Systems of ODEs

A continuous-time Dynamical System is defined by a system of differential equations: $\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, t; \theta)$ or $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t; \theta)$

- Initial conditions: where the system is at the beginning of the evolution: $\mathbf{x}(t_0)$
- Phase space: space of all possible states
- Trajectory: the curve traces by $\mathbf{x}(t)$ in the phase space starting from $\mathbf{x}(t_0)$
- Solution: is in the form $\mathbf{x}(t; t_0)$ that defines a family of time trajectories in the phase space. Once we fix t_0, we fix a unique trajectory

Vector fields and flows

How are solutions built? At any point, $\mathbf{f}$ assigns a vector that shows where the point is heading (direction of motion).
If we plot these arrows (vectors) in the phase space, we get can get an idea of how the system evolves.
Flow: $\Phi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is the collection of all trajectories generated by all possible starting conditions.
$\Phi(t, \mathbf{x}_0) = \mathbf{x}(t; \mathbf{x}_0)$
A fundamental theorem guarantees that **two orbits corresponding to two different initial solutions never intersect with each other** , except at equilibrium

Basic Properties

An ODE is linear if $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ (Homogeneous) or $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + b$ (Affine)
Linear ODE enjoys closed form solutions, non-linear ODEs usually not
A system is autonomous if time doesn't appear in expression of $\mathbf{f}$.

- Any n-order ODE can be rewritten as a system of 1st order ODEs in $\mathbb{R}^n$
- Any Non-Autonomous ODE can be rewritten as an autonomous one

Solving!

General form of linear ODE: $\dot{\mathbf{x}} = A\mathbf{x}, \quad x \in \mathbb{R}^n$
A solution is a function $\mathbf{x}(t)$ that satisfies the vector field A.

- Solve $\det(A - \lambda I) = 0$ for λ
- The roots λ_i are eigenvalues of A
- For each λ_i, there exists a non-null eigenvector $\mathbf{u}_i$
- Together they yield one solution: $\mathbf{x}(t) = \mathbf{u}_i e^{\lambda_i t}$
- Each distinct eigen-pair gives ONE independent vector solution
- The general solution is then the combination of these terms: $\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{u}_1 + \cdots + c_n e^{\lambda_n t}\mathbf{u}_n$ (at most n terms)

Important: the above is strictly true only if all eigenvalues are distinct

Matrix Exponential representation: $\mathbf{x}(t) = e^{At}\mathbf{x}(0)$ where $\mathbf{x}(0)$ is a generic initial condition

Exponentials and Asymtotic Behavior

Since the solution is a sum of exponentials, stuff is being pulled in the direction of the eigenvectors, weighted by their corresponding signed eigenvalues.
If the real part of $\lambda_i > 0$, mode i is unstable/diverging.
If the real part of $\lambda_i < 0$, mode i is stable/contracting.
At each point, the solution **mixes** the modes.

Equilibrium points

A state $\mathbf{x}_e$ is an equilibrium state of a system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ if when at a time t_0 the sytem is at $\mathbf{x}_e$ then it stays there FOREVER
For a linear ODE, the equilibrium points are the points of the **Null Space** (solutions to $A\mathbf{x} = 0$) Theres one trivial solution at $\mathbf{x} = \mathbf{0}$ if A is invertible, o.w. inifnitely many solutions

Linear System Classification by Eigenvalues

Eigenvalues	Critical Point	Stability
$r_1, r_2 > 0$	Node (real, distinct)	Unstable
$r_1, r_2 < 0$	Node (real, distinct)	Asymptotically stable
$r_1 r_2 < 0$	Saddle	Unstable
$r_1 = r_2 \neq 0$	Node / Improper node	Same as sign of r_1
$r_{1,2} = \lambda \pm i\mu$	Spiral (focus)	Same as sign of λ
$r_{1,2} = \pm i\mu$	Center	Neutrally stable

Perturbations

For Pure Imaginary Eigenvalue, small perturbations add a tiny real part to the eigenvalues:

- $\lambda > 0$ (positive real part) $\implies$ trajectories spiral outward (unstable spiral).
- $\lambda < 0$ (negative real part) $\implies$ trajectories spiral inward (stable spiral).

For Repeated Real Eigenvalues

- If the eigenvectors are linearly independent, the system stays a node, but may change to a saddle if the signs differents.
- If the eigenvectors are linearly dependent (a degenerate node), a small perturbation will typically turn it into either a spiral (if eigenvalues become complex) or a node (if eigenvalues become distinct real numbers).

Linearization around a Critical Point

Consider a nonlinear system: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n$
Let $\mathbf{x}_0$ be a critical point: $\mathbf{f}(\mathbf{x}_0) = 0$.
Step 1: Linear Approximation
$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + J_{\mathbf{f}}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$
Since $\mathbf{f}(\mathbf{x}_0) = 0$, the linearized system is $\dot{\mathbf{x}} \approx J_{\mathbf{f}}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0), \quad \mathbf{y} = \mathbf{x} - \mathbf{x}_0$
Step 2: Jacobian Matrix $J_{\mathbf{f}}(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{\mathbf{x}=\mathbf{x}_0}$
Step 3: Eigenvalues, Eigenvectors, and General Solution
Compute eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of $J_{\mathbf{f}}(\mathbf{x}_0)$. The general solution of the linearized system is $\mathbf{y}(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t}\mathbf{v}_i, \quad \mathbf{x}(t) = \mathbf{x}_0 + \mathbf{y}(t)$
Step 4: Stability Analysis
Stability is determined by the eigenvalues of the Jacobian $J_{\mathbf{f}}(\mathbf{x}_0)$, just like in the linear case.

Global Behavior and Nullclines

- **Basin of Attraction:** The set of all initial conditions that eventually lead a trajectory to the same stable equilibrium point.
- **Separatrix:** The boundary between different basins of attraction.
- **Isoline:** The set of points where a function takes the same value.
- **Isocline:** The set of points where the vector field has the same **slope**.
- **Nullcline:** A specific type of isocline. It's the set of points where the slope is either zero or infinite (i.e., where one component of the vector field is zero).

Nullcline fun facts:

- The x**-nullcline** is the curve where $\dot{x} = f(x, y) = 0$.
- The y**-nullcline** is the curve where $\dot{y} = g(x, y) = 0$.
- On the x-nullcline, the vector field has no horizontal component, so vectors can only point vertically (up or down).
- On the y-nullcline, the vector field has no vertical component, so vectors can only point horizontally (left or right).
- **Equilibrium points** are exactly at the intersection of the x-nullclines and y-nullclines (since $\dot{x} = 0$ and $\dot{y} = 0$ simultaneously).
- Nullclines divide the phase space into regions. In each region, the sign of $\dot{x}$ and $\dot{y}$ is constant.

Limit cycles

A periodic orbit is just any trajectory that forms a closed loop.
A limit cycle is an isolated closed trajectory. This means neighboring trajectories are not closed; they either spiral into the limit cycle (a stable limit cycle) or spiral away from it (an unstable limit cycle). There's also mixed scenarios (half-stable)
Example: Van der Pol Oscillator

2D is Boring

THM1: Any closed trajectory muyst encolse at least one equilibrium Poincare-Bendixson Theorem: IF a trajectory is trapped in a closed, bounded region R, AND this region R contains no equilibrium points, THEN the trajectory must eventually approach a limit cycle.
In other words, 2D systems cannot have chaos

Chaos and Strange Attractors

3 main types of attractors: Fixed Points, Limit Cycles and Strange Attractors (the last appears only in 3D+)
Properties of strange attractors:
1) Sensitive dependence on initial conditions: two initial conditions very close to each other become very far apart as time goes on (but remain confined in the set that defines the attractor)
2) Fractal dimensions (e.g. 2.06). Non-integer

Example: Lorenz Attractor

For a parameter r=21, trajectories spiral into one of two stable fixed points. For r=28, the fixed points become unstable. Trajectories are still bounded, but they never settle down. They move from one "wing" of the attractor to the other in an aperiodic, unpredictable way.
Chaos: **aperiodic long-term behavior** in a **deterministic** system that exhibits **sensitive dependence on initial conditions**

Iterated Maps

Discrete-time dynamical systems have a state which is only defined at integer steps.
The state is: $\mathbf{x}(n) = (x_1(n), \ldots, x_k(n))$ where k is the dimension of the system and n is an integer step parameter
and the rule is: $\mathbf{x}_n = \mathbf{f}(\mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-m})$
Next state is obtained by directly apply the map $\mathbf{f}$

1D Iterated Maps and Cobweb Plots

We focus on the simplest case: a 1D map $x_n = f(x_{n-1})$

- **Orbit:** sequence of points generated starting at x_0 and keep applying the map
- **Fixed point:** Point that maps to itself: $f(x^*) = x^*$

Sawtooth diagram: Plotting the $(x, f(x))$ diagram. No time shown, all it tells you is if the system is at x now, the next state will be at $f(x)$
Cobweb Plot construction algorithm:

- Draw $y = f(x)$ and $y = x$
- Start at initial point x_0 on horizontal axis
- Move **vertically** to $y = f(x)$ (this is point x_0, x_1)
- Move **horizontally** to $y = x$ (this is point x_1, x_1)
- Move **vertically** to $y = f(x)$ again (this is point x_1, x_2)
- Keep going lil bro

The intersections of $y = f(x)$ and $y = x$ are fixed points. We can analyze stability of a fixed point by looking at the plot.

Stability of fixed points

A linear approximation shows that $\epsilon_{n+1} = f'(x^*)\epsilon_n$. Let $\lambda = f'(x^*)$
Solving: $\epsilon_n = \lambda^n \epsilon_0$. If you know some basics you can infer stability from this alone but I will write a table.

- **Stable (Attracting)**
  - $0 < \lambda < 1$: Monotonic — Converges from one side
  - $-1 < \lambda < 0$: Oscillatory — Zig-zag convergence (alternating sides)
  - $\lambda = 0$: Superstable — Very fast convergence
- **Unstable (Repelling)**
  - $\lambda > 1$: Monotonic — Diverges from one side
  - $\lambda < -1$: Oscillatory — Alternating divergence
- **Marginal**
  - $|\lambda| = 1$: Neutral — Linearization inconclusive

Logistic Map: $x_{n+1} = rx_n(1 - x_n)$ where $x_n \in [0, 1]$ is population, $r \in [0, 4]$ is growth rate

- $1 < r < 3$: The population converges to a single, stable fixed point $x^* = 1 - 1/r$.
- $r = 3$: The fixed point becomes unstable.
- $3 < r < 3.449\ldots$: The system no longer settles to one point. It settles into a stable period-2 cycle, oscillating between two values. This is a bifurcation.

- $r > 3.449\ldots$: The 2-cycle becomes unstable and splits into a stable period-4 cycle. This continues, creating an 8-cycle, 16-cycle, etc., in a period-doubling cascade.
- $r > r^{\infty} \approx 3.5699\ldots$: The cascade finishes, and the system enters the chaotic regime. The orbit becomes aperiodic, never settling down and seemingly random.

This behavior can be summarized in an **orbit diagram**.
- The x-axis is the parameter r.
- The y-axis plots the long-term attractor points for that r.

A **bifurcation** is a qualitative change in the long-term behavior of a system with a smooth variation of a parameter
$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln |f'(x_k)|$
The approximation for λ can be constructed numerically, by iterating the map!
- $\lambda < 0$: For stable fixed points and cycles
- $\lambda > 0$: For chaotic attractors
- $\lambda = 0$: This is the marginal case, which occurs at bifurcation points.
- λ is the same for all points in the basin of attraction of an attractor

A **Cellular Automaton** is a multi-dimensional discrete-time dynamical system that is defined by the principle of locality. properties:
- **System state:** n-dimensional $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$
- **Time:** Discrete updates at time steps
- **Topology:** State components arranged according to a given topology
- **Neighborhood** $N(x_i)$**:** Set of neighbors influencing x_i Example 1D: $N(x_i) = \{x_{i-1}, x_i, x_{i+1}\}$ Example 2D:
  Von Neumann: cell + 4 neighbors (up, down, left, right)
  Moore: cell + 8 surrounding neighbors (3x3 box)
- **Local updates:** Each x_i evolves according to a function of itself and neighbors: $F_i : S(N(x_i)) \to S_i$, where S_i is the possible states of x_i and S are states of neighbors

Lattices and Boundaries:

Infinite/adaptive lattice: The grid grows as the pattern propagates
Finite lattice
- Hard boundary: fixed, edge cells have a fixed state
- Hard boundary: reflective, leftmost (rightmost) cell only diffuse right (left)
- Soft boundary: periodic boundary conditions, edges wrap around

Updating Schedules:

Synchronous, Asynchronous
Some Math

We assume they are homogeneous (same lattice, same N, and same rule F for all cells) and use synchronous updating.
Combinatorics Let: $k = |S|$ is the number of states per cell
M is the number of cells
r is the range $(floor(|N(a)|/2))$

Then: Number of possible state configs: k^M ,Number of possible neighborhood configurations: $k^{|N|}$, Number of possible evolution functions (rules): $k^{\left(k^{|N|}\right)}$
Wolfram Code

For Wolfram's 1D CA: $S = \{0, 1\}$ (so $k = 2$) and the neighborhood is $r = 1$ (i.e., $N(x_i) = \{x_{i-1}, x_i, x_{i+1}\}$, so $|N| = 3$).

Using our formula, the number of rules $= k^{\left(k^{|N|}\right)} = 2^{(2^3)} = 2^8 = 256$.
- **Class 1:** Evolves to a simple, stable, homogeneous state (all 0s or all 1s). Information dynamics: Small changes die; information is lost. Lyapunov exponent: $\lambda \leq 0$
- **Class 2:** Evolves to simple periodic structures (stripes, oscillators). Information dynamics: Small changes may persist locally but do not spread. Lyapunov exponent: $\lambda = 0$
- **Class 3:** Evolves to chaotic, aperiodic patterns (e.g., Rule 30). Information dynamics: Small changes spread out and affect distant regions. Lyapunov exponent: $\lambda > 0$
- **Class 4:** Creates complex, localized, moving structures (e.g., Rule 110 is **Turing complete**). Information dynamics: Small changes may or may not spread; irregular but structured dynamics. Lyapunov exponent: $\lambda > 0$ (tends to 0)

Networks

For Undirected graphs, Average degree: $\langle k \rangle = \frac{2L}{N}$

For Directed Graphs: Average degree (either in or out): $\frac{L}{N}$

Degree Distribution: $p_k = \frac{N_k}{N}$ where N_k is number of nodes with degree k
Connected graph is an undirected graph where there exists a path between every pair of nodes
For directed graphs there is
- Strongly connected if for every pair there is a path
- Weakly connected if every pair there is a path **when you ignore edge directions**

If the largest component in a graph has a large fraction of the nodes, we call it the giant component
- Adjacency list: a mapping from each node to it's neighborhood
- Adjacency matrix: $N \times N$ matrix A such that $A_{ij} = 1$ if link (i, j) exists, and 0 otherwise.
- For undirected graphs they are symmetric (the A)
- Number of walks between nodes i, j can be calculated as follows: $(A^l)_{ij}$ gives #walks of length l.

Average path length: $h = \frac{1}{2E_{\max} \sum_{i,j\neq i} h_{ij}}$ where h_ij is the distance from i to j and $E_{\max} = n(n-1)/2$
Clustering Coefficient
Let L_i be the number of links between neighbors of i. Then the clustering coefficient of node i is $c_i = \frac{2L_i}{k_i(k_i-1)}$
Global clustering coefficient: $C = \frac{1}{N} \sum_{i=1}^{N} c_i$
Real networks are:
- Scale-free (from p_k): they have hubs
- Small world (from h): average distance is very small
- Locally dense (from C): clustering is very high

We will try to have a process that can create a network with these properties
Model 1: Erdos-Renyi (ER)

Generative process: each of the $\binom{N}{2}$ possible edges, an edge is created with probability p **Degree Distribution:** $p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
Mean degree is $\bar{k} = p(N-1)$, Variance is $(N-1)p(1-p)$

For large, sparse networks that simplifies to Poisson distribution: $p_k = e^{-\bar{k}} \frac{\bar{k}^k}{k!}$

Clustering Coefficient: $E[C] = p = \frac{\bar{k}}{N-1}$, i.e. small

Average path length grows as: $O(\log N)$
Verdict:
- Scale-free: NO, p_k is Poisson (should be power law)
- Small world: YES, low h
- Locally dense: NO, C is very low (should be high)

Model 2: Watts-Strogatz (WS) (Small World)

This model was made to fix the clustering problem.
Generative process:
- Start with a regular ring lattice, where each node is connected to m nearest neighbors. Graph starts with high C and h
- Go through each edge, and with prob p rewire one end of the edge to a randomly chosen node.

What happens for different p?
- $p = 0$: Just a regular lattice, high C and h
- $p = 1$: Just a random graph, low C and h
- $p = 0.01$: A few rewired links act as shortcuts, dropping h dramatically (to around $logN$), while keeping a high C

Verdict:
- Scale-free: NO, p_k is peaked at around m
- Small world: YES, low h
- Locally dense: YES, C is high

Model 3: Barabasi-Albert (BA) Scale-Free Model

Generative Process
- **Growth:** The network is not a fixed size. It starts with a small "seed" network, and at each time step, a new node is added.
- **Preferential Attachment:** New node connects to m existing nodes. The probability $\Pi(i)$ of it connecting to an existing node i is not uniform. It is proportional to that node's current degree k_i: $\Pi(i) = \frac{k_i}{\sum_j k_j}$

Verdict:
- Scale-free: YES, naturally produces a power-law degree distribution (p_k around k^{-3})
- Small world: YES, low h
- Locally dense: KINDA, C is a typically lower than in real networks

Importance of a node: Network Centrality Measures

Certain positions within the network give nodes more power or importance. We will explore different types
Classic (Non-Recursive) Centrality Measures
Degree Centrality
The number of other nodes n is connected to.
Meaning: A node with high degree centrality has high potential communication activity
Betweenness Centrality
Number of shortest paths connecting all pairs of other nodes that pass through n
Meaning: A node with high betweenness centrality acts as a "bridge" or "gatekeeper" and has significant control over the flow of information.
Closeness Centrality
Measures how "close" a node is to all other nodes in the network.
$C(n) = \frac{N}{\sum_m^{d(m,n)}}$ where $d(m, n)$ is distance between m and n
Meaning: Efficiency of information spread. It answers the question, "How fast can I reach everyone else from this node?"
Self-Consistent (Recursive) Centrality

A node's importance is determined by the importance of its neighbors. This creates a recursive, self-referential definition.
Eigenvector centrality
The centrality of a node i is the scaled sum of centralities of its neighbors
$c_i = \frac{1}{\lambda} \sum^{c_j}_{j\in N(i)}$
Some math leads it to $A\mathbf{c} = \lambda\mathbf{c}$ where $\mathbf{c}$ is the principal eigenvector of the adjacency matrix
Computing eigvenector centrality
For a network with billions of nodes, we can't just "solve" this equation. We must compute it iteratively.
1. Power Method (Power Iteration)
This is the standard centralized algorithm. It uses iterative matrix multiplication to converge to the principal eigenvector.
- **Initialize:** Choose an arbitrary nonzero vector $\mathbf{c}^{(0)}$ (e.g., all ones).
- **Iterate:** Multiply by the adjacency matrix: $\mathbf{c}^{(t+1)} = A\mathbf{c}^{(t)}$.
- **Normalize:** To prevent numerical overflow, normalize at each step: $\mathbf{c}^{(t+1)} = \frac{A\mathbf{c}^{(t)}}{\|A\mathbf{c}^{(t)}\|}$.
- **Converge:** As $t \to \infty$, $\mathbf{c}^{(t)}$ converges to the principal eigenvector $\mathbf{c}_1$.

2. Gossip Algorithms (Decentralized Power Method)
Gossip algorithms implement the same idea in a distributed and asynchronous way, without a central coordinator.
- Each node i maintains its own estimate $c_i^{(t)}$.
- Nodes periodically *gossip* (push or pull) their current c_i values to their neighbors.
- Each node updates asynchronously: $c_i^{(t+1)} = \sum_{j\in N(i)} c_j^{(t)}$. Over time, these local updates collectively approximate the global power iteration, converging to the same principal eigenvector.

Alpha-Centrality
The idea is that the centrality is a combination of two things:
- Recursive influence from its neighbors (scaled by α).
- An "exogenous" or baseline importance e.

The defining equation is: $\mathbf{c} = \alpha A\mathbf{c} + \beta\mathbf{e}$
Solving for $\mathbf{c}$: $\mathbf{c} = \beta(I - \alpha A)^{-1}\mathbf{e}$
Since $(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \dots$, we can interpret each term as a contribution from paths of increasing length:
- $\beta I\mathbf{e}$ — baseline importance (paths of length 0),
- $\beta\alpha A\mathbf{e}$ — influence from direct neighbors (paths of length 1),
- $\beta\alpha^2 A^2\mathbf{e}$ — influence from neighbors-of-neighbors (paths of length 2),
- and so on for longer paths.

The parameter α controls the extent of influence propagation:
- For small α, only local structure matters (nearby nodes dominate).
- As α approaches $1/\lambda_1$ (where λ_1 is the largest eigenvalue of A), global structure dominates, and the measure converges to **eigenvector centrality**.

Swarm Intelligence is the study and design of Complex Systems/ Multi-Agent Systems that
- Potentically large number of locally interacting, decentralized, and distributed components, a swarm
- Each component has a purpose that contributes to the peformance of the whole
- Under some condiitons, the system displays emergent forms of collective / swarm intelligence

The method of communication defines the type of SI system:
- Point-to-Point: Direct contact (e.g., ant antennation).
- Limited-Range Broadcast: A signal propagates locally (e.g., fish lateral lines detecting waves, visual cues).
- Indirect (Stigmergy): An agent modifies the environment, and another agent responds to that change later.
- Physical Mobility: Agents move through space, and their "network" is their set of current neighbors, which changes constantly.
- Static positioning, state evolution: connection topology and/or positioning in the environment do not change over time.

Particle Swarm Optimization (PSO)

Multi-agent black-box optimization framework inspired by social and roosting behavior of flocking birds
Background: Boids!

Realistic model of coordinated animal motion, agents follow three simple rules:
- Separation: steer to avoid crowding local flockmates
- Alignment: steer towards the average heading of local flockmates
- Cohesion: steer to move toward the average position of local flockmates

Later, a **roost** was added. An attraction point in a simplified Boids-like simulation, such that each agent
- Is attracted to the location of the roost
- Remembers where it was closer to it
- Tells its gang about its closest location to the roost

Eventually, (almost) all agents will land on the roost.
What if the roost is the unknown global optimum (min/max) of a mathematical function? And the "distance to the roost" is simply the quality of the function's value at that point?
Black-box optimization with PSO

We do not know the formula for $f(x)$. All we can do is "query" the function: provide an input x and observe the output $f(x)$. We must find the optimum by intelligently sampling the search space.
PSO is a derivative-free, black-box optimization algorithm.
Core Mechanic (In Words)

Swarm of particles are made, each particle i represents a potential solution.
- Position $\mathbf{x}_i \in \mathbb{R}^n$: Candidate solution of particle
- Velocity $\mathbf{v}_i \in \mathbb{R}^n$: Current direction and speed of particle

Each particle also has a memory:
- pbest, position with the best $f(x)$ particle i ever visited
- lbest, position with the best $f(x)$ any particle in ith social neighborhood ever visited (will be used later)
- gbest, position with the best $f(x)$ any particle ever visited

At each time step, the particle updates its velocity and position based on 3 influences:
- Inertia component: keep going where its going already
- Personal component: pull towards pbest
- Social Component: pull towards gbest

Core Mechanic (In Math) :)

Velocity Update: $\mathbf{v}_i(t+1) = \omega\mathbf{v}_i(t) + c_1 r_1(pbest_i - \mathbf{x}_i(t)) + c_2 r_2(gbest_i - \mathbf{x}_i(t))$ where
- ω is inertia weight (influence of old velocity)
- c_1, c_2 are acceleration coefficients (weighting the "pull" of the personal and social components).
- r_1, r_2 are random numbers, adding stochasticity to the search.

Position Update: Just move. $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1)$
Parameters and Variants

What if only one agent? If left alone, each individual agent would behave like a stochastic hill-climber when moving in the direction of a local optimum, and then it will have a quite hard time to escape it.
Neighborhood Topology
There's no clear way tp know which topology is best.
- gbest (Global Best): Leads to fast convergence but can get trapped in local optima.
- lbest (Local Best): Each particle is pulled toward the best solution found by its immediate topological neighbors. Slower but explores the space more effectively, making it better for complex, multi-modal problems.

Acceleration Coefficients
The balance between c_1, c_2 two defines the swarm's search strategy:
- $c_1 > 0, c_2 = 0$: (Independent). No social influence, set of independent hill climbers
- $c_1 = 0, c_2 > 0$: (Social-only). No personal memory, entire swarm is pulled only towards single best-known point. One stochastic hill-climber
- $c_1 = c_2 > 0$: Attracted to average of pbest, gbest
- $c_2 > c_1$: (More Social). Better for unimodal problems
- $c_1 > c_2$ (More Personal). Better for multimodal problems

Inertia Coefficient
The weight was added to control balance between exploration and exploitation:
- $\omega \geq 1$: velocities increase over time, swarm diverges. Exploration.
- $0 < \omega < 1$: particles decelerate, convergence depends on c_1, c_2. Explotation.

Constriction Coefficient: idk what this is, TODO understand it.
Fully Informed PSO (FIPS)
Like **lbest**, FIPS uses neighborhood information, but more democratically. Instead of moving toward the single best particle, a particle is attracted to a weighted average of all neighbors' personal bests. This slows convergence slightly but reduces the chance of getting stuck in local optima.
Binary / Discrete PSO
PSO can be adapted for binary search spaces ($\mathbf{x}_i \in \{0,1\}^n$):
- Velocity $\mathbf{v}_i$ is continuous and updated normally.
- Each component v_{ij} is interpreted as the probability of a 1.
- Probability is computed via a sigmoid: $s(v_{ij}) = 1/(1 + e^{-v_{ij}})$.
- Positions are updated stochastically: $x_{ij}(t+1) = \begin{cases} 1 & \text{if rand()} < s(v_{ij}(t+1)) \\ 0 & \text{otherwise} \end{cases}$

Optimization

Comparing PSO with single optimizing agents. We saw for PSO if left alone, each agent would behave like a memory-based random searcher, biased toward its own best-so-far.
A conservative blind explorer, with no reliable signal to use for improving.
Hill-Climber Optimizing Agent
Smarter Optimizing Agent. Like climbing Everest in thick fog with amnesia.
- Start at a random point
- Sample the neighborhood of the current point
- Always moves in direction of local imporovement
- Repeat until no neighbor is better (it has reached a peak).

Variants to improve it:
- Stochastic hill-climber: introduces randomness in the selection of neighbor points
- Random Restarts: escape mechanism for local optima

Hill-climbing is powerful for Constraint Satisfaction Problems (CSPs), like n-Queens. It can solve large instances of n-queens ($n = 106$) in a few seconds
Gradient-Based Agents

The Gradient ($\nabla f(\mathbf{x})$)

The gradient is the vector of partial derivatives $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)$, and it points in the direction of the steepest increase of the function f at $\mathbf{x}$.

Gradient Descent Algorithm
To minimize a function, an agent iteratively moves in the direction of the negative gradient:
- Start at an initial point $\mathbf{x}_0$.
- Compute the descent direction: $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$.
- Choose a step size α_k (how far to move).
- Update the position: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$.
- Repeat until the gradient is approximately zero ($\nabla f(\mathbf{x}) \approx 0$), indicating a minimum.

For convex functions we are guaranteed to get to the local and global minimum in a finite (possibly small) number of iterations
For non-convex functions, the final local optimum depends on where we start from
Applications for ML
Optimization function represents the classification Error of a ML model during training, i.e., the total Loss on on training set $\sum_{i=1}^m l(\hat{y}^i(\mathbf{w}^i, \mathbf{x}^i), y^i)$
Gradient Descent Procedure for OLSR
For a linear model with quadratic loss: $E(\mathbf{w}; X, Y) = \frac{1}{2}\sum_{i=1}^N (\mathbf{w}^\top x_i - y_i)^2$
The gradient is: $\nabla E(\mathbf{w}) = \sum_{i=1}^N (\mathbf{w}^\top x_i - y_i)x_i$
Gradient descent update: $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha\sum_{i=1}^N (\mathbf{w}^{(t)\top} x_i - y_i)x_i$

- Initialize $\mathbf{w}^{(0)}$ (e.g., random or zero vector), $t \leftarrow 0$
- Repeat until convergence or maximum iterations N: Compute gradient ∇E, choose step size α, update parameters: $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha\nabla E$, break if $\|\mathbf{w}^{(t+1)} - \mathbf{w}^{(t)}\| <$ tolerance
- Return $\mathbf{w}^{(t+1)}$ as the solution

Batch vs Stochastic GD
- Batch Mode (Regular): Gradient at each step is computed over the entire function, This is accurate but impossibly slow if $m = 1$ billion.
- Stochastic Gradient Descent: Gradient at each step is computed over one single component / data point selected out of the m ones. Takes many more steps (but each step is computationally light). The steps are very fast but "erratic".
- Mini-batch (stochastic) Gradient Descent: Gradient at each step is computed over one subset of data points. This is good balance of speed and stability.

Momentum and Adam

SGD can be slow or get stuck
Momentum (MGD)
Adds an inertia term by combining the current gradient with the previous update, allowing the agent to roll through small bumps and accelerate along gentle slopes: $\mathbf{v}_t = \gamma\mathbf{v}_{t-1} + \alpha\nabla f(\mathbf{w}_t), \quad \mathbf{w}_{t+1} = \mathbf{w}_t - \mathbf{v}_t$
Adam (Adaptive Moment Estimation)
A popular optimizer combining two ideas:
- **Momentum ($\mathbf{m}_t$)**: exponentially weighted average of past gradients (1st moment) to determine direction.
- **Adaptive scaling ($\mathbf{v}_t$)**: exponentially weighted average of past squared gradients (2nd moment) to adapt learning rates per parameter.

Comparing ADAM against PSO

One core difference: ADAM is a Local Optimizer, PSO is a Global optimizer! In general Adam worse, unless the shape is nice (Adam did well on ellispoid, got smoked in griewank, rastrigin, schwefel)

Hybrid?

If PSO is a great "explorer" and Adam is a great "exploiter," the obvious solution is to combine them.

The Strategy:

- PSO Phase: Let the PSO swarm run for several iterations. The particles spread out and find the most promising valleys.
- Local Search Phase: Take the best particles and run a local optimizer starting from their positions.
- Cooperation: The local optimizer will find the exact bottom of that valley. This new accurate position is then fed back to the swarm as the new gbest.

Local Search

Local Search is a fundamental optimization framework

Core Concept

They are iterative algorithms that each step consider a single current state, and try to improve it by moving to one of its neighbor state that has a better evaluation score

- Search for an improvement is done in local neighborhood only
- Hill climbing, gradient are specific instances of LS

Local Search methods work with complete problem states/solutions, s, i.e., all necessary variables are assigned

Neighborhoods in discrete domains

- 1-flip N for 0-1 vectors
- 2-swap N for permutation vectors
- k-exchange N, $N(s)$ of a state s is the set of states s' that differ from s up to k solution components/values

A small neighborhood is fast to search at each step, but the search is less powerful.

A large neighborhood is slower to search but can find better solutions.

Local Search for TSP: Neighborhood Structures

Local Search for the TSP relies on defining a neighborhood

2-opt

- Select two non-adjacent edges, e.g., (A, B) and (C, D).
- Remove them, creating two separate paths.
- Reconnect the paths as (A, C) and (B, D).

Effect: Reverses the segment between the two edges, effectively "uncrossing" the tour.

Complexity: $O(N^2)$ possible edge pairs.

3-opt

- Delete three edges, splitting the tour into three paths.
- Reconnect them in one of $2^3 = 8$ possible configurations (including the original).

Effect: Allows both reversed and non-reversed segments, enabling solutions unreachable by 2-opt.

Complexity: $O(N^3)$ possible triples of edges.

Note: Necessary for Asymmetric TSP (ATSP) where $d(A, B) \neq d(B, A)$.

4-opt (Double Bridge)

- Does not revert tour segments (useful for ATSP).
- Achieves $O(N^2)$ complexity with clever implementation.
- Commonly combined with 2-opt for improved performance.

Solution Modification vs Construction

- Solution Modification (Local Search): You start with a complete solution (e.g., a full TSP tour) and make local modifications to iteratively improve it.
- Solution Construction: You start with an empty solution and incrementally build it, one piece at a time, until it is complete.

Cyber-Physical Multi-Agent Systems (CP MAS)

Systems of multiple interacting cyber-physical agents (robots, sensors) combining mechatronics, communication, and computation.

Design Objectives of Robot Swarms

- **Scalability:** Performance degrades gracefully with swarm size
- **Robustness:** Tolerates individual failures / environmental changes
- **Flexibility:** Handles different tasks/environments
- **Distributed Control:** Local decision-making, no single point of failure

Swarm Intelligence Approach

- Simple individual controllers, complex interactions
- Local interactions / decentralized communications
- Emergence and self-organization

Challenges

- Hard to predict behavior or provide formal guarantees
- Finite-time performance hard to characterize
- Efficiency may be mediocre under full decentralization

CPMAS Taxonomy

- **Members:** Homogeneous **or** Heterogeneous
- **Coupling:** Loosely **or** Tightly coupled
- **Goals:** Non-cooperative **or** Cooperative
- **Control:** Centralized **or** Decentralized
- **Coordination:** Explicit **or** Implicit

Note: Real systems may combine any of these features.

Communication Issues

- Global schemes do not scale
- Need ad hoc networking in infrastructure-less environments
- Decide what to communicate, how frequently, and to whom

Example: AntHocNet (MANET routing via ACO)

- Monte Carlo sampling and full-path updates via ant agents (reactive)
- Local periodic exchange of routing info and pheromone updates (proactive)

Task Allocation is the problem of deciding "who does what, where, when and how" in a multi-robot system.

It involves finding a mapping / allocation A: T → R, which assigns a set of n_t tasks (T) to a set of n_r robots (R).

The goal is to find an assignment that maximizes overall global system utility.

Division of labor = the pattern or state resulting from that process (how work is distributed)

Tasl allocation can be:

- Explicit: a central controller explicitly assigns task 1 to robot A, etc
- Emergent: a division of labor emerges from local interactions, without any central controller

Formal Definition of MRTA

We are given:

- Set of robots, R
- Set of tasks, T
- $\mathtt{R} = 2^R$ set of all possible robot subteams
- A utility function: $U_{rt} = Q_{rt} - C_{rt}$ for example, where Q is quality and C is cost

We need to find an allocation A (mapping from tasks to robots) that maximizes global objective $U(A)$

MRTA Taxonomy

- **Robot Tasking (R):** Single-Task (ST) **or** Multi-Task (MT) — Can a single robot work on one task (ST) or multiple tasks simultaneously (MT)?
- **Task Type (T):** Single-Robot (SR) **or** Multi-Robot (MR) — Does a task require one robot (SR) or a coalition/team of robots (MR)?
- **Assignment Time (A):** Instantaneous Assignment (IA) **or** Time-Extended Allocation (TA) — Is the assignment instantaneous (IA) or does it involve planning a schedule/route over time (TA)?

Optimization Models for MRTA Problems

The "Hello World": ST-SR-IA

Linear Assignment Problem (LAP) Solution: The optimal assignment can be found in polynomial time using the Hungarian Algorithm with a complexity of O(n^3), (where n is the number of robots/tasks).

- **Adding Task Priorities:** Tasks have priority weights w_t indicating importance, added to the objective function. *Insight:* Weighted assignment maximizes total utility $\sum_r \sum_t w_t U_{rt} x_{rt}$.
- $|R| \neq |T|$ **(Unequal sets):** Number of robots and tasks differ. Use dummy robots/tasks to balance. *Insight:* Two IA rounds preserve polynomial-time solvability without affecting optimal real assignments.
- **Iterated Assignment:** Utilities or task info change over time → recompute or adaptively update assignments. *Insight:* BLE (2001): 2-competitive (≥ 50% of optimal); L-ALLIANCE (1998): learns best assignments iteratively.
- **Online Assignment:** Tasks revealed one-by-one; robots may not be reassignable. *Insight:* MURDOCH (2002): greedy, 3-competitive; best possible for online assignment without reassignment.

Other MRTA Scenarios

- **ST–SR–TA:** Find disjoint routes or schedules for multiple robots to visit all tasks, minimizing total travel cost. *Model/Complexity:* Multiple Traveling Salesperson Problem (mTSP), NP-Hard.
- **MT–SR–IA:** Assign multiple tasks to a single robot, constrained by robot capacity/resources, to maximize total utility. *Model/Complexity:* Generalized Assignment Problem (GAP), NP-Hard.
- **MT–SR–TA: VRP:** Vehicle Routing Problem (VRP). Find optimal routes for a fleet of vehicles to serve a set of customers/tasks, including constraints like limited capacity and time windows. *Model/Complexity:* VRP, NP-Hard.
- **ST–SR–TA + Dependencies:** ST–SR–TA with precedence constraints (e.g., Task A must be completed before Task B can begin). *Model/Complexity:* Complex Scheduling / Precedence-Constrained Scheduling, NP-Hard.
- **Traveling Salesman Problem (TSP):** Find the shortest single tour that visits a set of n tasks (cities) exactly once and returns to the origin (single-robot route planning). *Model/Complexity:* TSP, NP-Hard ($O(2^n)$).
- **Vehicle Routing Problems (General):** Generalization of mTSP, where vehicles must respect real-world constraints like capacity limits, time windows, and precedence. *Model/Complexity:* VRP, NP-Hard.

Orienteering Problems (OP)

- **Classic Orienteering Problem (Single Vehicle):** Also called Selective TSP, Maximum Collection, or Bank Robber Problem. Maximize profit collected within a route duration constraint.
- **Profitable Tour Problem:** Maximize profit minus travel cost for a single vehicle.
- **Prize Collecting Traveling Salesman Problem (PCTSP):** Minimize travel cost while achieving a given route profit, for a single vehicle.
- **Team Orienteering Problem (Multiple Vehicles):** Also called Multiple Tour Maximum Collection Problem. Maximize total profit across multiple vehicles under route duration constraints.

Set-Based Formulations

- **Set Covering:** Select a subset of activities so that all requirements are covered at least once, minimizing cost. → Models MT–MR–IA (robots may appear in multiple coalitions). $\min \sum c_j x_j \quad s.t. \sum a_{ij} x_j \geq 1$
- **Set Packing:** Select disjoint subsets to maximize profit — each task covered by *at most one* robot. → Models ST–MR–IA (exclusive task allocation). $\max \sum p_j x_j \quad s.t. \sum a_{ij} x_j \leq 1$
- **Set Partitioning:** All tasks covered *exactly once* by disjoint subsets. → Base for routing/scheduling models (e.g., VRP). $\sum a_{ij} x_j = 1$

Set Models and Coalition Formation in MRTA

- **Set Covering:** All tasks covered, robots may overlap → models **MT–MR–IA**.
- **Set Packing:** Maximize unique task coverage (no overlap) → models **ST–MR–IA**.
- **Set Partitioning:** All tasks done once, strict non-overlap → used in routing/scheduling (**VRP, TSP**).

Task allocation often combines:

- *Selection* — choosing which agents or coalitions perform which tasks (set-based).
- *Ordering* — sequencing or scheduling tasks for each agent (routing/scheduling).