Neural circuits for cognition

MIT 9.49/9.490/6.S076

Instructor: Professor Ila Fiete
TA: Gregg Heller
Senior Instruction Assistant: Adnan Rebei

Logistics/reminders

- HW # 1 due next week (Tuesday)

Today

- Quick review
- From single rate-based neuron to networks of rate-based neurons
- Simplest nonlinear network: single neuron with feedback (autapse)
- Graphical and linear stability analysis
- Begin: simplest multi-neuron networks: linear and symmetric

From single neurons to networks

Rate-based equations

synaptic activation: output

Firing rate

$$\frac{ds_i}{dt} + \frac{s_i}{\tau} = f\left(\overbrace{\underbrace{\sum_j W_{ij}s_j}_{\text{Network input}} + b_i(t)}^{\text{Total input } (g_i) \text{ or input conductance}}\right) \equiv r_i(t)$$

External input

Biophysical time-constant (cell or synapse, depending on which is slow for method of averaging)

Discrete-time network dynamics (for numerical integration)

$$\frac{ds_i}{dt} + \frac{s_i}{\tau} = f(\sum_j W_{ij} s_i + b_i(t)) \equiv r_i$$

Replace derivative by discrete time-difference:

$$\frac{s_i(t + \Delta t) - s_i(t)}{\Delta t} = -\frac{1}{\tau} s_i(t) + f(g_i(t))$$

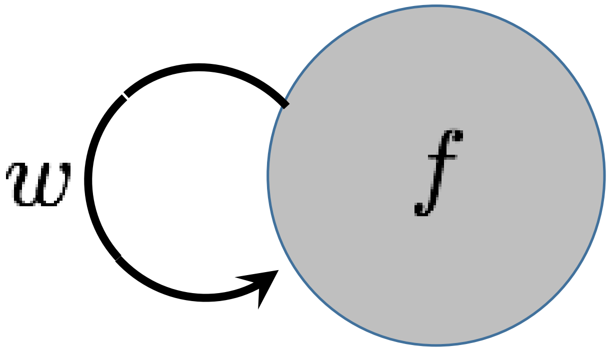
Iteration equation for numerical integration:

$$s_i(t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right) s_i(t) + \Delta t f(g_i(t))$$

The simplest nonlinear network

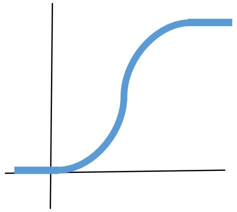
A single-neuron network: the “autapse”

Bistable switch dynamics with an autapse



$$\tau \frac{ds}{dt} + s = f(ws + b)$$

$$f(x) = \frac{e^x}{1 + e^x}$$



Fixed point condition:

$$\frac{ds}{dt} = 0 \qquad \Rightarrow \bar{s} = f(w\bar{s} + b)$$

Solve this equation numerically or graphically to find the fixed points.
And how about the stability of the fixed points?

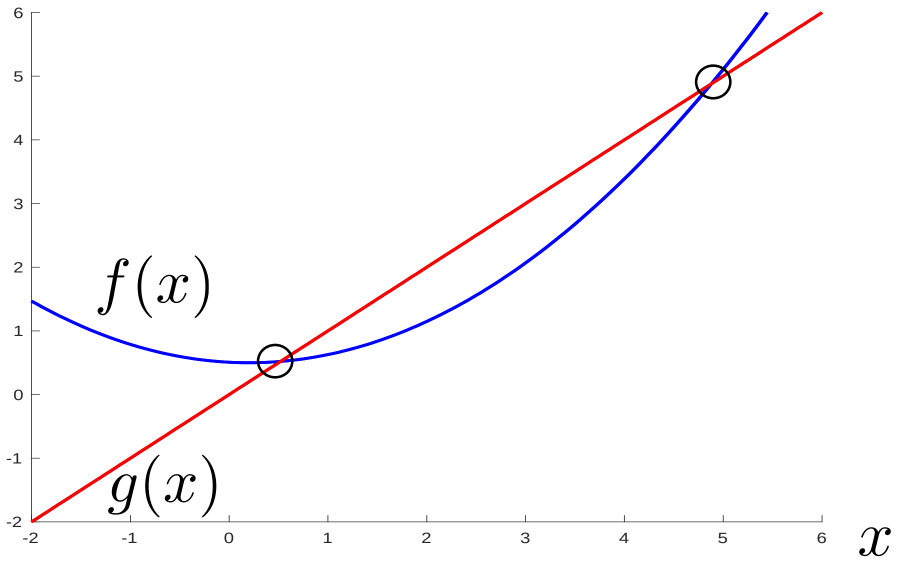
Graphical stability analysis

$$\frac{dx}{dt} = f(x) - g(x)$$

Fixed points:

$$f(\bar{x}) = g(\bar{x})$$

Graphical stability analysis



$$\frac{dx}{dt} = f(x) - g(x)$$

Fixed points:

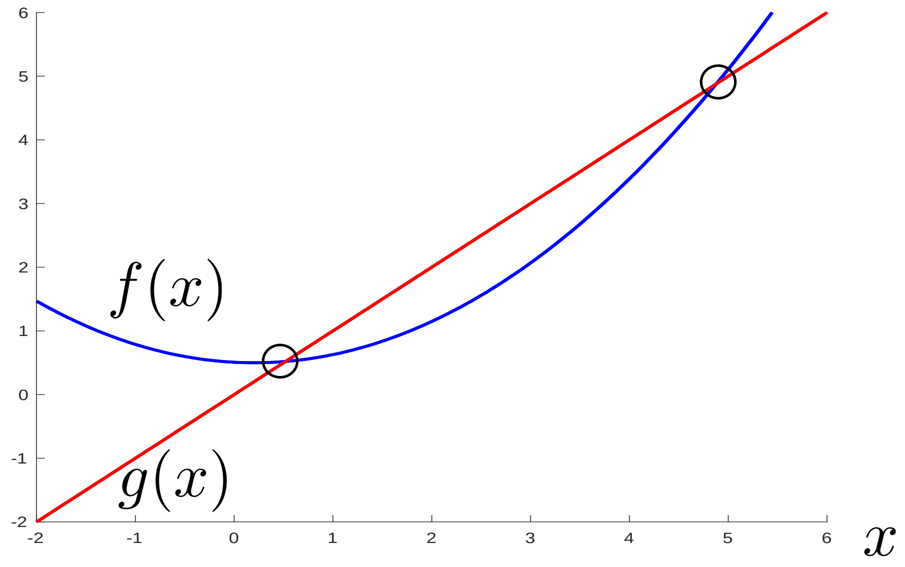
$$f(\bar{x}) = g(\bar{x})$$

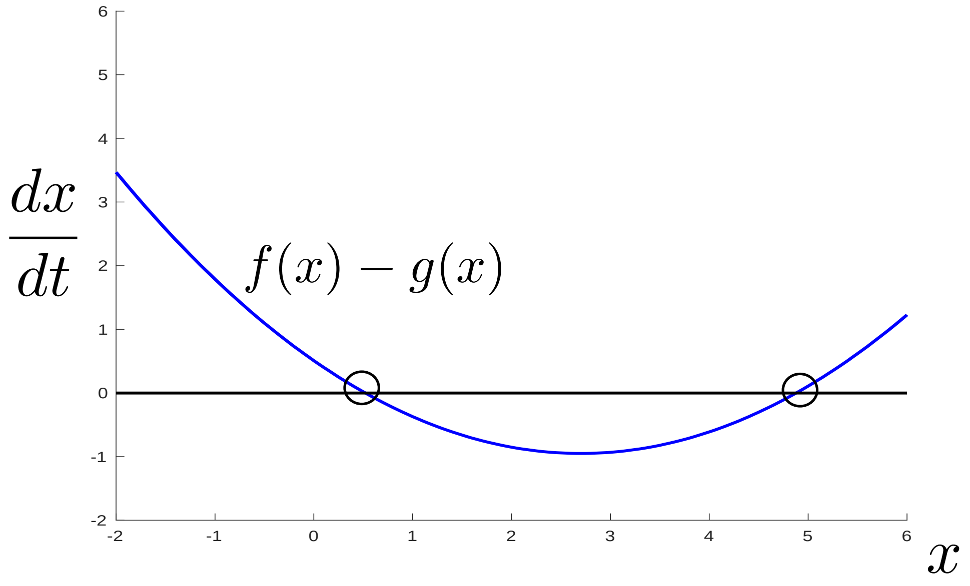
Graphical stability analysis

$$\frac{dx}{dt} = f(x) - g(x)$$

Fixed points:

$$f(\bar{x}) = g(\bar{x})$$



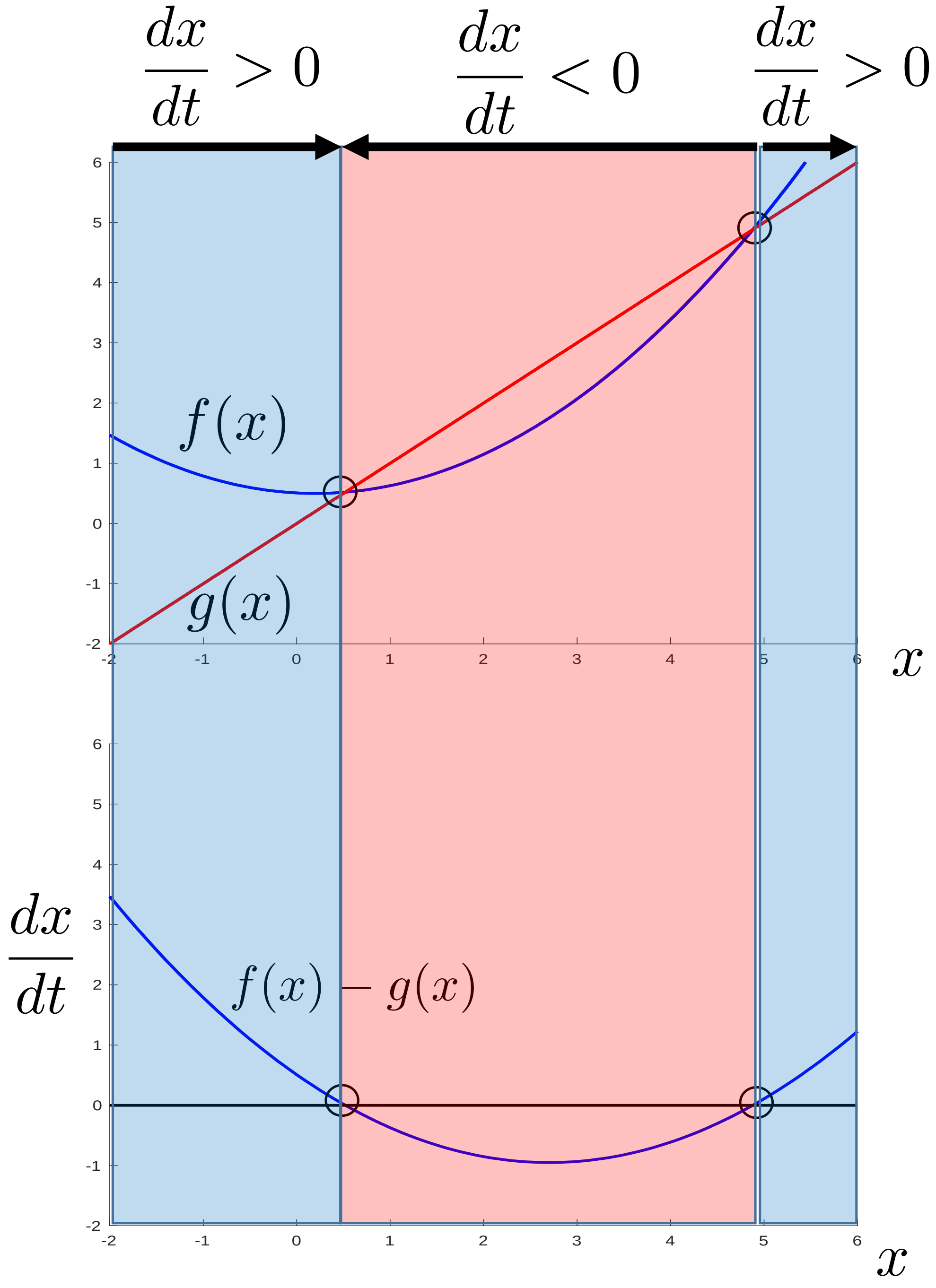


Graphical stability analysis

$$\frac{dx}{dt} = f(x) - g(x)$$

Fixed points:

$$f(\bar{x}) = g(\bar{x})$$

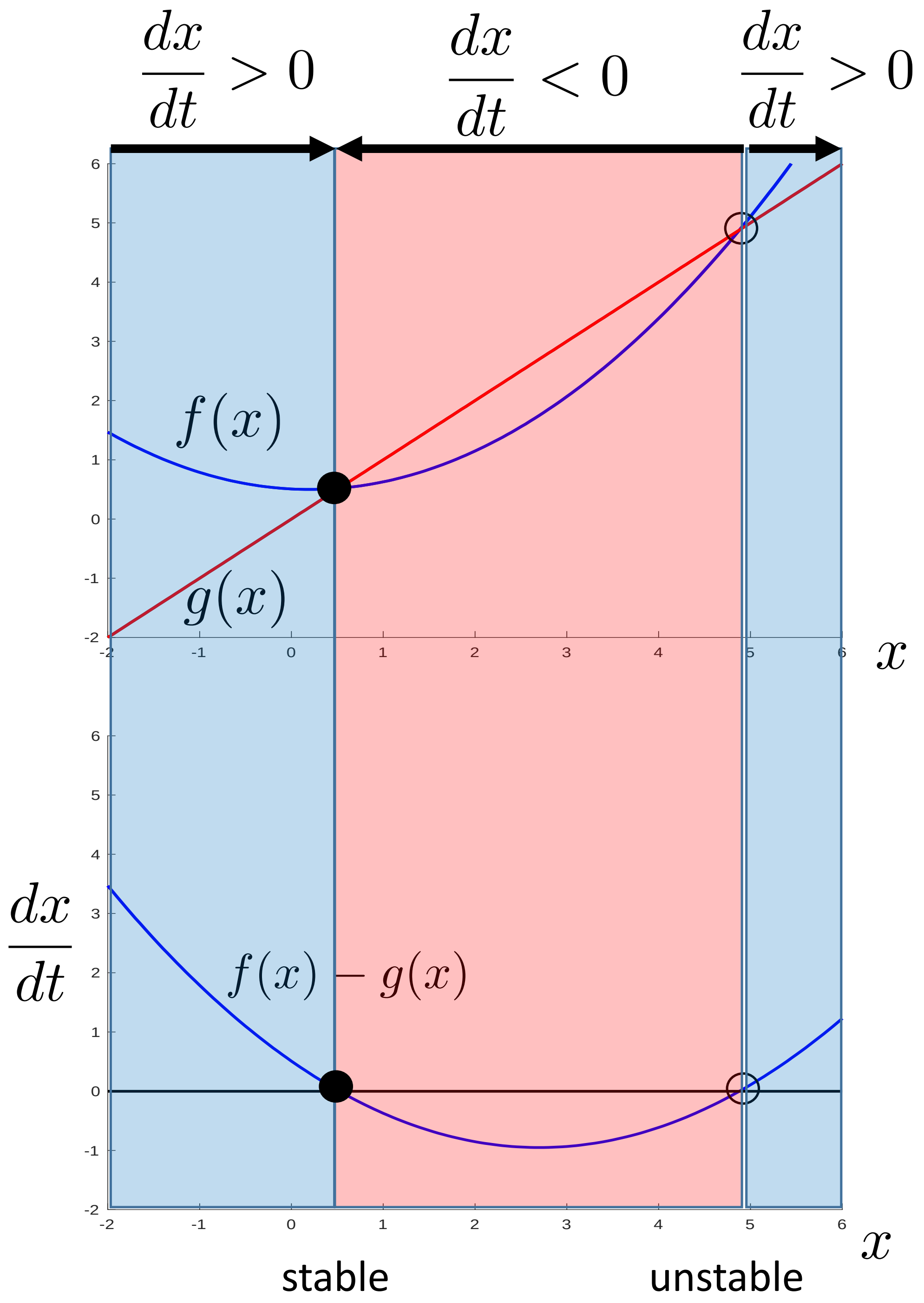


Graphical stability analysis

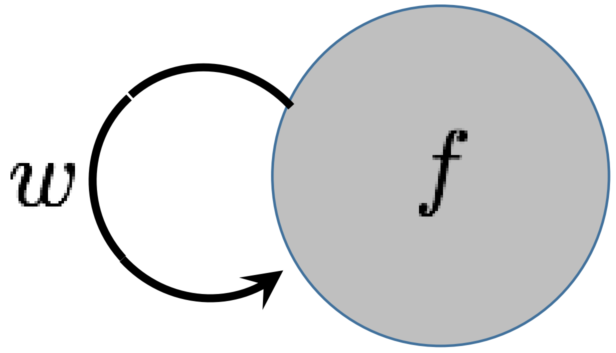
$$\frac{dx}{dt} = f(x) - g(x)$$

Fixed points:

$$f(\bar{x}) = g(\bar{x})$$



Back to bistable switch: autapse



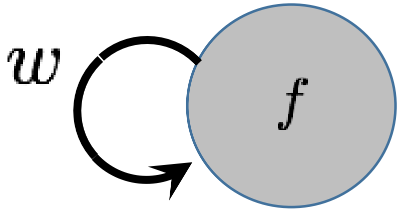
$$\tau \frac{ds}{dt} + s = f(ws + b)$$

$$f(x) = \frac{e^x}{1 + e^x}$$

Fixed point condition:

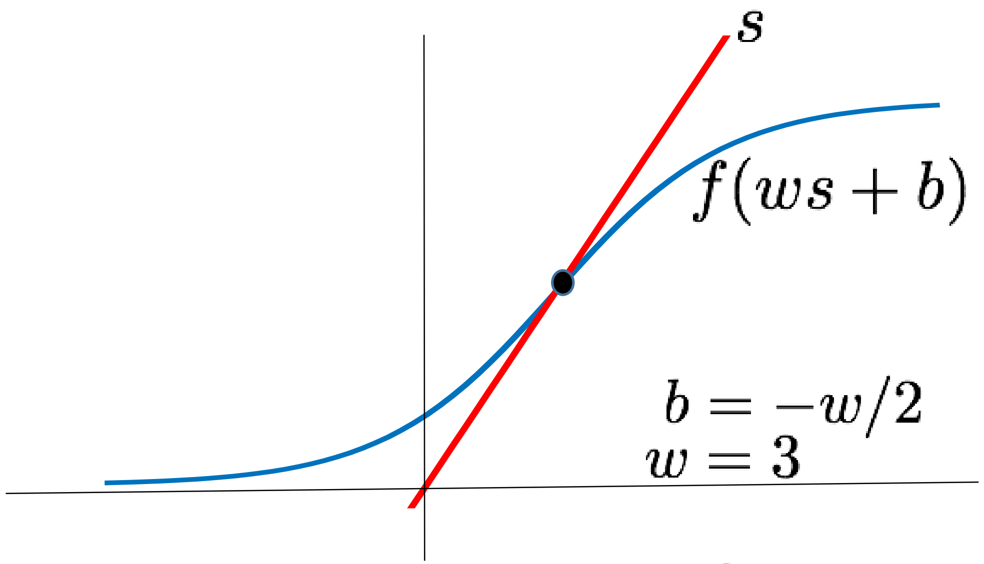
$$\frac{ds}{dt} = 0 \qquad \Rightarrow \bar{s} = f(w\bar{s} + b)$$

Finding fixed points of autapse graphically

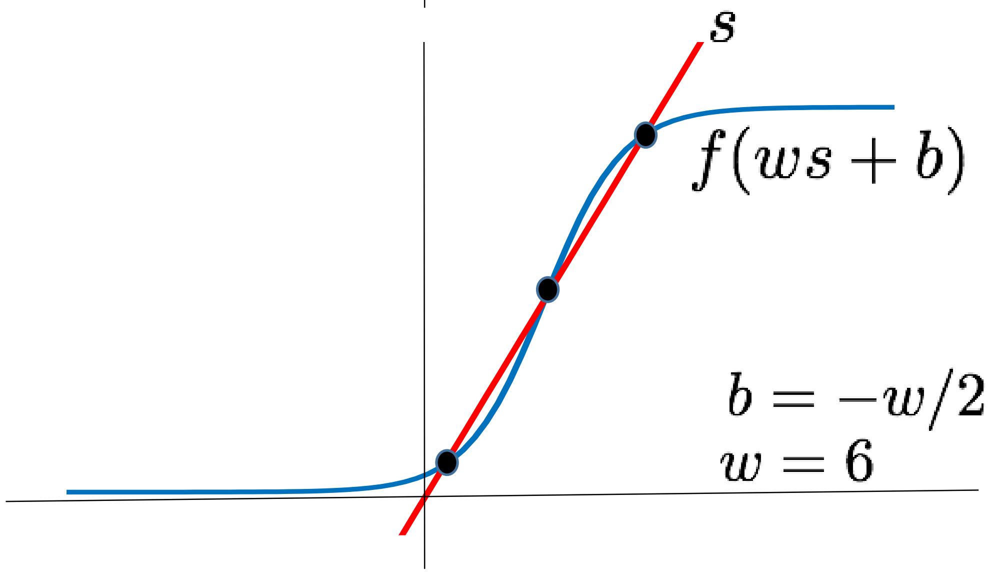


$$\bar{s} = f(w\bar{s} + b)$$

Fixed points equation



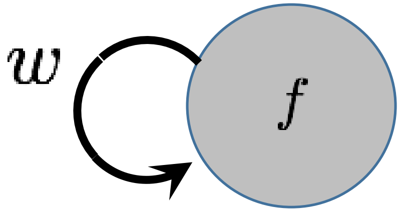
Shallow sigmoid:
one fixed point



Steep sigmoid:
three fixed points

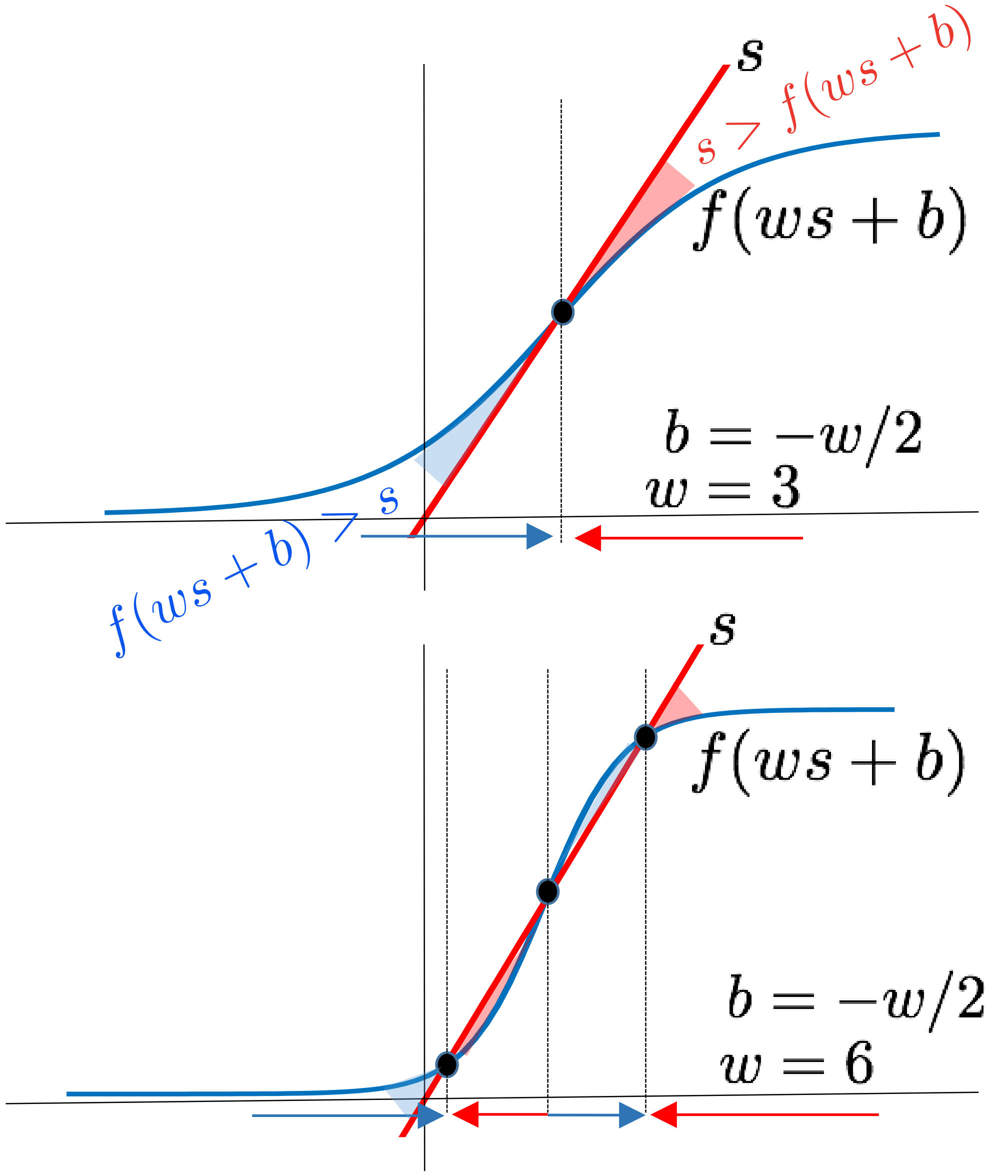
“Pitchfork bifurcation” as a function of parameter *w*: from 1 to 3 fixed points.

Stability of fixed points: graphical analysis



$$\bar{s} = f(w\bar{s} + b)$$

$$\tau \frac{ds}{dt} + s = f(ws + b)$$



fixed point is stable

central fixed point becomes unstable; other two are stable.

Linear stability analysis of non-linear systems

Linear (local) stability analysis

Arbitrary autonomous nonlinear dynamical system (autonomous – no driving term *g(t)*):

(single variable for now; will generalize to multi-dimensional case later)

$$\frac{dx}{dt} = -x + f(x)$$

Linear (local) stability analysis

Arbitrary autonomous nonlinear dynamical system (autonomous – no driving term g(t)):

(single variable for now; will generalize to multi-dimensional case later)

$$\frac{dx}{dt} = -x + f(x)$$

Suppose there is a set of fixed points, indexed by i:

$$\bar{x}_i = f(\bar{x}_i)$$

Linear (local) stability analysis

Arbitrary autonomous nonlinear dynamical system (autonomous – no driving term g(t)):

$$\frac{dx}{dt} = -x + f(x)$$

Suppose there is at least one fixed point:

$$\bar{x} = f(\bar{x})$$

We can examine the dynamics equation at values of *x* near this fixed point:

$$x = \bar{x} + \delta x$$

where δx is very small ($\delta x \rightarrow 0$). This is why the approach is called "local". Idea: Taylor expand the non-linear function f around the fixed point to lowest order to get linear equation.

Linear (local) stability analysis

$$\frac{d(\bar{x}+\delta x)}{dt} = -\bar{x} - \delta x + f(\bar{x}+\delta x)$$

Linear Taylor approximation: $f(\bar{x}+\delta x) \approx f(\bar{x}) + f'(x)|_{\bar{x}}\delta x$

Obtain linear equation for the dynamics near the fixed point:

$$\frac{d\delta x}{dt} = \cancel{-\bar{x}} + \cancel{f(\bar{x})} - \delta x + f'(\bar{x})\delta x \;\; = -\underbrace{\left(1 - f'(\bar{x})\right)}_{\text{constant}}\delta x$$

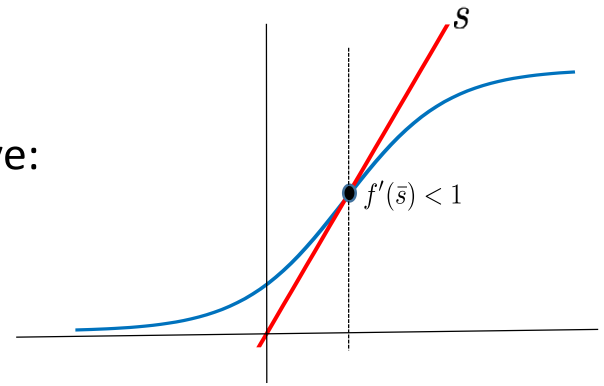
Gives a simple solution for how the perturbations δx will evolve:

$$\delta x(t) = \delta x(0)e^{-(1-f'(\bar{x}))t}$$

Linear (local) stability analysis of $\bar{x}$

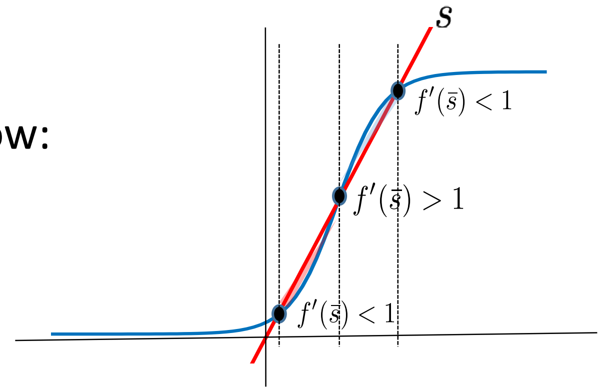
Linearization around $\bar{x}$

$$\frac{dx}{dt} = -x + f(x) \longrightarrow \frac{d\delta x}{dt} = -\left(1 - f'(\bar{x})\right)\delta x$$



Simple exponential growth/decay solution describing how the perturbations δx evolve:

$$\delta x(t) = \delta x(0) e^{-(1-f'(\bar{x}))t}$$



$\bar{x}$ STABLE if perturbations decay:

$$(1 - f'(\bar{x})) > 0$$

$\bar{x}$ UNSTABLE if perturbations grow:

$$(1 - f'(\bar{x})) < 0$$

Relative merits: graphical vs linear stability analysis

Graphical	Linear
Global; can get basins	Local
Low-dimensional systems	High-dimensional systems
Not easy to use in general	Easy to use
Qualitative	Quantitative

Bistable switch summary

- A system with nonlinear positive feedback that is superlinear/accelerating, with saturation, can exhibit bistability.
- Bistability occurs for a sufficiently steep positive feedback curve, out of a pitchfork bifurcation as a function of the steepness parameter.
- A single neuron exciting itself is a “cartoon” of positive feedback within a network: it can exhibit switch dynamics.
- Graphical stability analysis and linear stability analysis are tools to examine stability of fixed points.
- Further analysis of this system: homework.

Simplest multi-neuron networks: linear networks

Notation

- Matrices: upper-case A, B, U, W $\mathbf{A}, \mathbf{B}, \mathbf{U}, \mathbf{W}$
- Column vector: **bold**, (usually) lower-case $\mathbf{x}, \mathbf{y}, \mathbf{v}, \mathbf{w}$

  (handwriting: $\mathbf{x} \to \underline{x}$)
- Scalars a, b, c, γ, α
- Discrete indices $i, j, k, l, m, n; \alpha, \beta$

The rate-based network equation (vector-matrix form)

$$\frac{ds_i}{dt} + \frac{s_i}{\tau} = f(\sum_j W_{ij} s_i + b_i(t))$$

$$\frac{d\mathbf{s}}{dt} + \frac{\mathbf{s}}{\tau} = f(\mathbf{W}\mathbf{s} + \mathbf{b})$$

Some notation: vectors and matrices

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

size *(m x 1)* column vector

$$v_i \in \mathbb{R}$$
$$\mathbf{v} \in \mathbb{R}^m$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

size *(n x m)* matrix

$$A \in \mathbb{R}^{n \times m}$$

Some more notation

- Matrices: upper-case A, B, U, W
- Column vector: **bold**, (usually) lower-case $\mathbf{x}, \mathbf{y}, \mathbf{v}, \mathbf{w}$

  (handwriting: $\mathbf{x} \to \underline{x}$)

- Scalars a, b, c, γ, α
- Discrete indices $i, j, k, l, m, n; \alpha, \beta$

The rate-based network equation (vector-matrix form)

$$\frac{ds_i}{dt} + \frac{s_i}{\tau} = f(\sum_j W_{ij} s_i + b_i(t))$$

$$\frac{d\mathbf{s}}{dt} + \frac{\mathbf{s}}{\tau} = f(\mathbf{W}\mathbf{s} + \mathbf{b})$$

Linear algebra: basics to review

Please look at linear algebra slides and primer on course website if you'd like a quick refresher/introduction to some basic definitions and concepts for vectors and matrices.

- Eigenvalues, eigenvectors
- Orthogonality
- Properties of real, symmetric matrices ($\mathbf{M}=\mathbf{M}^T$)

- Notes available: Linear algebra primer

Eigenvectors and eigenvalues

If for a linear operator (matrix) M, there exists a non-zero vector **v** such that:

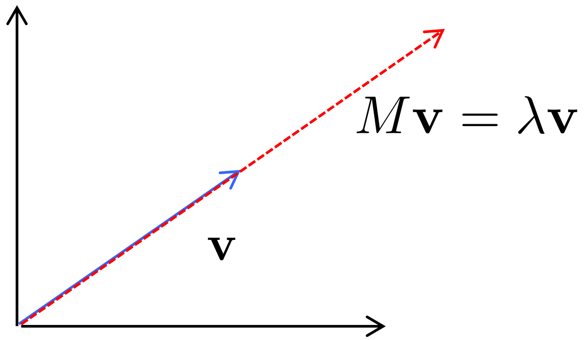
$$M\mathbf{v} = \lambda\mathbf{v}$$

where λ is a scalar, then **v** is an eigenvector of M, with eigenvalue λ.

Eigenvalues are given by the roots of the characteristic equation: $|M - \lambda I| = 0$

In other words, the eigenvectors of a matrix are a special set of vectors for which the matrix product acts simply as a scalar product:

geometric view



same direction, different length

Eigenvectors and eigenvalues of square matrices

- The eigenvalues of real-valued matrices are generally complex and eigenvectors need not be orthogonal.

- The eigenvalues of real symmetric matrices are real eigenvalues and their eigenvectors are orthogonal.

- Asymmetric matrices in general have different left and right eigenvectors, but have a common set of eigenvalues.

Linear and linearized networks, relationship to linear systems

Linear(ized) neural networks

Linearized dynamics of a *nonlinear neural network* around a point $\bar{\mathbf{s}}$:

$$\frac{d\delta\mathbf{s}}{dt} + \frac{\delta\mathbf{s}}{\tau} = \mathbf{DW}\delta\mathbf{s} \qquad \mathbf{D}_{ij} = \left(\frac{\partial f}{\partial g_i}\bigg|_{\bar{\mathbf{s}}}\right)\delta_{ij}$$

A linear neural network:

$$\tau\frac{d\mathbf{s}}{dt} + \mathbf{s} = \mathbf{W}\mathbf{s} + \mathbf{b} \qquad (\mathbf{D} = \mathbb{I})$$

The rate-based network equation (vector-matrix form)

$$\frac{ds_i}{dt} + \frac{s_i}{\tau} = f(\sum_j W_{ij} s_j + b_i)$$

Linearized dynamics in the vicinity of some state $\bar{\mathbf{s}}$: $\mathbf{s} = \bar{\mathbf{s}} + \delta \mathbf{s}$

$$\frac{d\delta s_i}{dt} + \frac{\delta s_i}{\tau} = \left(\left. \frac{\partial f}{\partial g_i} \right|_{\bar{\mathbf{s}}} \right) \sum_j W_{ij} \delta s_j$$

$$\frac{d\delta \mathbf{s}}{dt} + \frac{\delta \mathbf{s}}{\tau} = \mathbf{DW} \delta \mathbf{s}$$

$$\mathbf{D}_{ij} = \left(\left. \frac{\partial f}{\partial g_i} \right|_{\bar{\mathbf{s}}} \right) \delta_{ij}$$

Linear(ized) dynamical system fixed points correspond to the roots of corresponding linear systems

Fixed points of $$\frac{d\mathbf{x}}{dt} = W\mathbf{x}$$

$$\Updownarrow$$

Solutions of $$W\mathbf{x} = 0$$

Linear systems review

n equations in *m* unknowns *(v_1,…v_m)*:

$$
\begin{aligned}
a_{11}v_1 + \cdots + a_{1m}v_m &= b_1 \\
a_{21}v_1 + \cdots + a_{2m}v_m &= b_2 \\
&\cdots\cdots\cdots \\
a_{n1}v_1 + \cdots + a_{nm}v_m &= b_n
\end{aligned}
$$

$$
\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}}_{(n \times m)}
\underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}}_{(m \times 1)}
=
\underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}}_{(n \times 1)}
$$

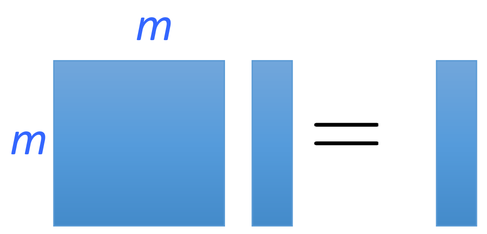
$$A\mathbf{v} = \mathbf{b}$$

System of equations: when does unique solution exist?

n equations (constraints) in *m* unknowns: *generically (though not exactly always!)*, a unique solution exists when, *n=m or A* is square.

$$\underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}}_{(m \times m)} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}}_{(m \times 1)} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{(n \times 1)}$$

$$\underset{(m \times m)\,(m \times 1)}{A\mathbf{v}} = \underset{(m \times 1)}{\mathbf{b}}$$



For a square matrix, when is a unique solution guaranteed to exist? Time for some geometric insight.

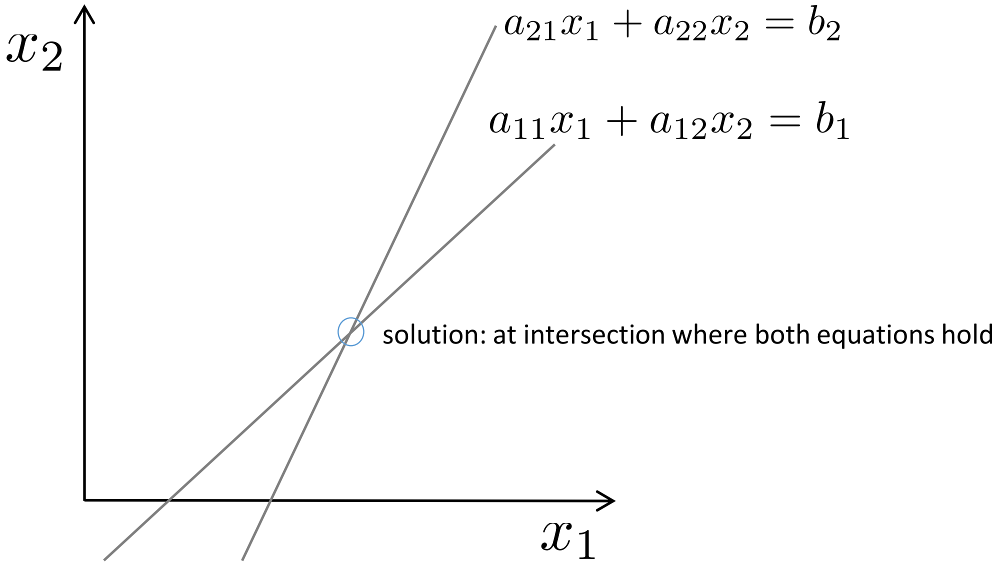
Geometric view: when does a unique solution exist?

E.g. 2-dimensional problem: 2 unknowns, 2 equations

equation of a line

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

unknowns x_1, x_2



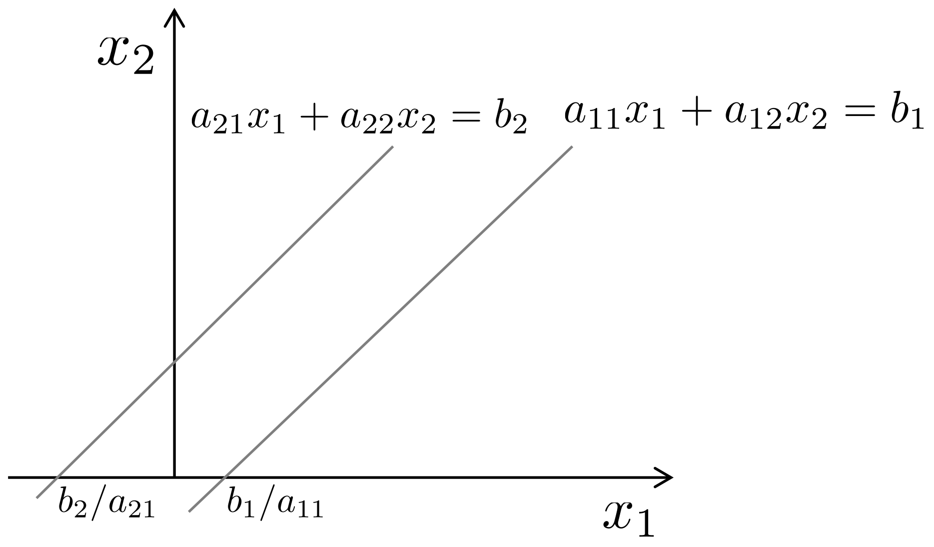
Two lines in 2D *generically* intersect at a (single) location thus generically a unique solution exists.

Geometric view: Two ways that a unique solution *does not* exist in 2D

What are these?

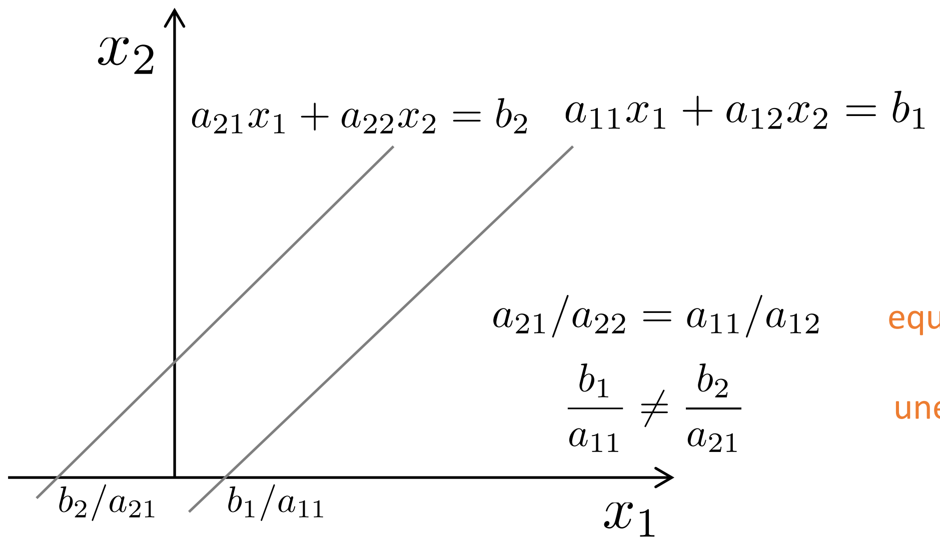
Geometric view: Two ways a unique solution *does not* exist in 2D

1. Offset parallel lines: **no solution**



Algebra: when does a unique solution *not* exist?

1. Offset parallel lines: **no solution**



$a_{21}/a_{22} = a_{11}/a_{12}$ equal slopes

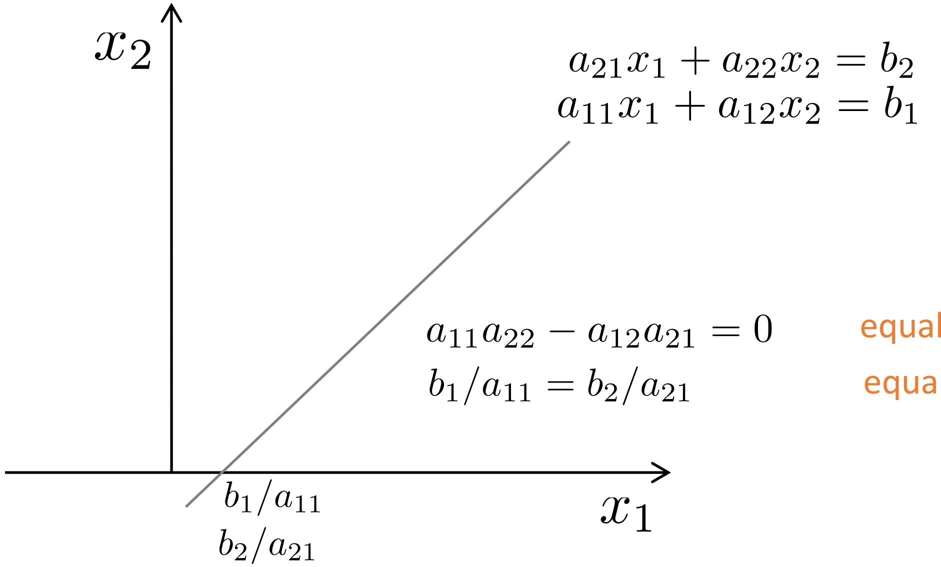
$\frac{b_1}{a_{11}} \neq \frac{b_2}{a_{21}}$ unequal intercepts

$$a_{11}a_{22} - a_{12}a_{21} = 0$$

$$\det(A) \equiv a_{11}a_{22} - a_{12}a_{21}$$

Algebra: when does a unique solution *not* exist?

2. Aligned parallel lines**: infinitely many solutions**



$a_{11}a_{22} - a_{12}a_{21} = 0$ equal slopes

$b_1/a_{11} = b_2/a_{21}$ equal intercepts

Back to algebraic view: existence of unique solution in terms of coefficient matrix *A*

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

determinant: $\det(A) \equiv a_{11}a_{22} - a_{12}a_{21}$

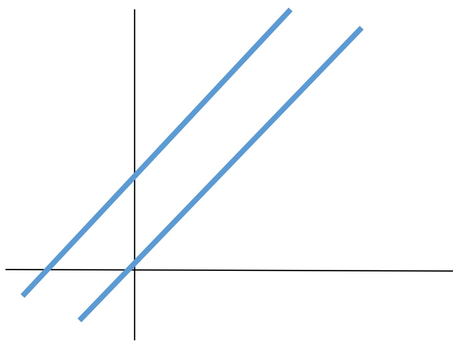
2-dim system of equations with square coefficient matrix *A* has a unique solution when:

$$\det(A) \neq 0$$

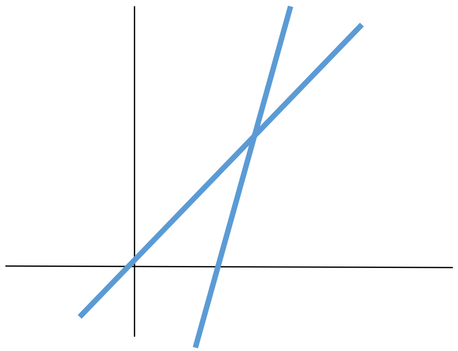
Same condition for *m*-dim system of equations with square coefficient matrix: need non-singular determinant.

Fixed points of any linear dynamical system

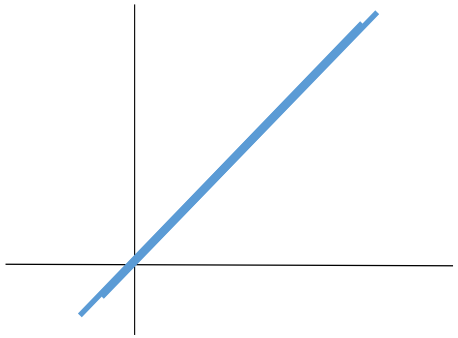
- A linear system (of any dimension) admits exactly 0, 1, or infinitely many fixed points.



0 solutions
NOT generic



1 solution
(generic case)
square matrix, non-zero determinant

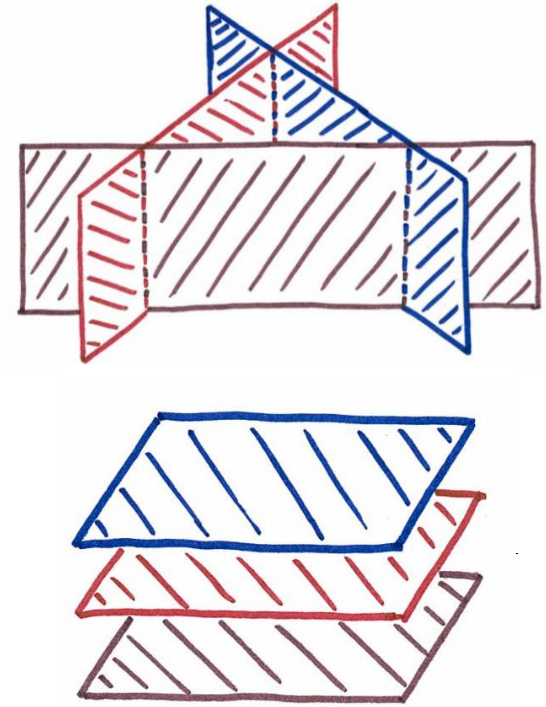


Infinitely many solutions
NOT generic

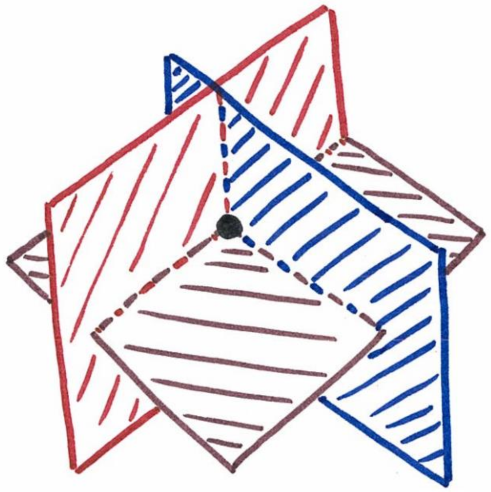
Corollary: A linear system cannot exhibit a finite number >1 of fixed points (cf. our bistable switch)

Linear dynamical systems: all possibilities

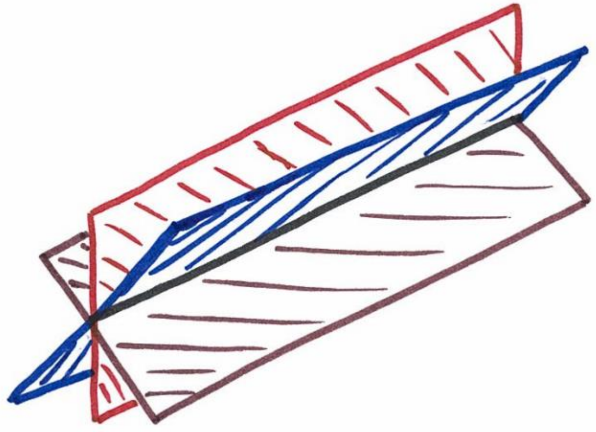
- A linear system admits 0, 1, or infinitely many fixed points.
- **Regardless of system dimension**: these are the only possibilities.



0 solutions
NOT generic



1 solution
(generic case)
square matrix, non-zero determinant



Infinitely many solutions
NOT generic

Image credit: https://www.math.utah.edu/~wortman/1050-text-lei3v.pdf

Summary

- Global and linear stability analysis
- Accelerating positive feedback + saturation → bistability
- Linear dynamical systems and relationship with linear systems of equations: fixed points of dynamical system are roots of linear system
- Linear dynamical systems admit 0,1, or infinitely many fixed points