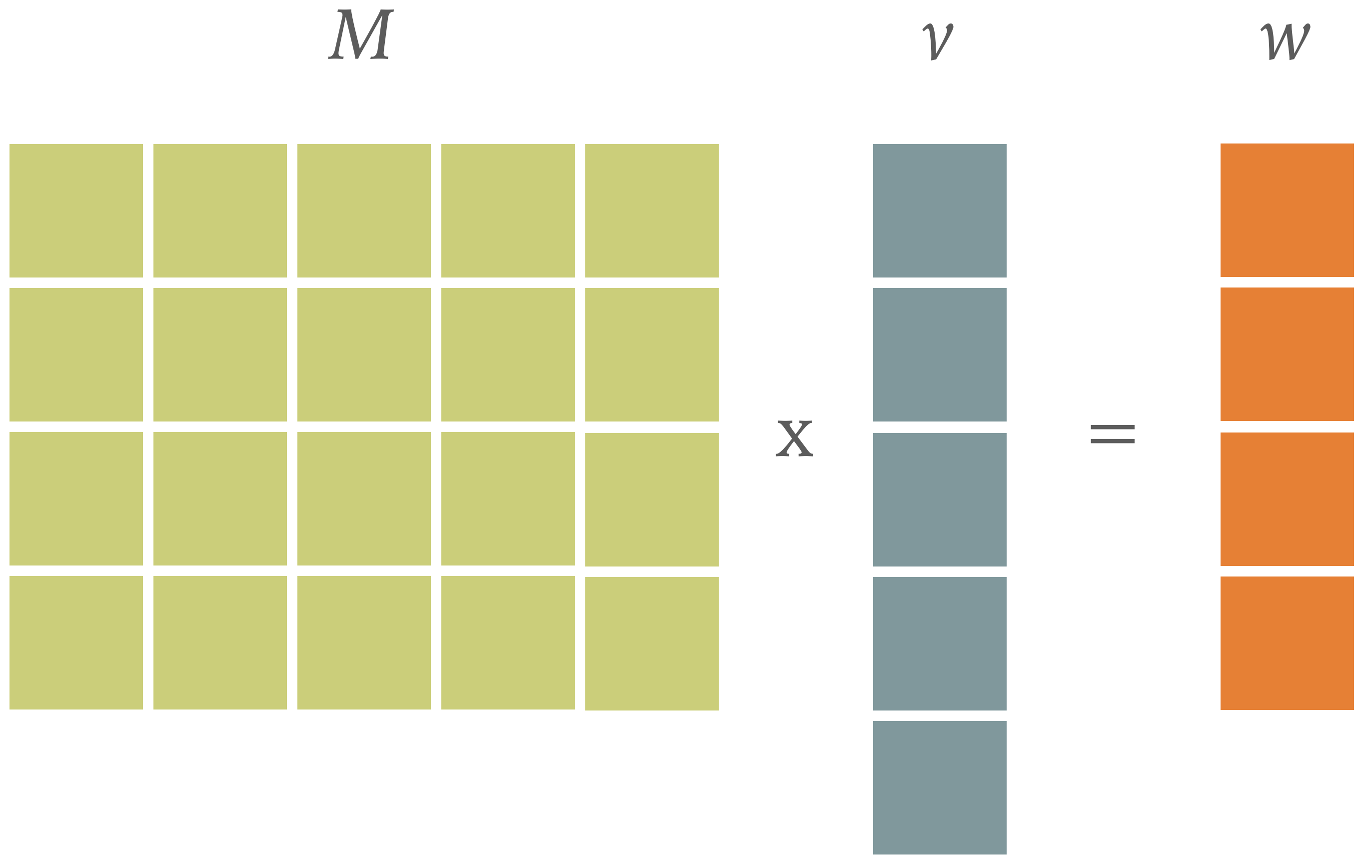
GENERIC

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

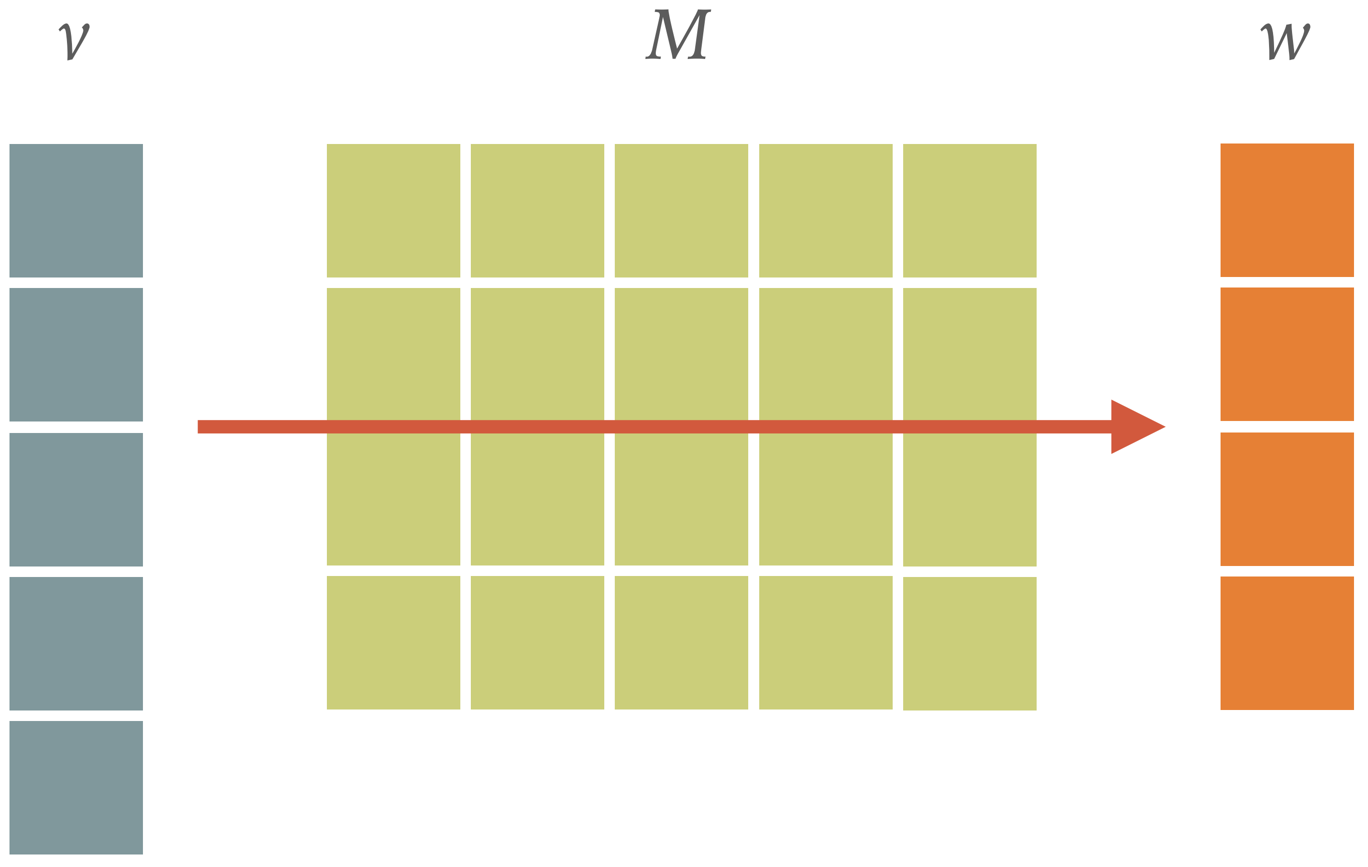


ACTION OF A MATRIX

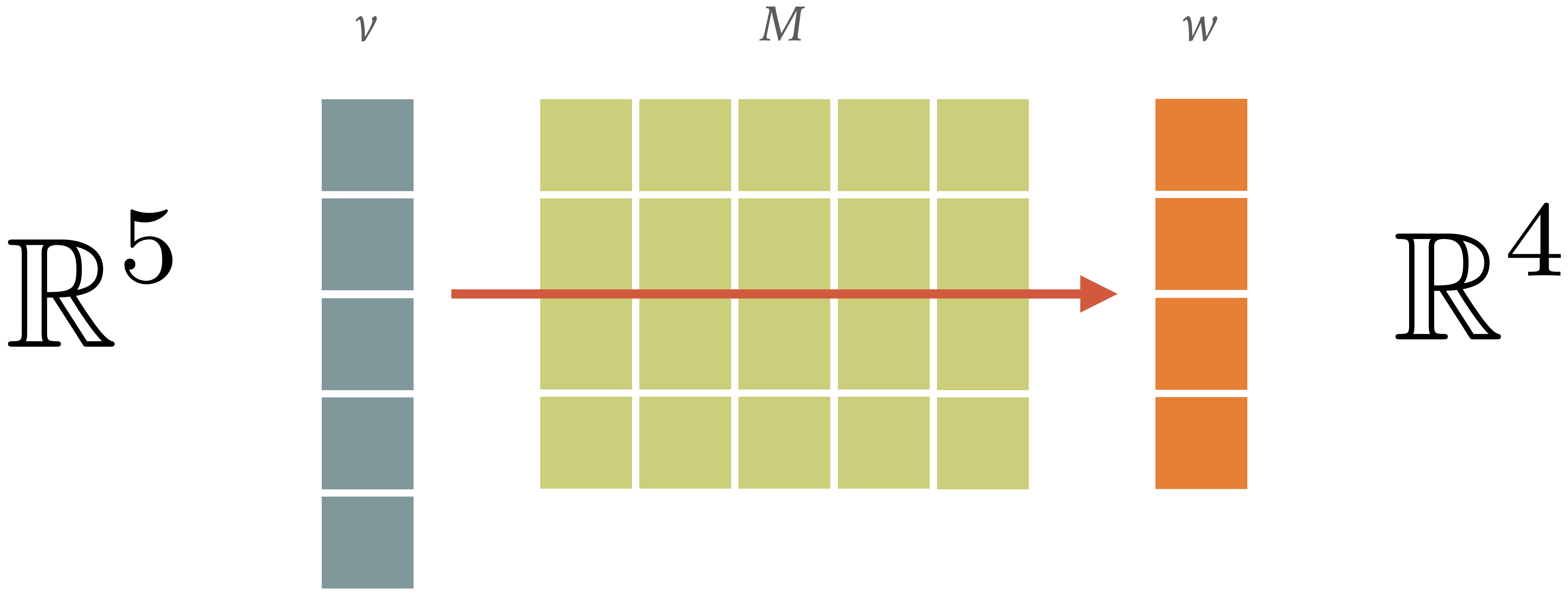
MATRIX-VECTOR MULTIPLICATION



LINEAR MAP



LINEAR MAP



$$\mathbb{R}^n \xrightarrow{M_{m \times n}} \mathbb{R}^m$$

Domain of the function? — Where does the input originate from?

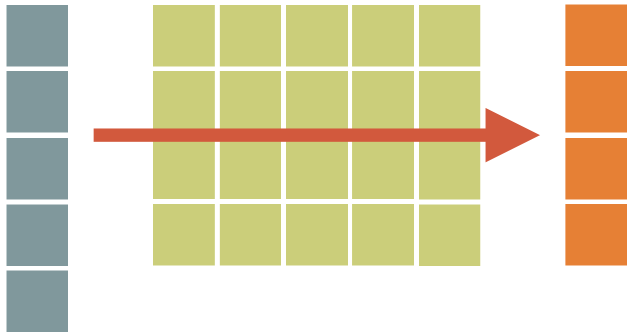
Codomain of the function? — What are possible values of output?

Range of the function? — What are the actual values of output?

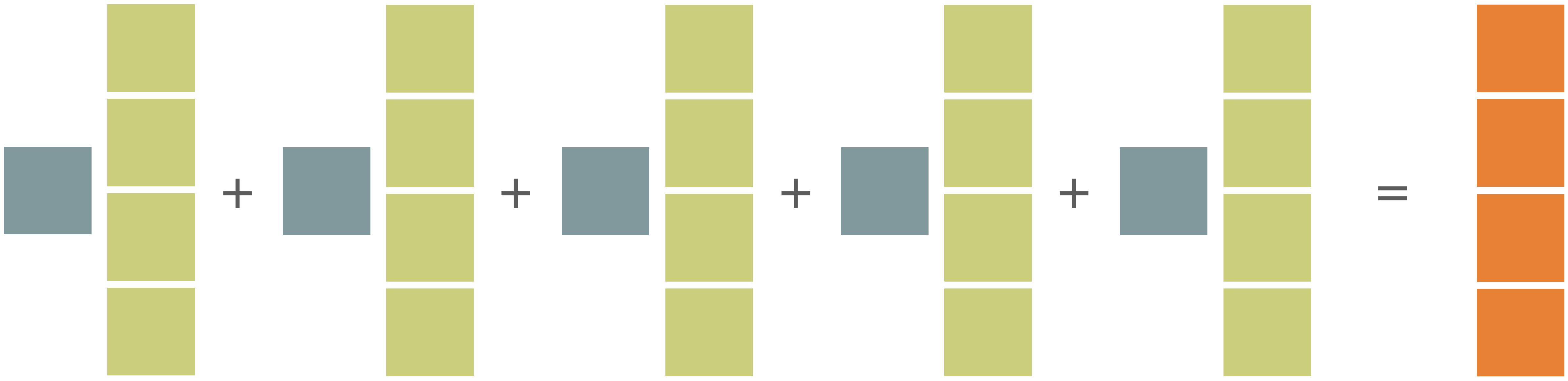
Kernel of the function? — Which inputs map to zero in the range?

SUBSPACES

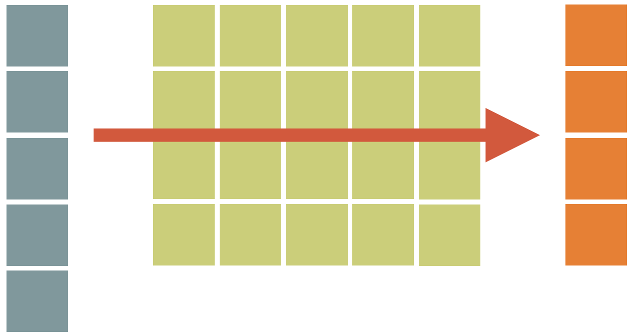
RANGE



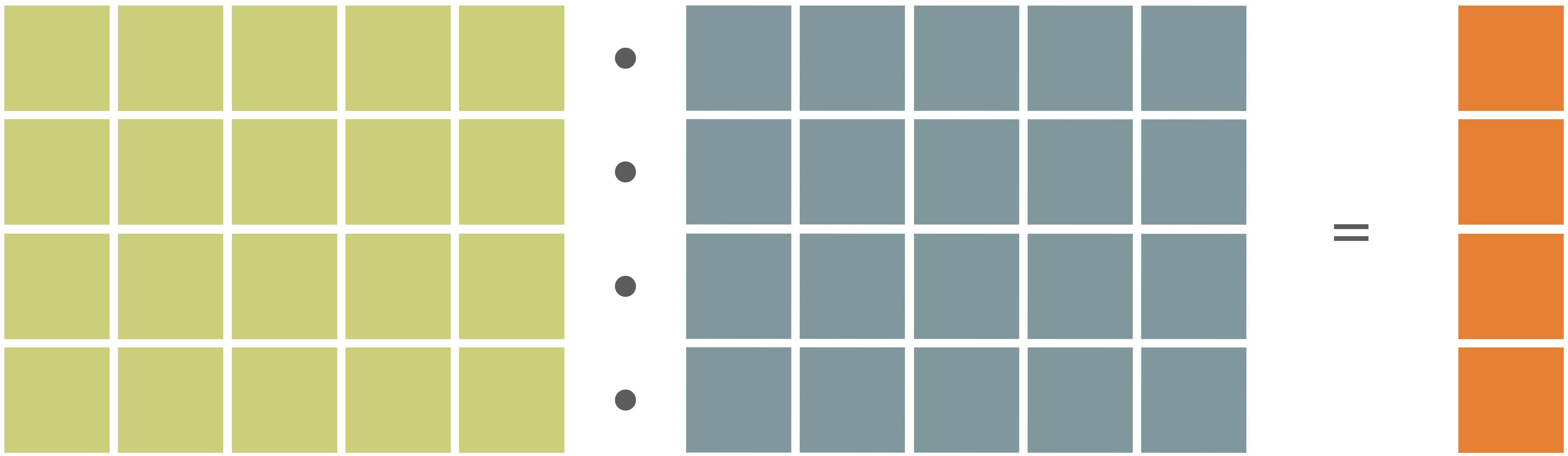
Must be the linear combination of Columns



KERNEL



Must be the orthogonal to all Rows

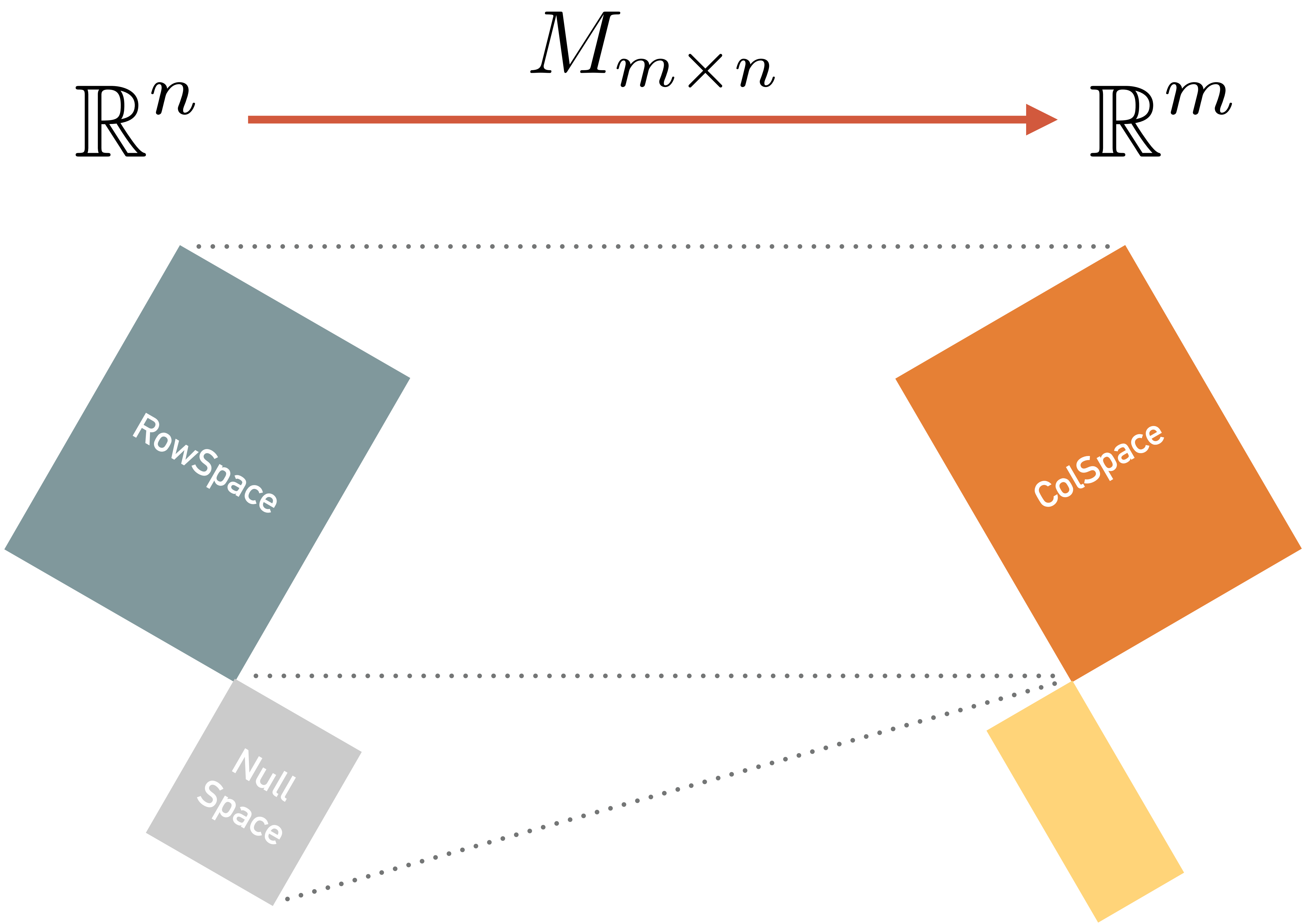


$$\mathbb{R}^n \xrightarrow{M_{m\times n}} \mathbb{R}^m$$

Column Space of the Matrix = Linear span of all Columns = Range

Row Space of the Matrix = Linear span of Rows = Non-zero output

Null Space of the Matrix = Map to zero = Orthogonal to Row Space

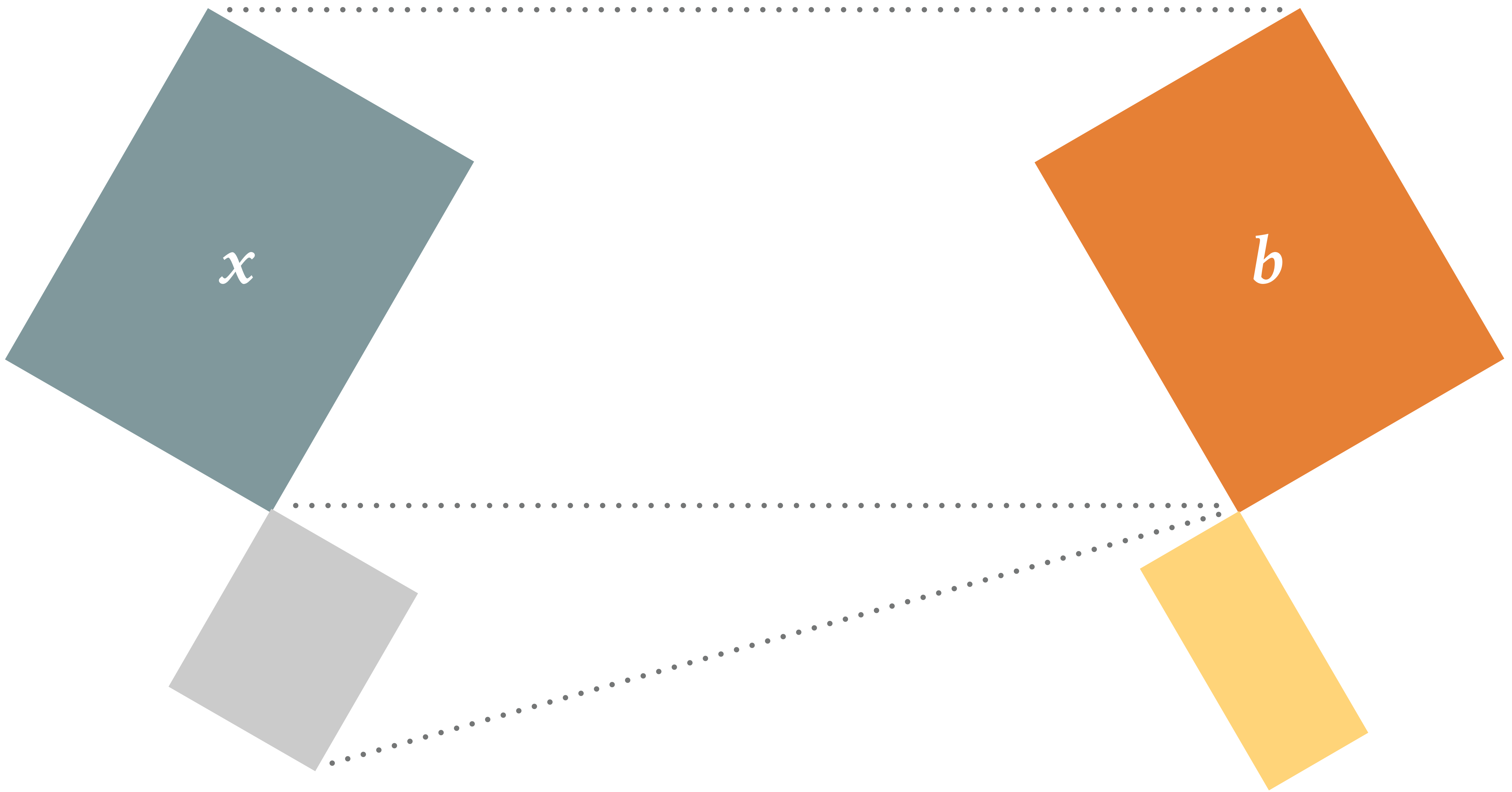


$M_{m \times n}$
$\mathbb{R}^n$
$\mathbb{R}^m$
RowSpace
ColSpace
Null Space

LINEAR EQUATIONS

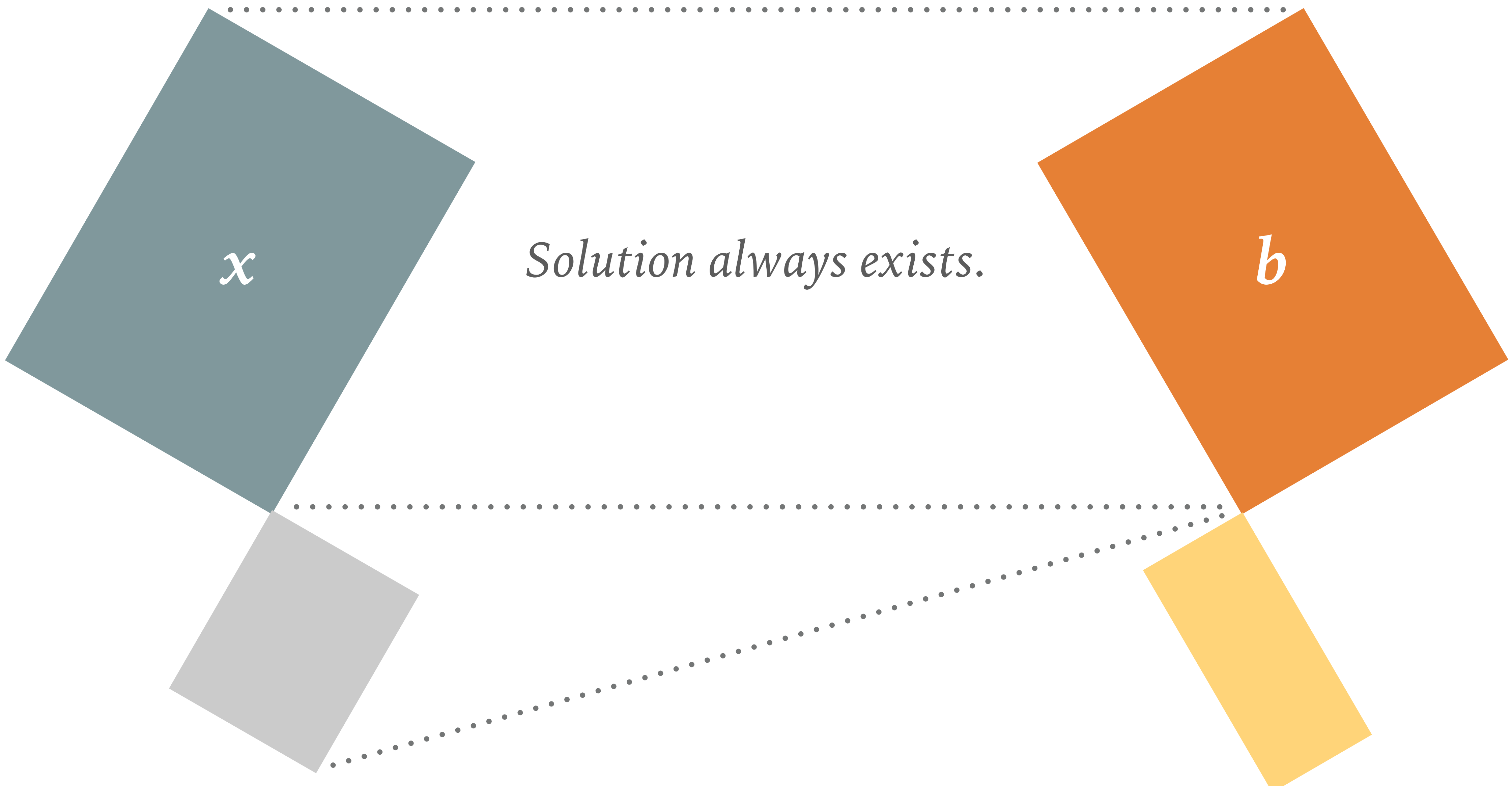
SOLVABILITY OF LINEAR EQUATION

$Mx = b$



SOLVABILITY OF LINEAR EQUATION

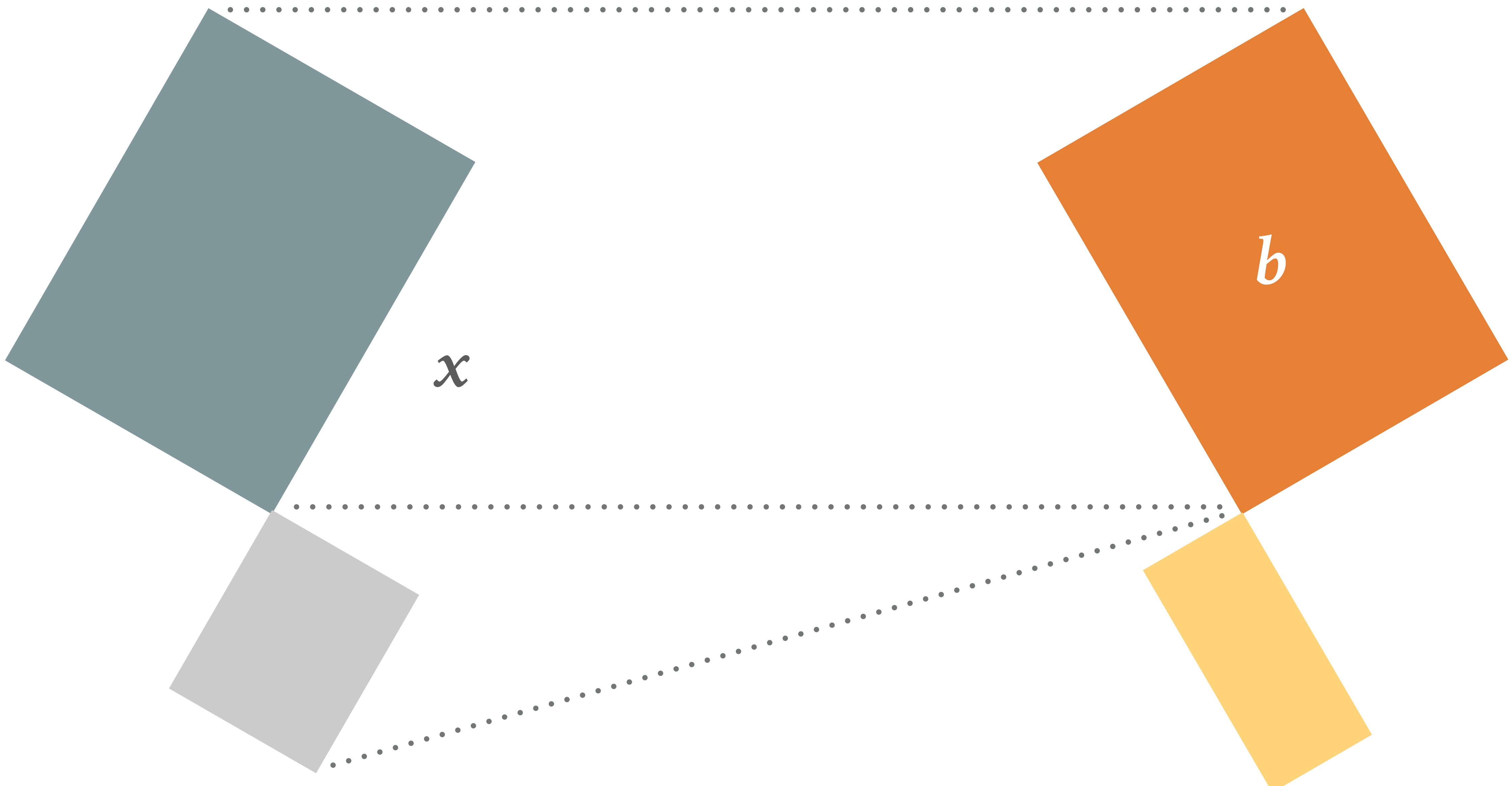
$Mx = b$



Solution is unique as $Mx = b$ and $My = b$ implies $(x\text{-}y) = 0$.

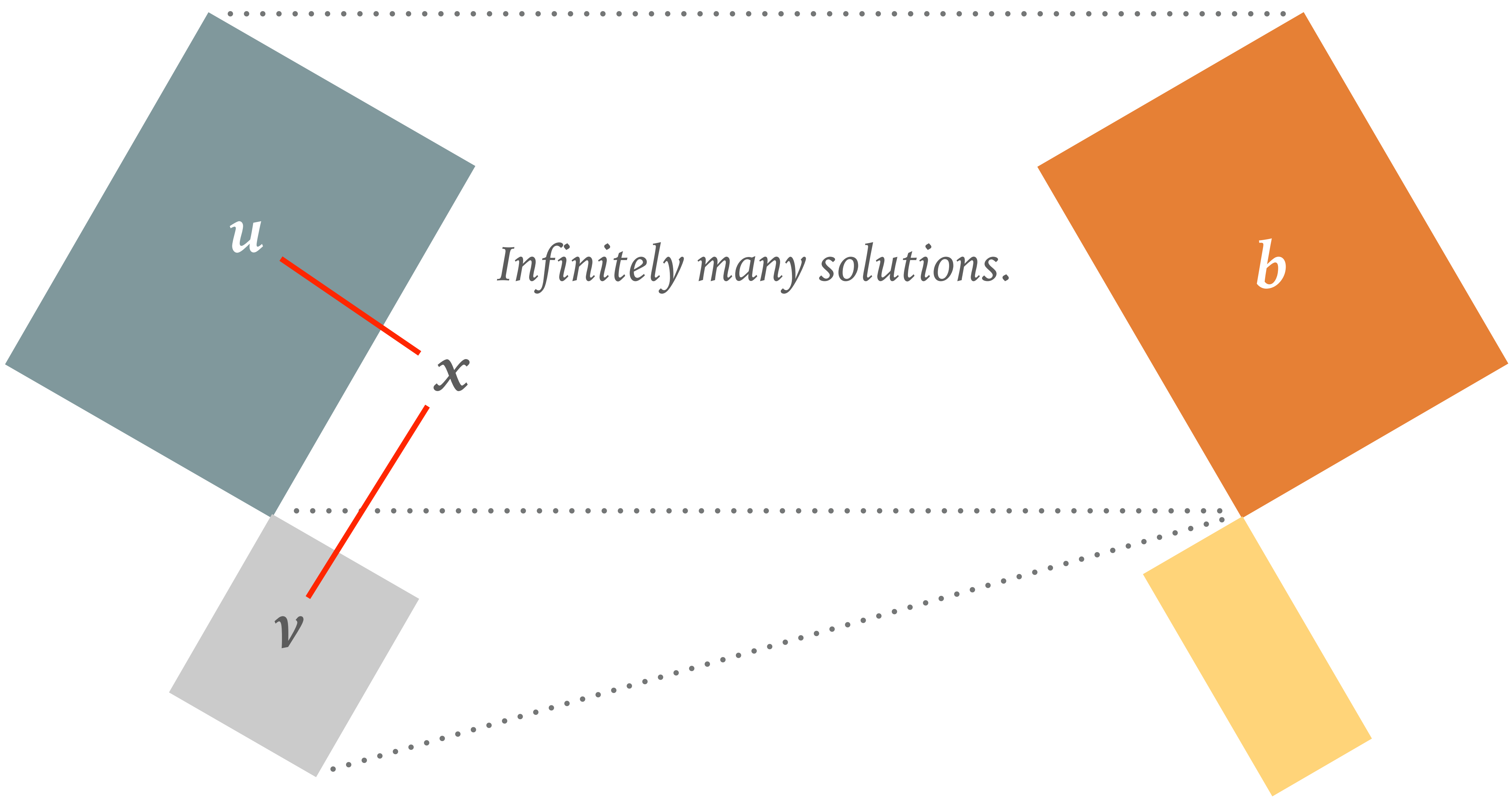
SOLVABILITY OF LINEAR EQUATION

$$Mx = b$$



SOLVABILITY OF LINEAR EQUATION

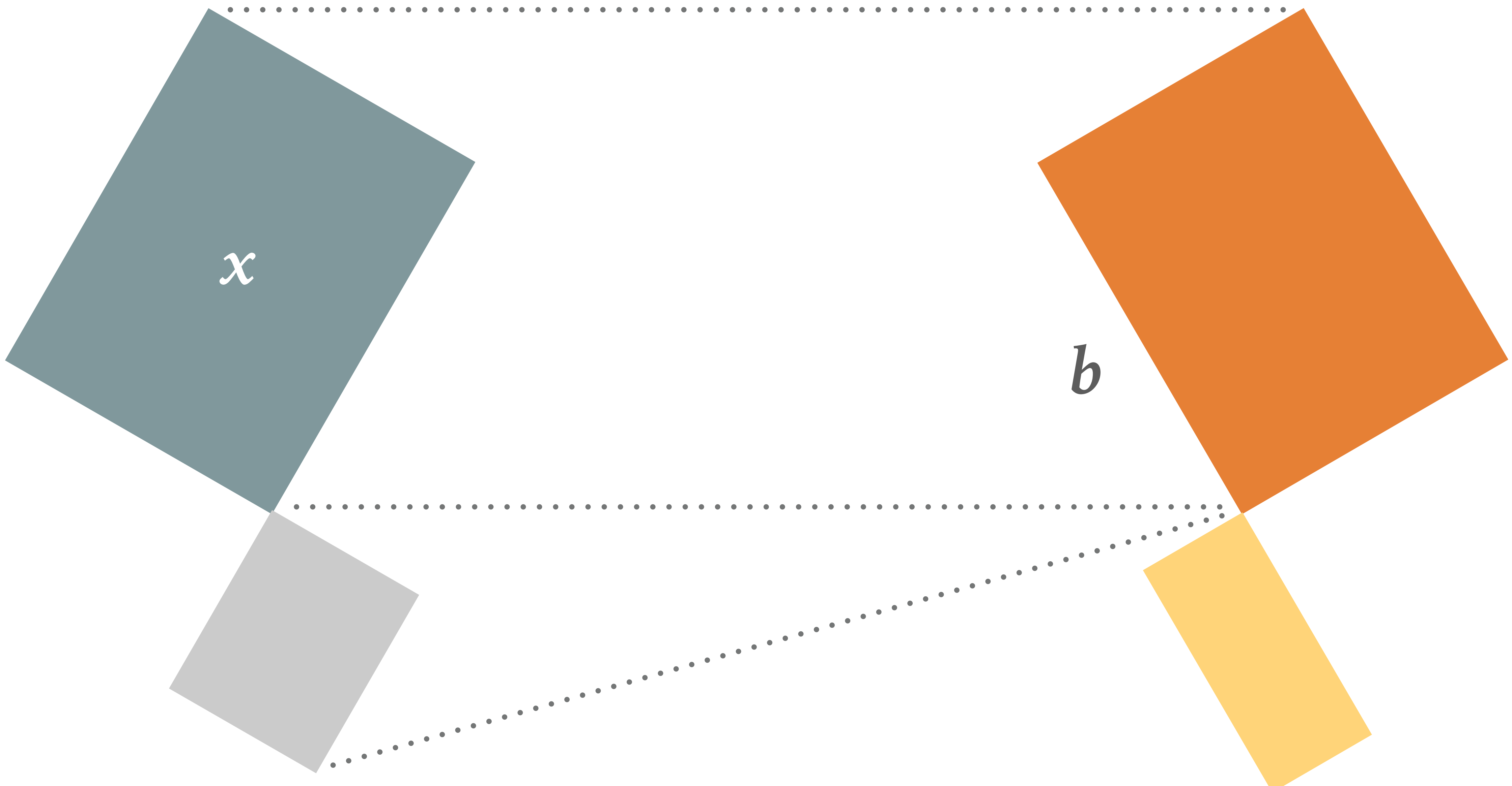
$Mx = b$



If $Mu = b$, *then* $M(u+v) = b$, *for any* v *from Null Space.*

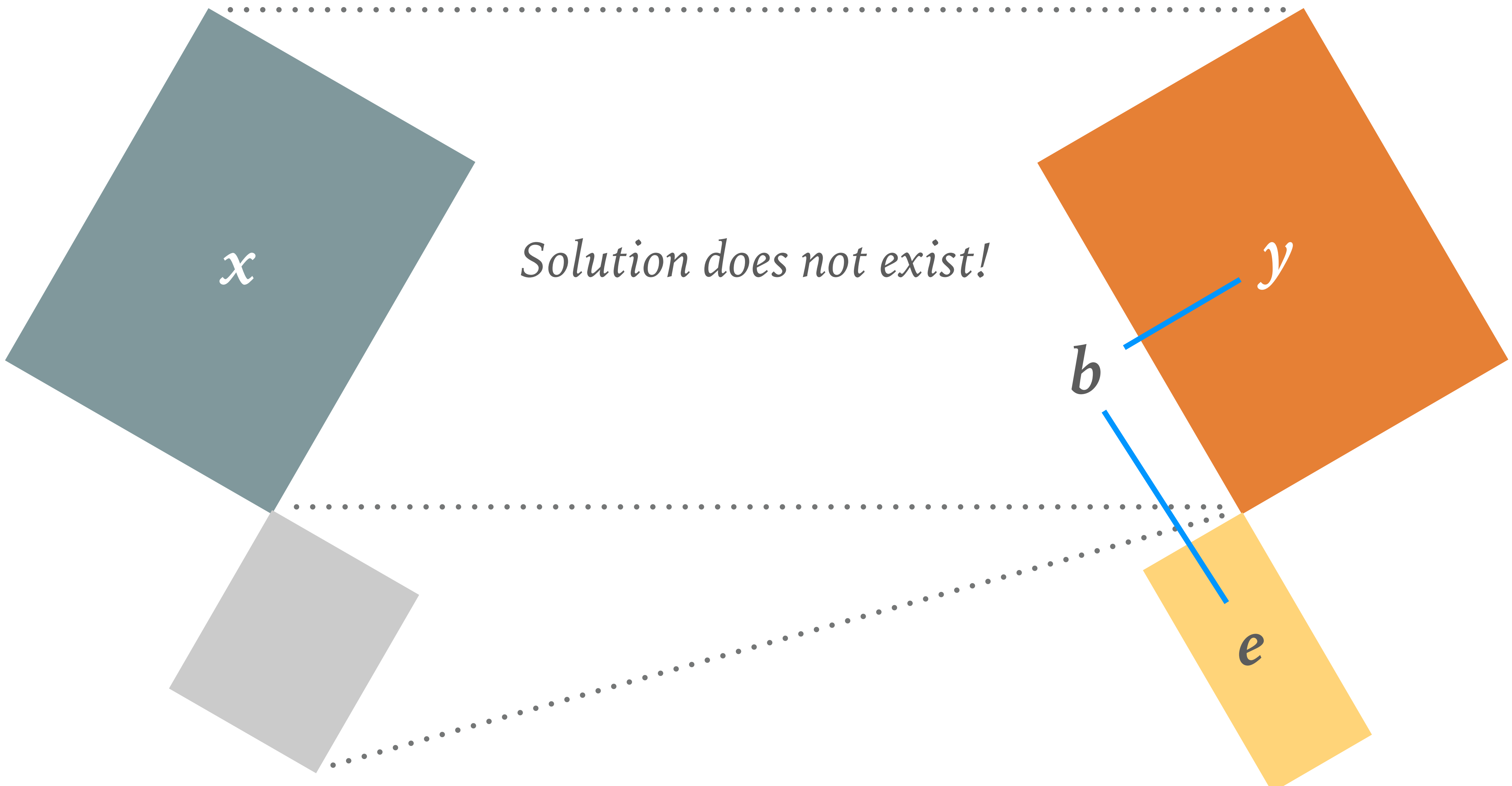
SOLVABILITY OF LINEAR EQUATION

$Mx = b$

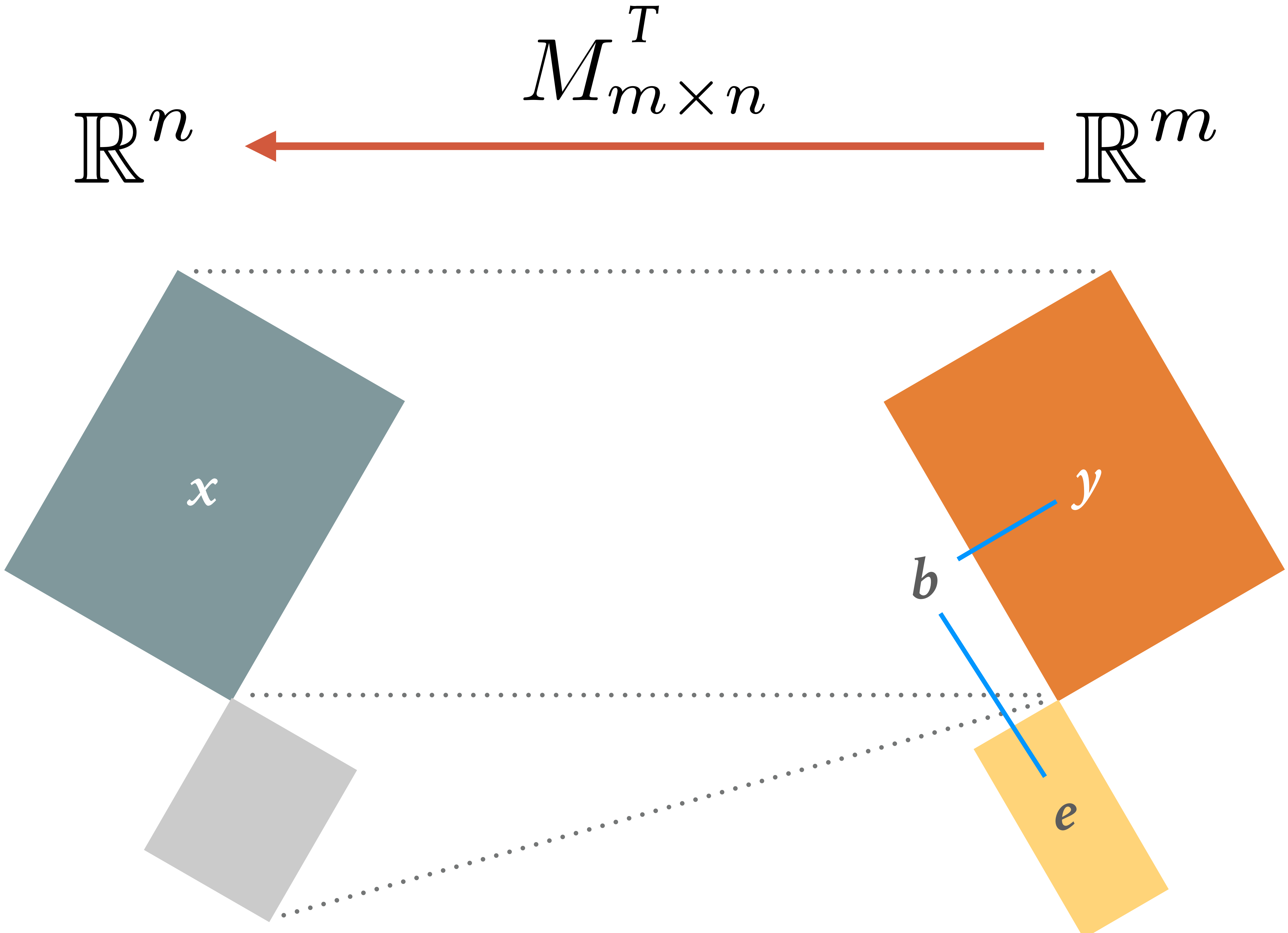


SOLVABILITY OF LINEAR EQUATION

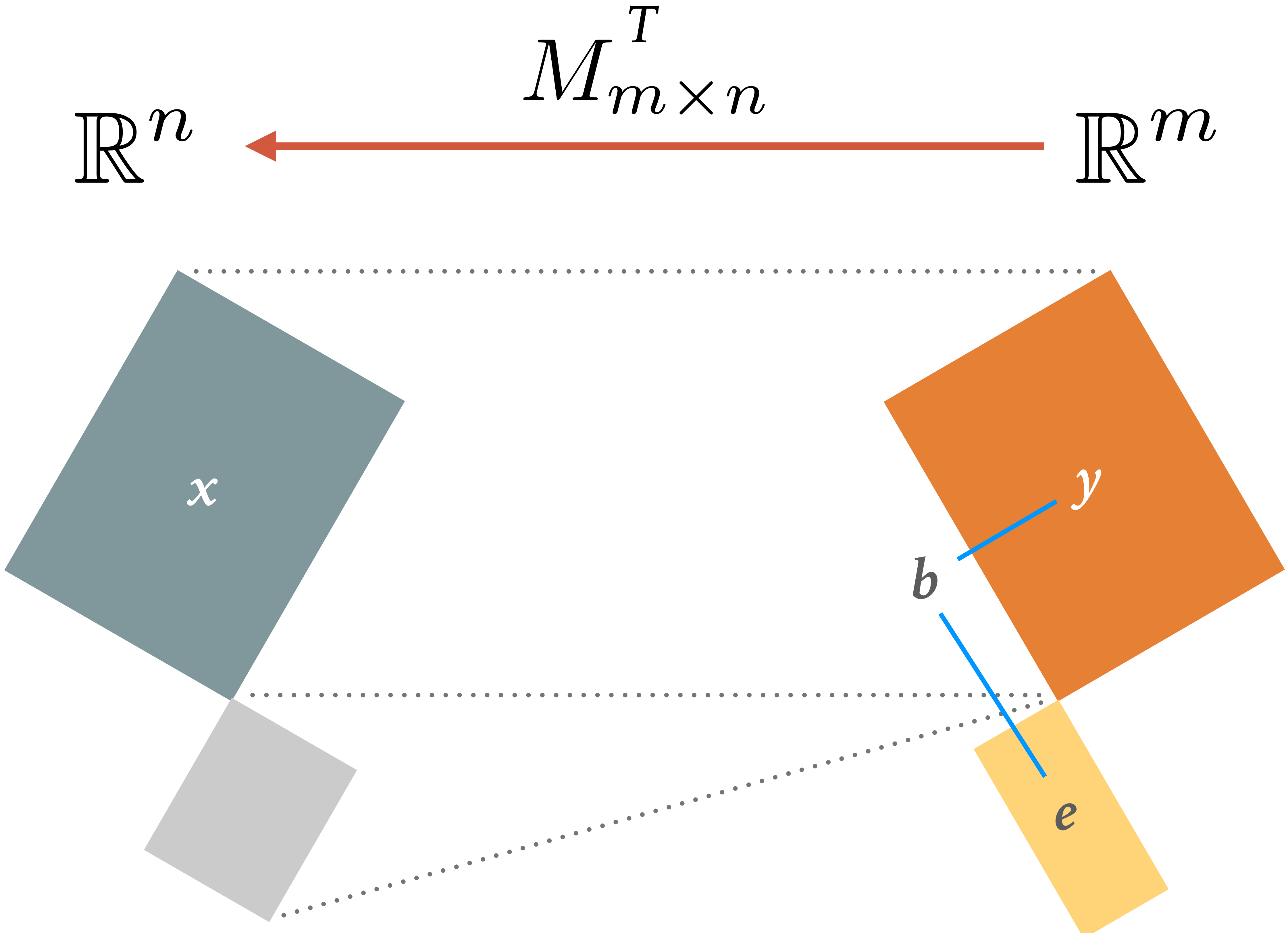
$Mx = b$



If $Mx = y$, *then* $Mx \sim b + e$, *with irreducible error* e.



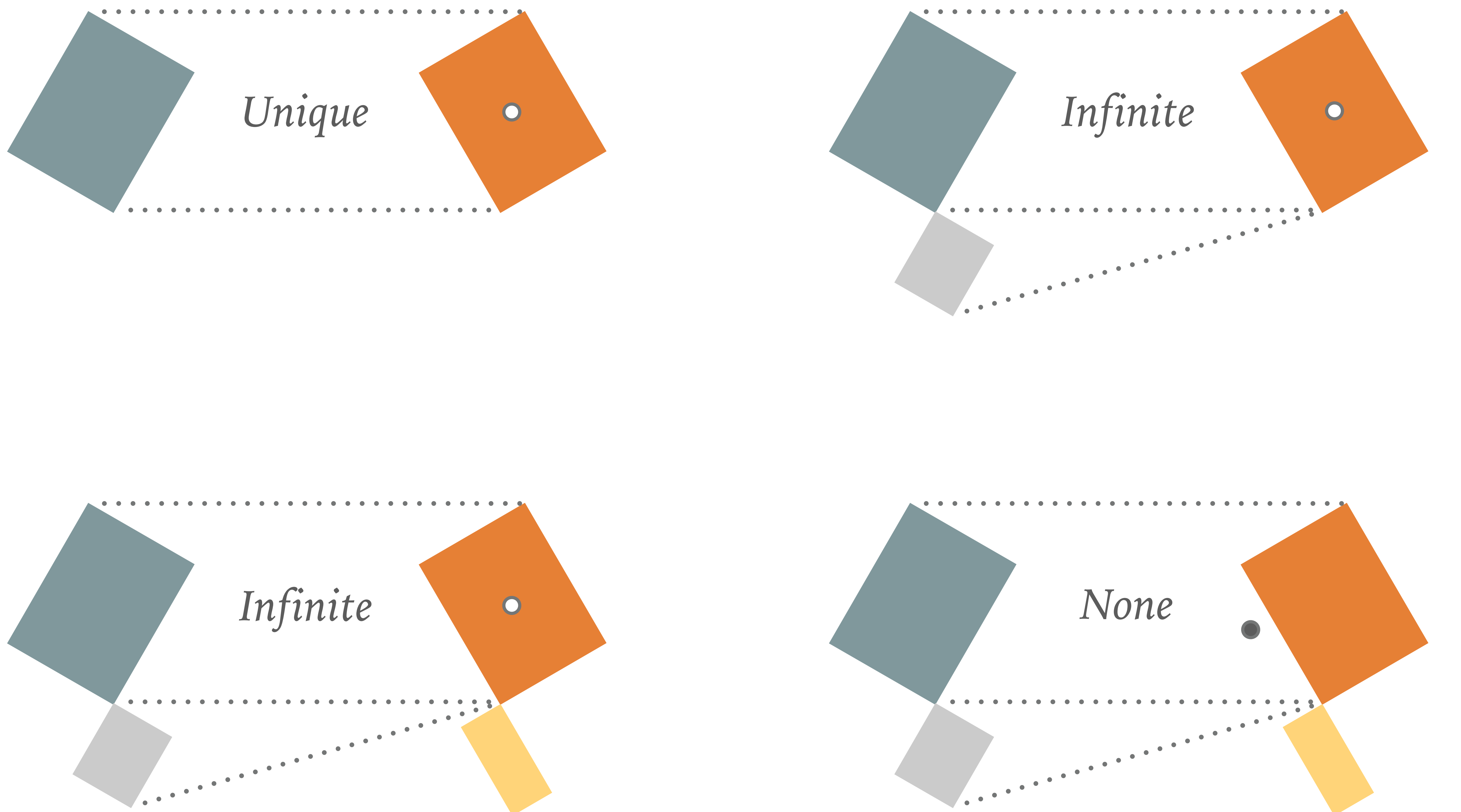
$M_{m\times n}^{T}$
$\mathbb{R}^n$
$\mathbb{R}^m$
x
y
b
e



$M^TMx = M^Tb = M^Ty$ provides solution $x = (M^TM)^{-1}M^Ty$

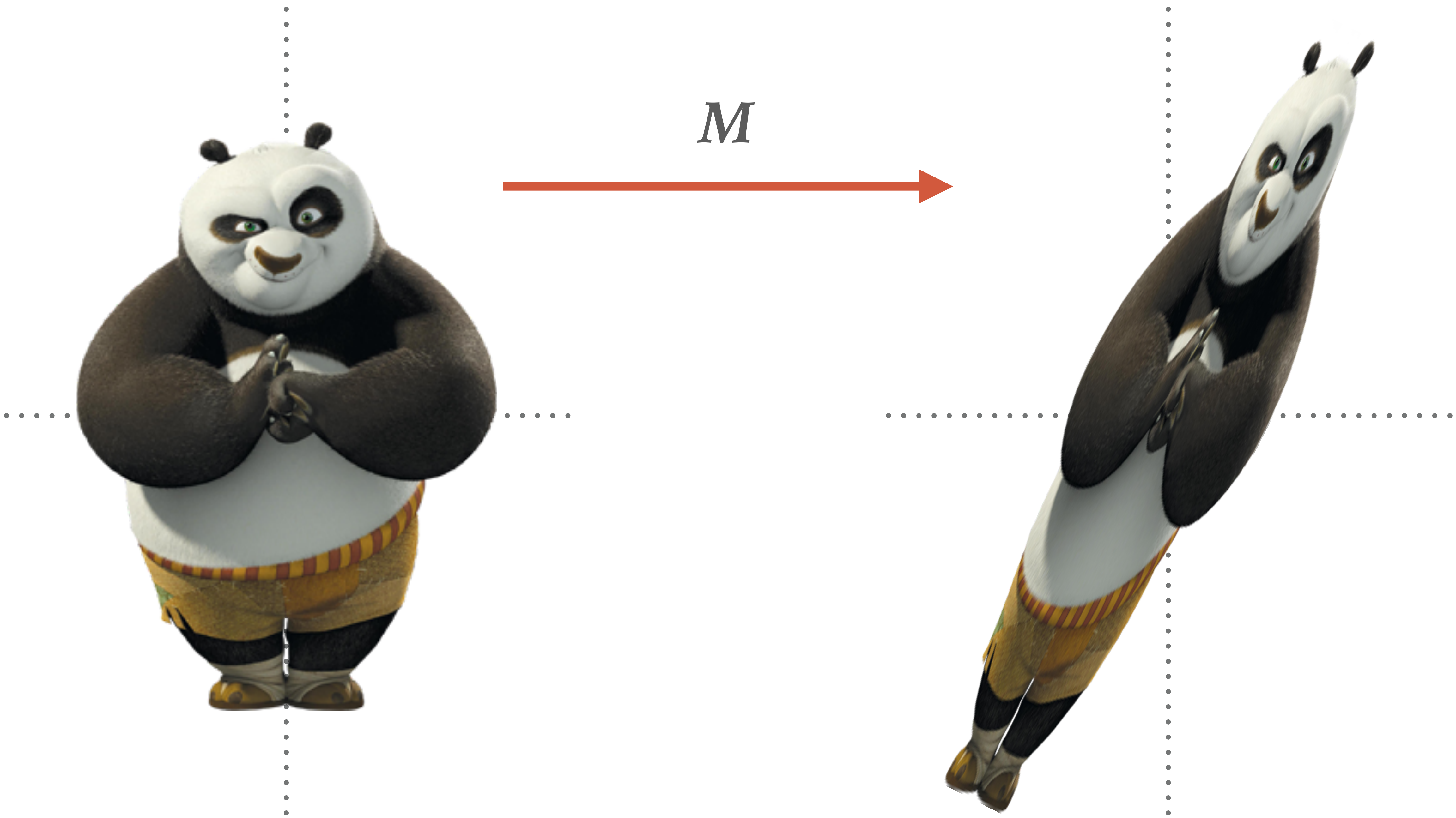
SOLVABILITY OF LINEAR EQUATION

$$Mx = b$$

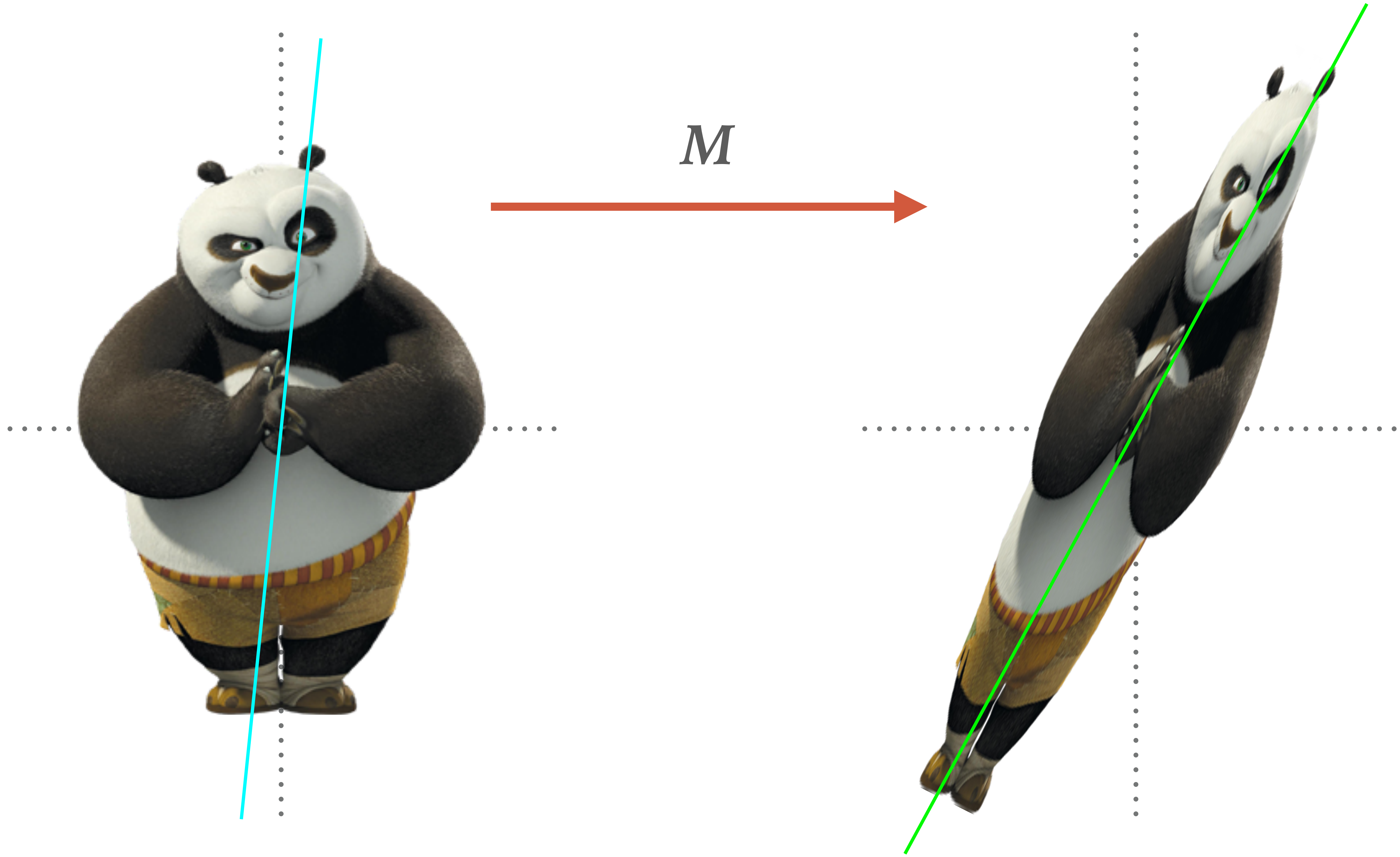


SINGULAR VALUES

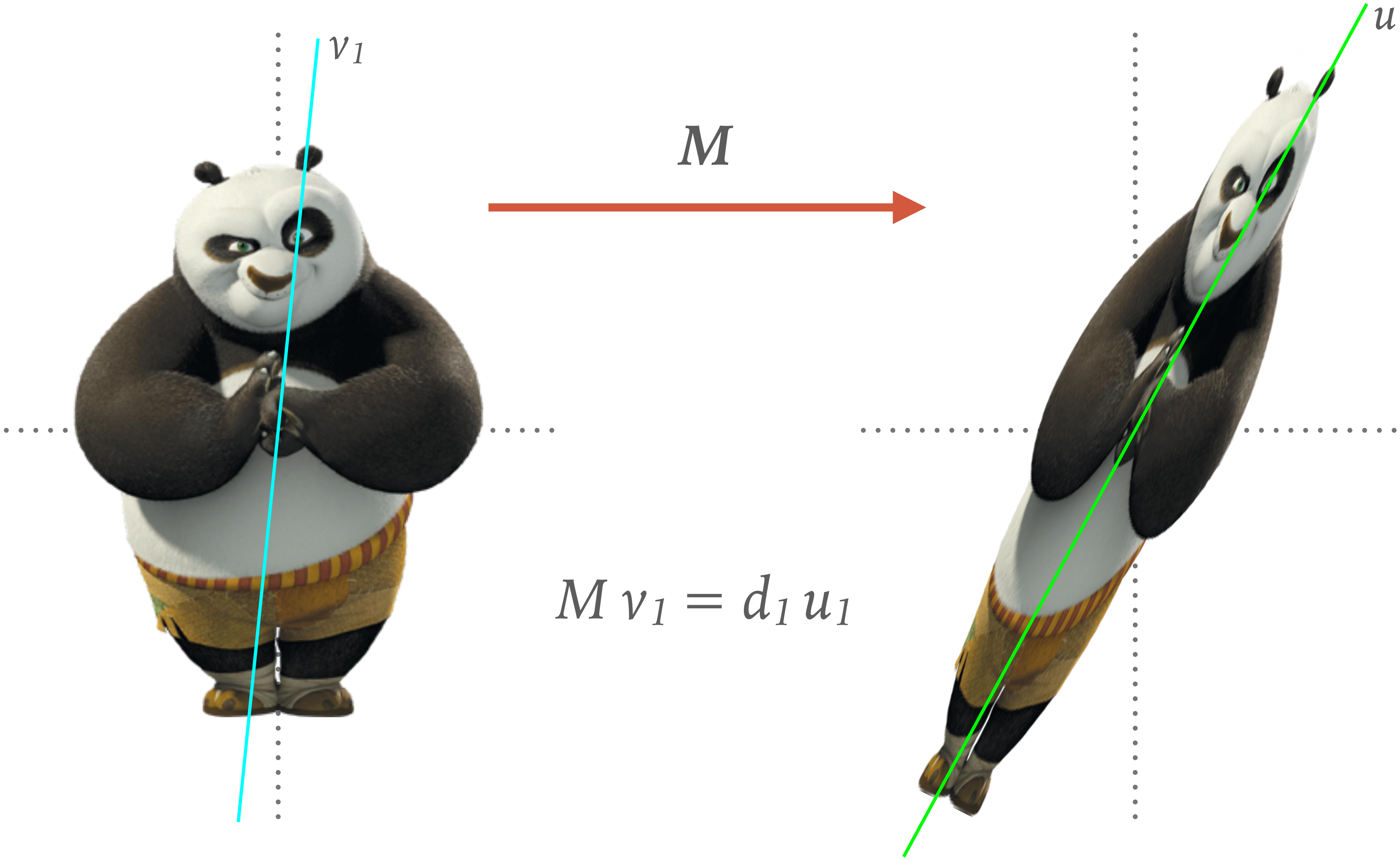
ACTION OF A MATRIX



ACTION OF A MATRIX

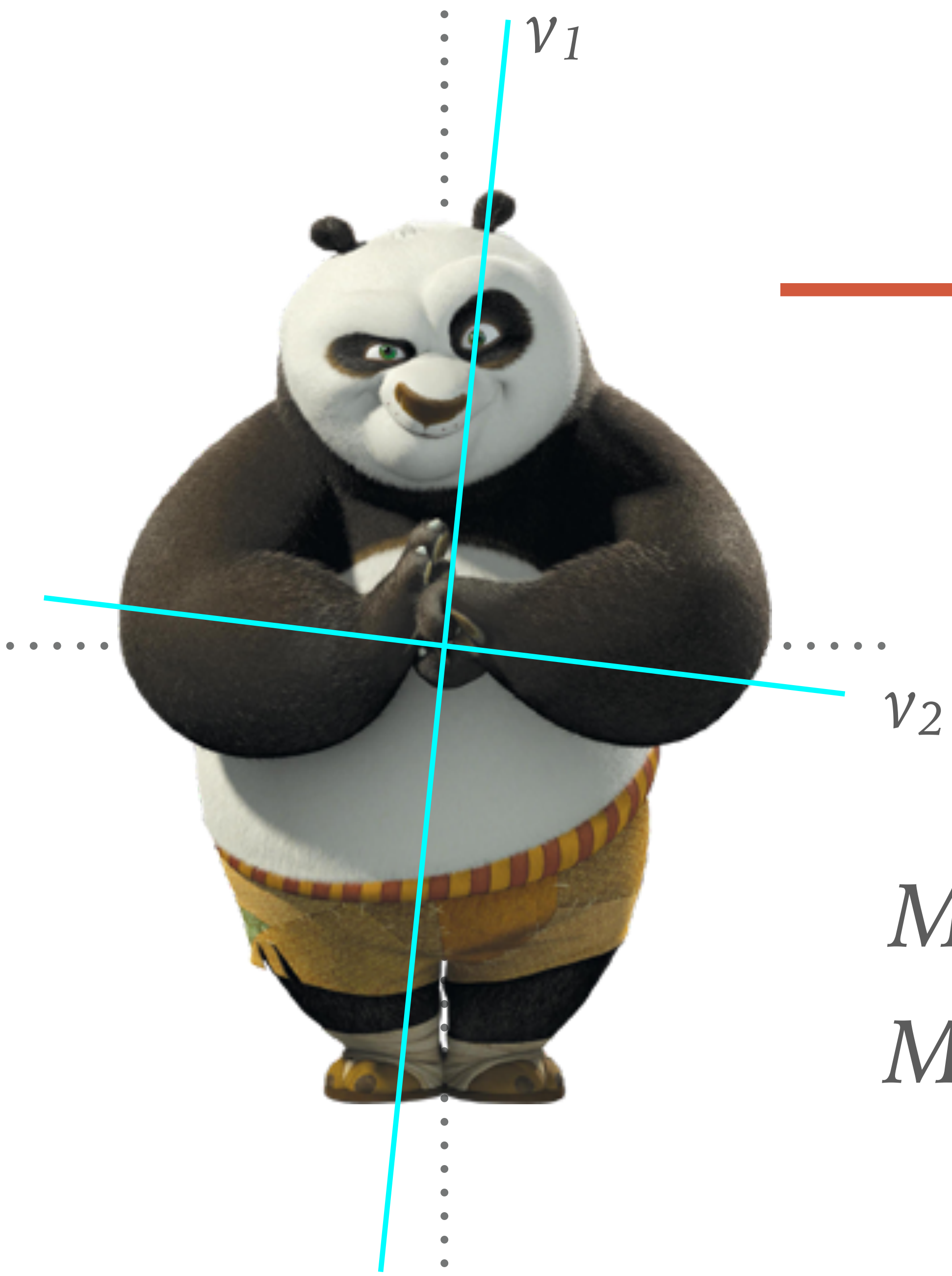


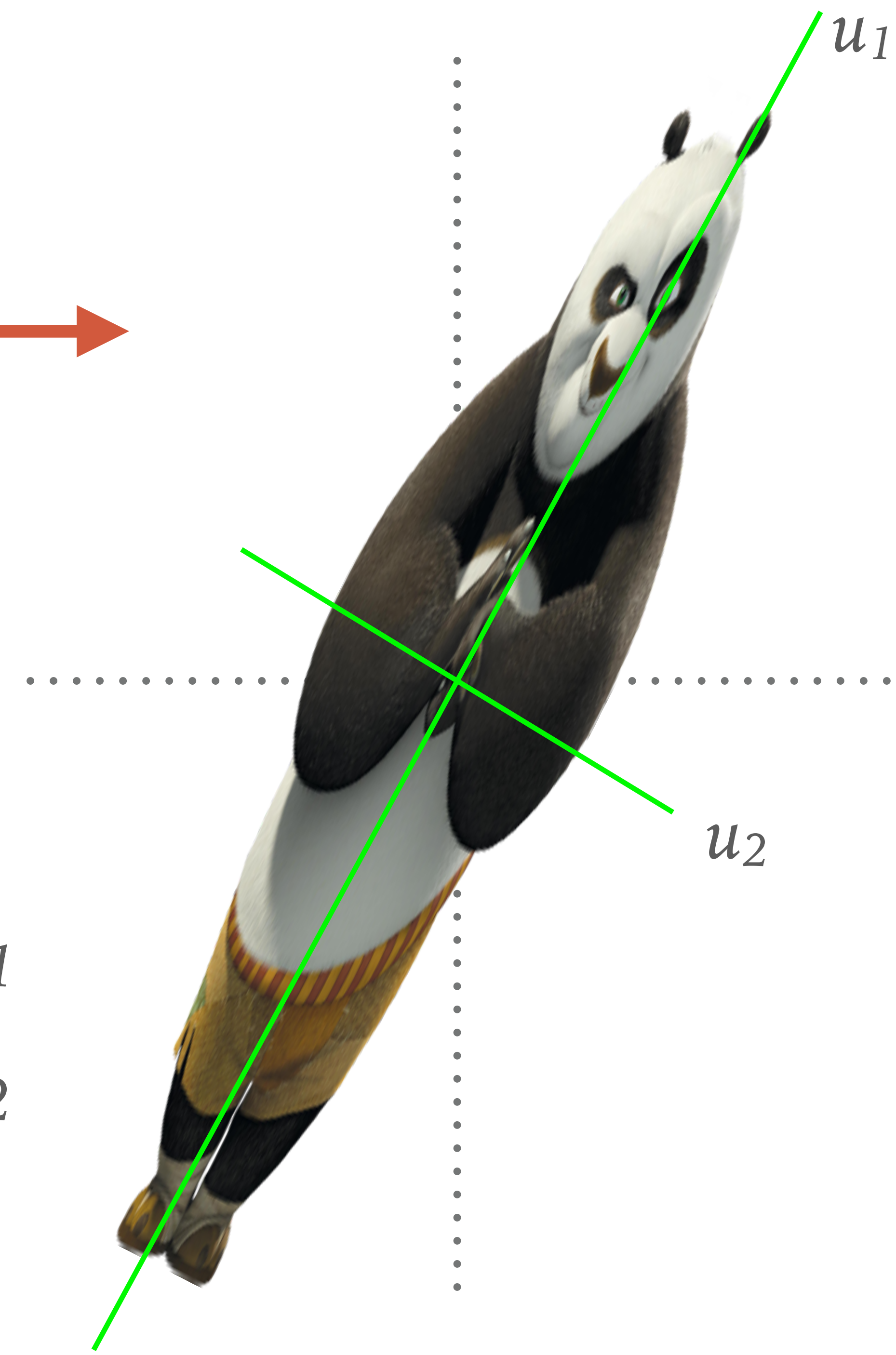
ACTION OF A MATRIX



$$M\, v_1 = d_1\, u_1$$

ACTION OF A MATRIX

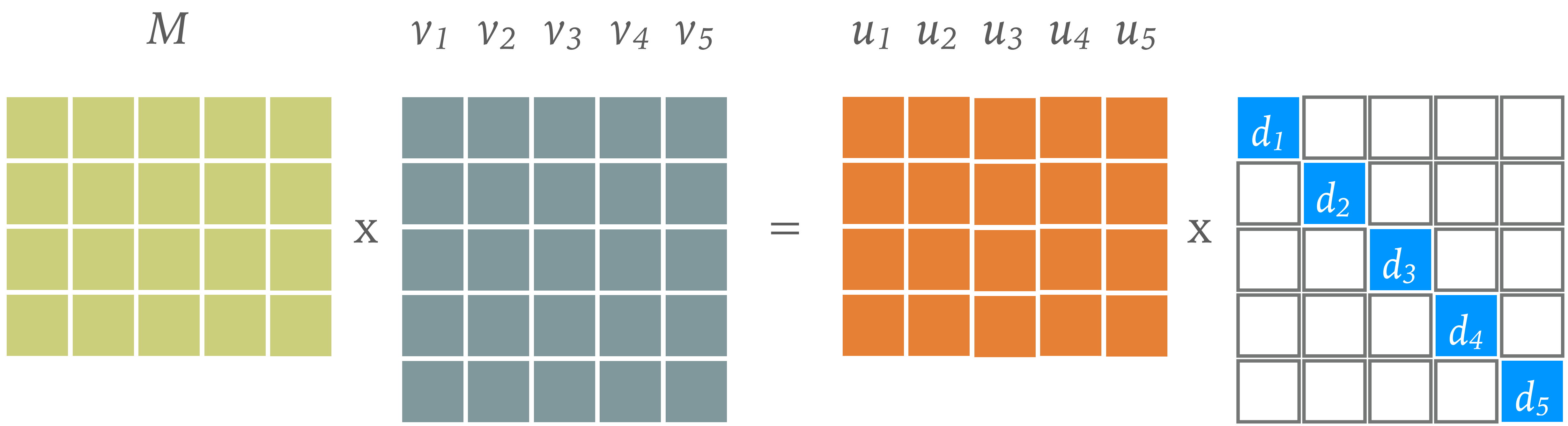




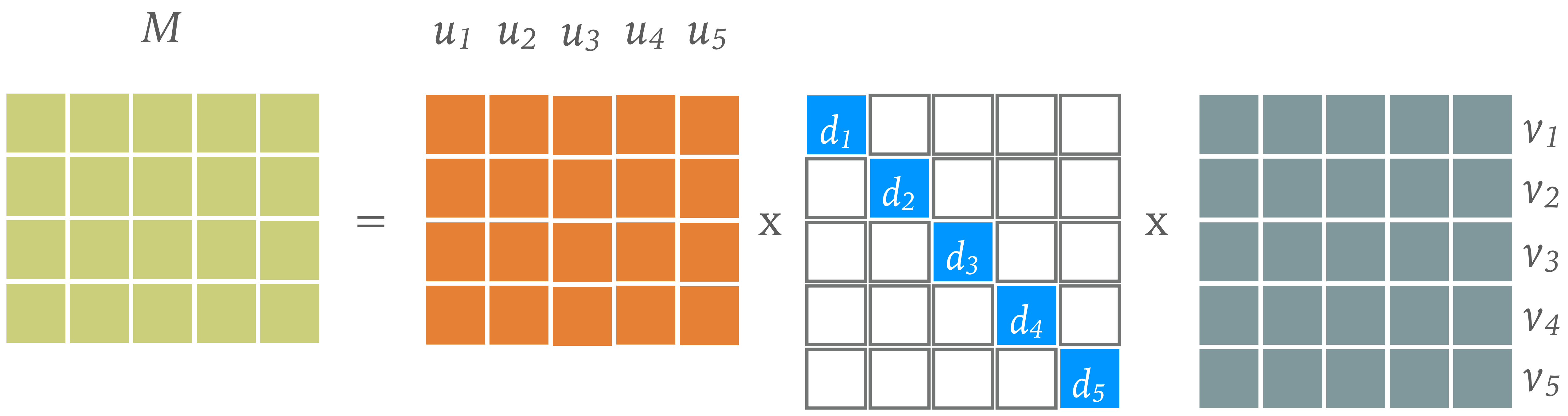
$$M\, v_1 = d_1\, u_1$$

$$M\, v_2 = d_2\, u_2$$

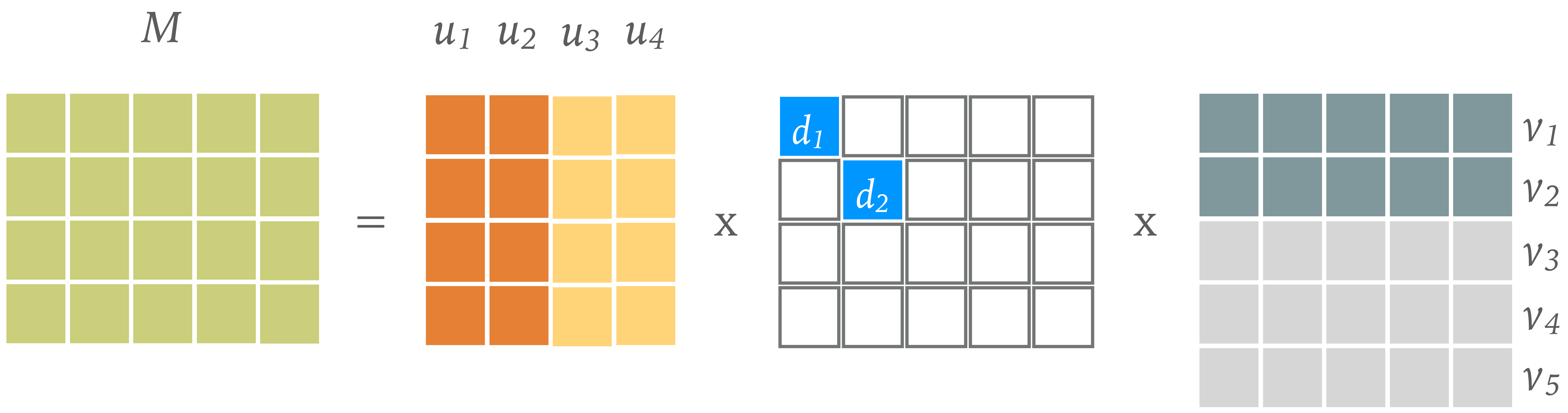
SINGULAR VECTORS



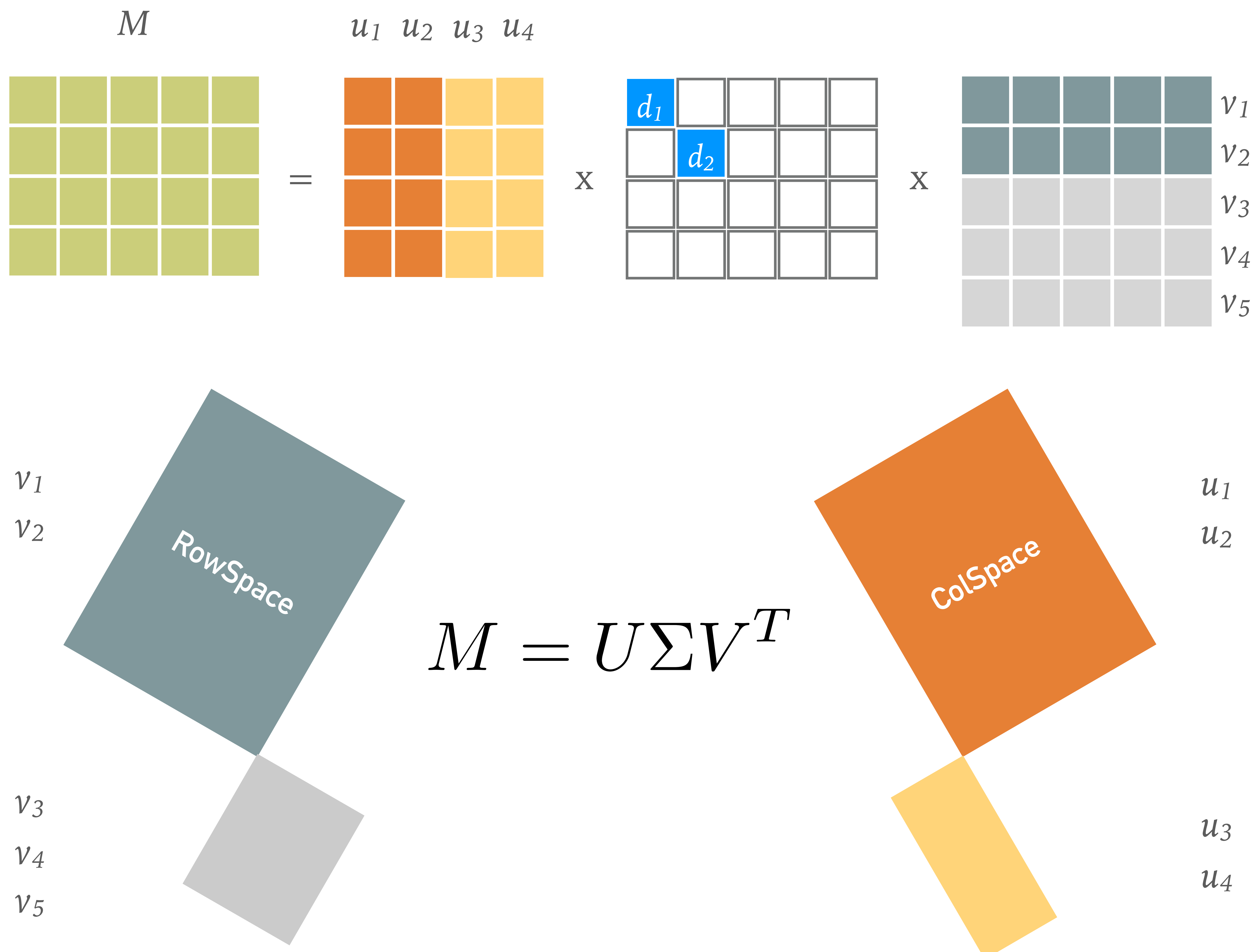
SINGULAR VECTORS



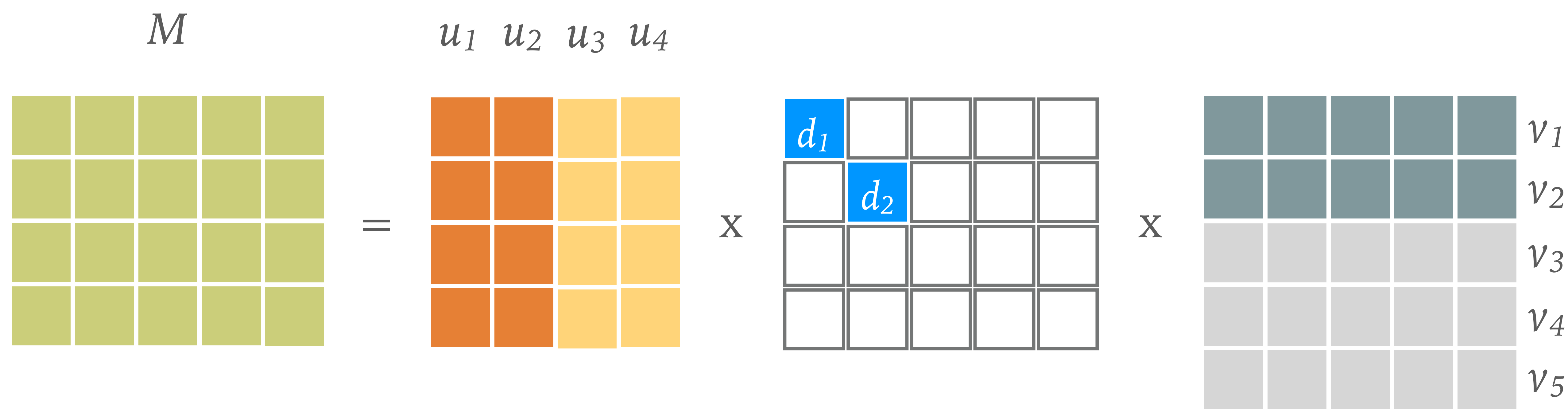
SINGULAR VALUE DECOMPOSITION

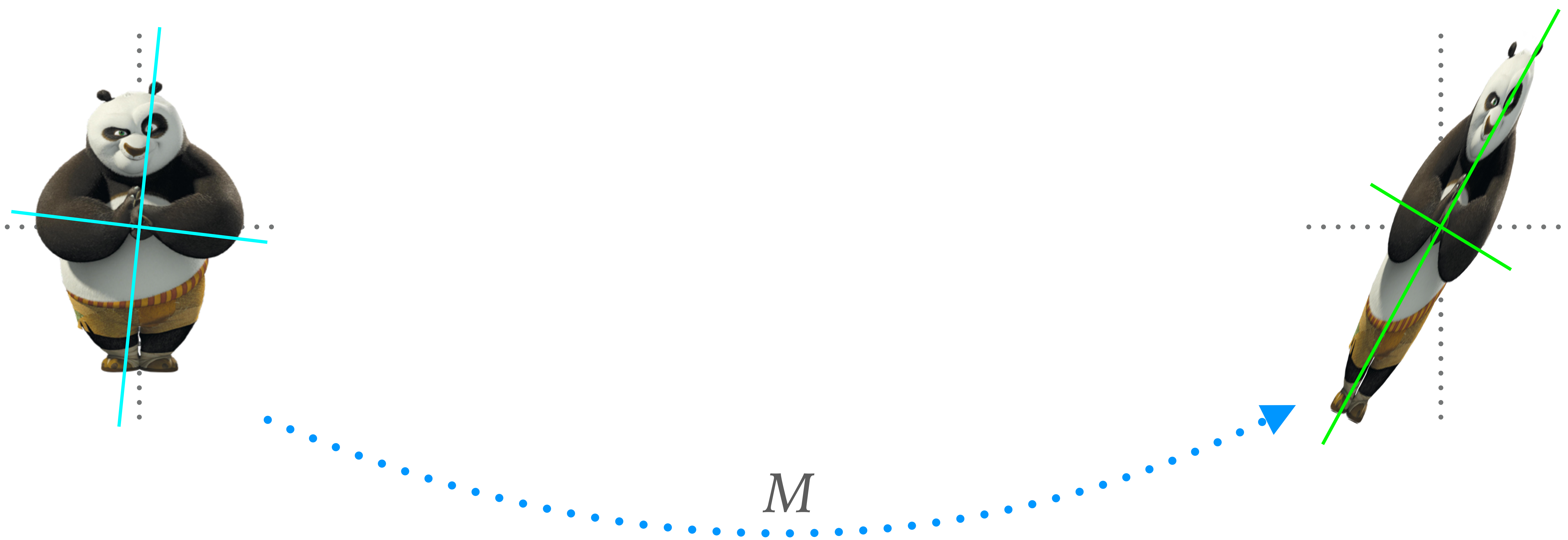


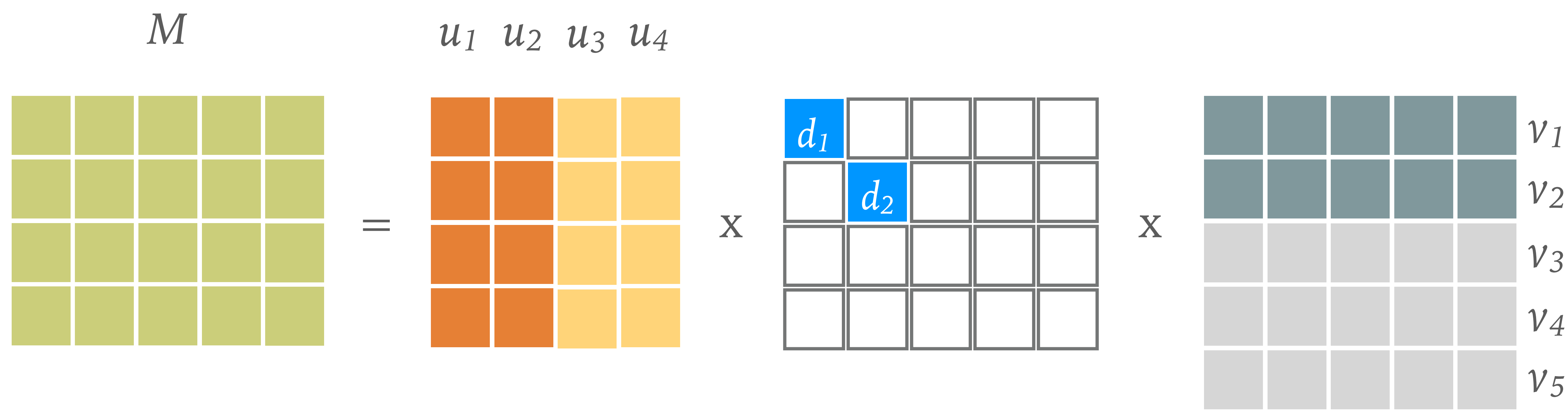
$$M = U\Sigma V^T$$

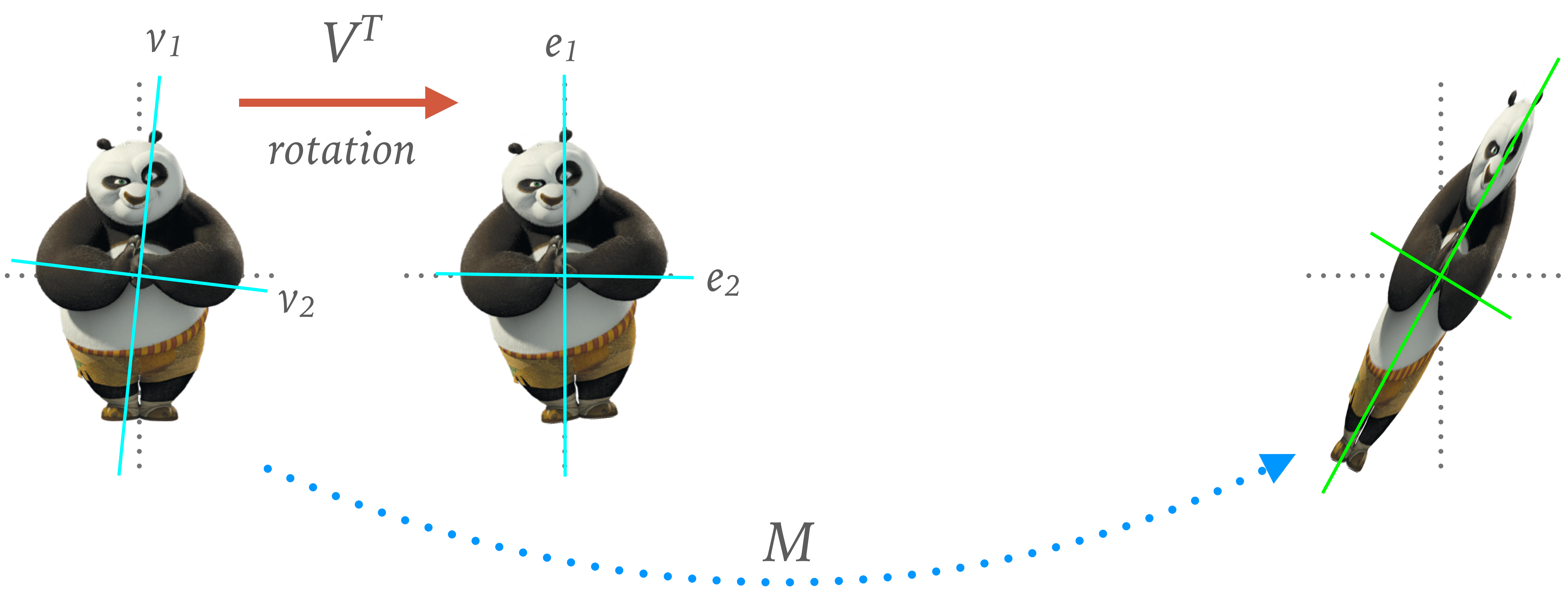


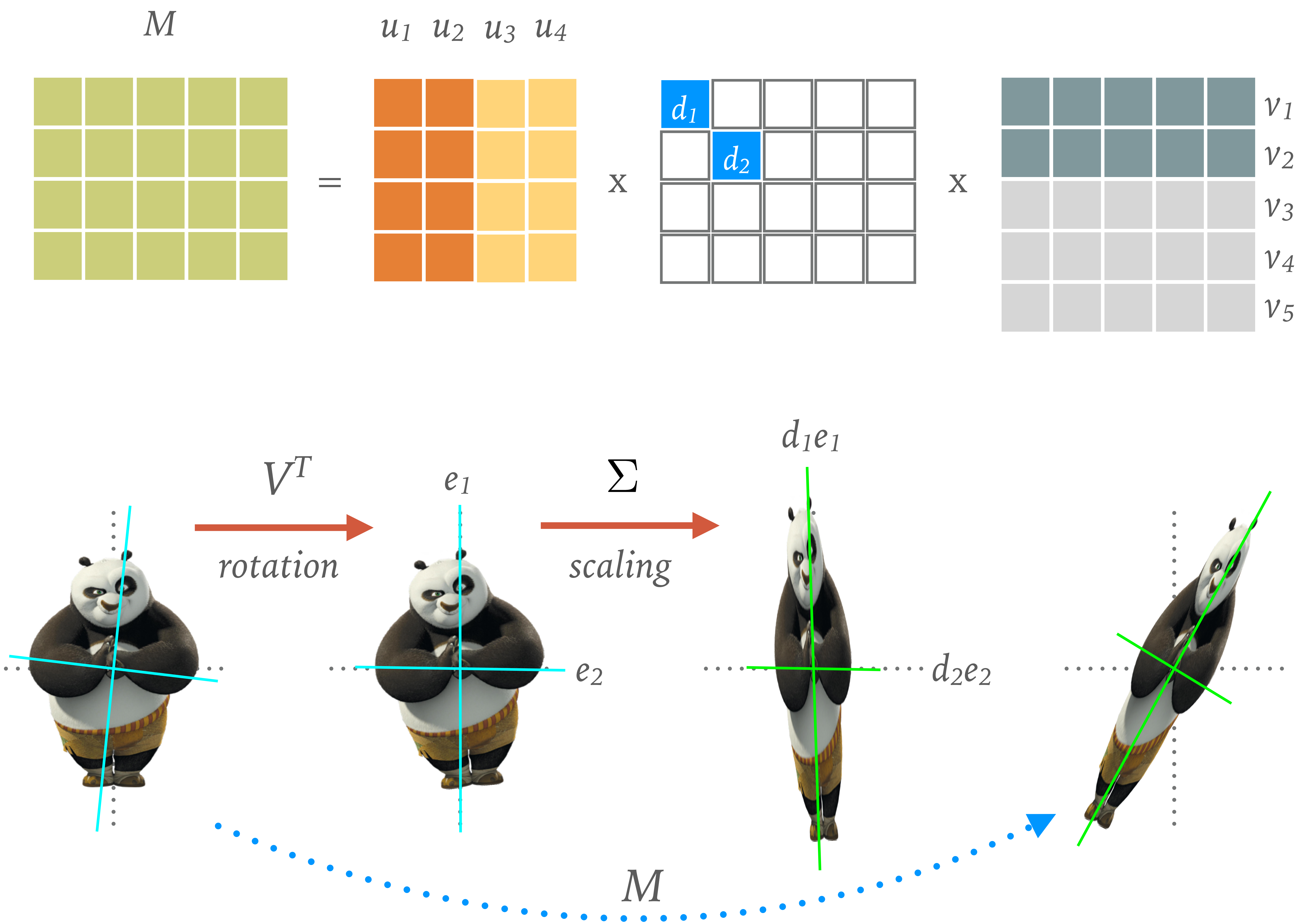
M
u_1 u_2 u_3 u_4
=
x
d_1
d_2
x
v_1
v_2
v_3
v_4
v_5
RowSpace
ColSpace
$M = U\Sigma V^T$


M
=
u1 u2 u3 u4
x
d1
d2
x
v1
v2
v3
v4
v5

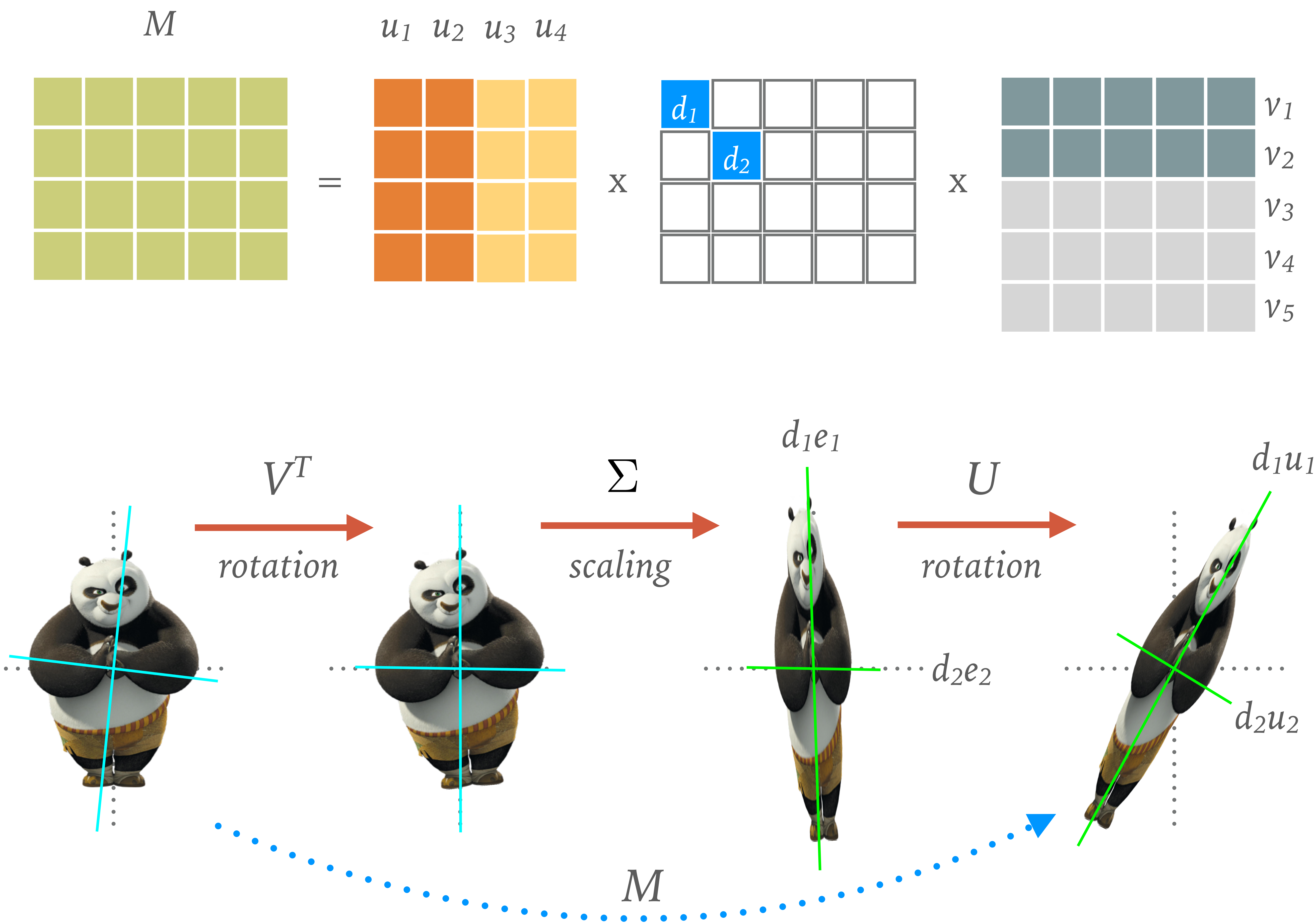

M


M
u_1 u_2 u_3 u_4
=
x
d_1
d_2
x
v_1
v_2
v_3
v_4
v_5

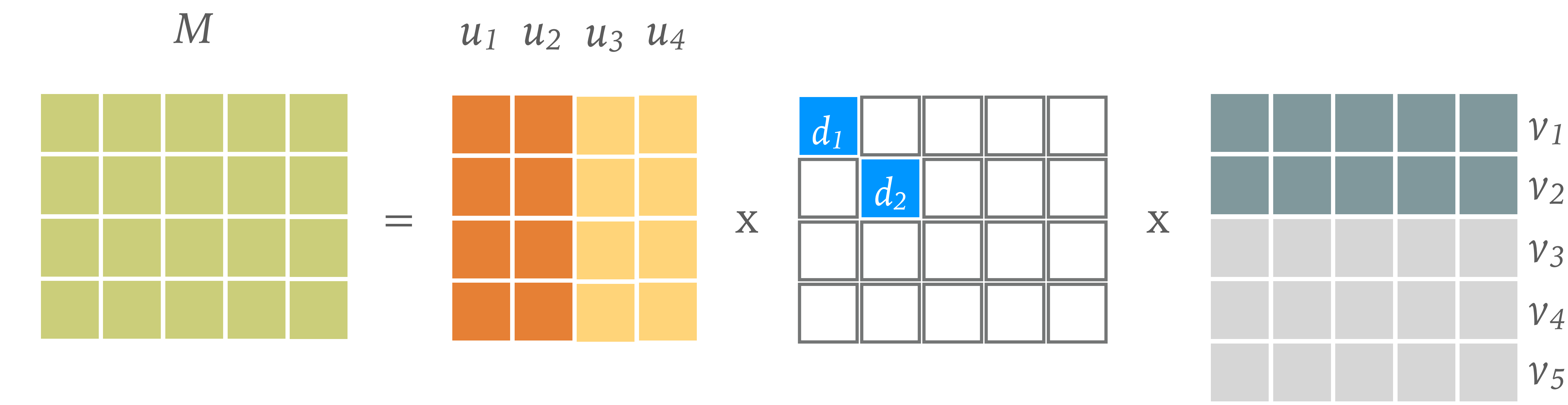

v_1
V^T
rotation
e_1
v_2
e_2
M

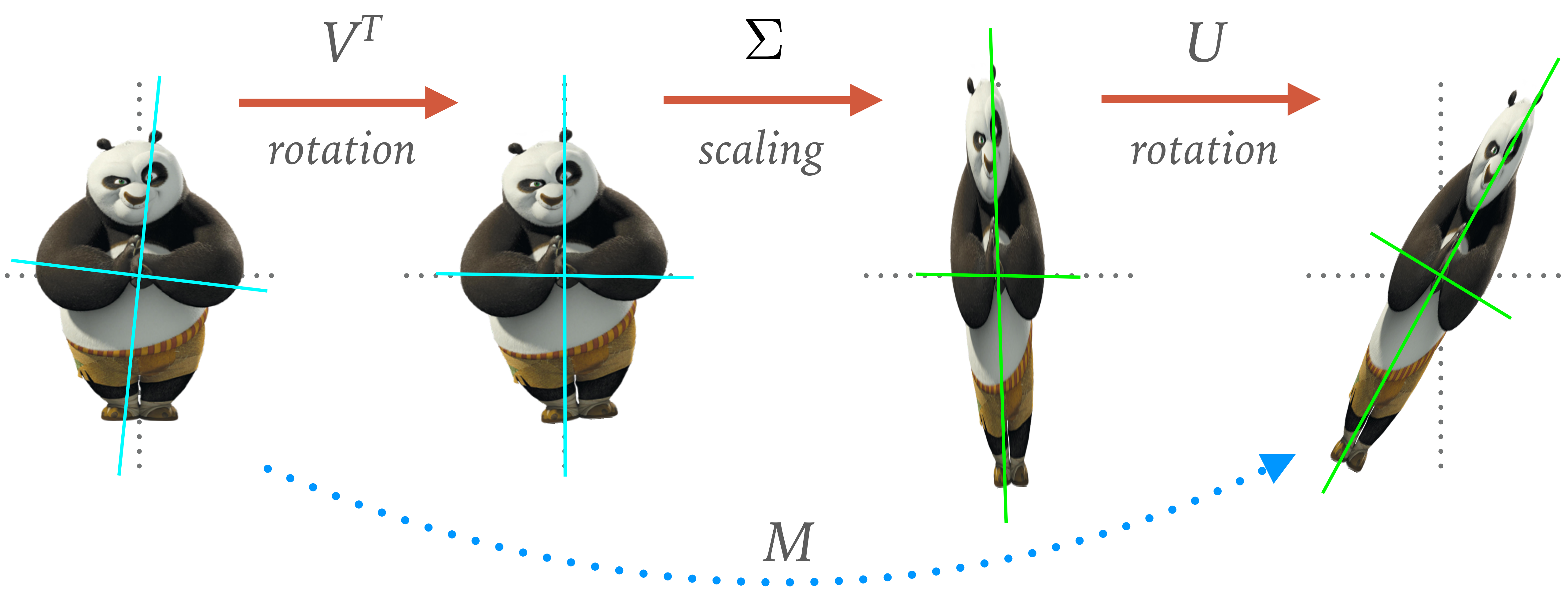


M
=
u1 u2 u3 u4
x
d1
d2
x
v1
v2
v3
v4
v5
V^T
rotation
e1
e2
Σ
scaling
d1e1
d2e2
M



M
u1 u2 u3 u4
=
x
d1
d2
x
v1
v2
v3
v4
v5
V^T
rotation
Σ
scaling
d1e1
d2e2
U
rotation
d1u1
d2u2
M


M
$=$
u_1 u_2 u_3 u_4
x
d_1
d_2
x
v_1
v_2
v_3
v_4
v_5


V^T
rotation
Σ
scaling
U
rotation
M

REFERENCES

Gilbert Strang — *Lectures in Linear Algebra* — *MIT OCW 18.06 (or YouTube)*

Gilbert Strang — *The Fundamental Theorem of Linear Algebra* — *AMM 1993*

Lloyd Trefethen and David Bau — *Numerical Linear Algebra* — *SIAM 1997*