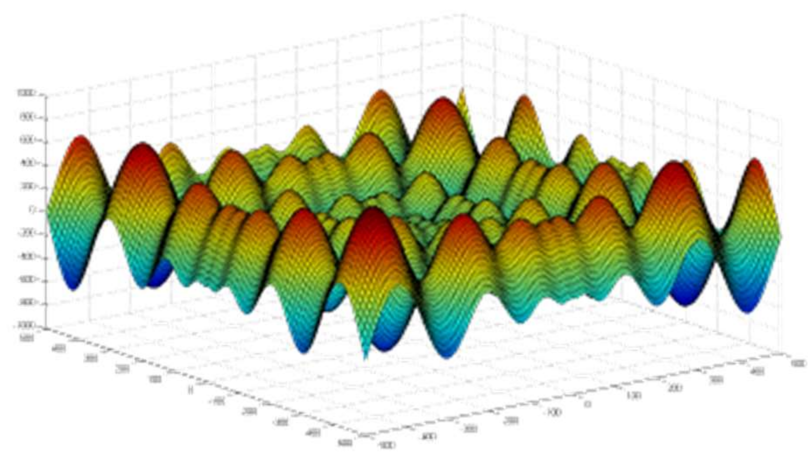
China Winter School on LISA MLDC – Lecture 3

Soumya D. Mohanty

The University of Texas Rio Grande Valley

Terminology

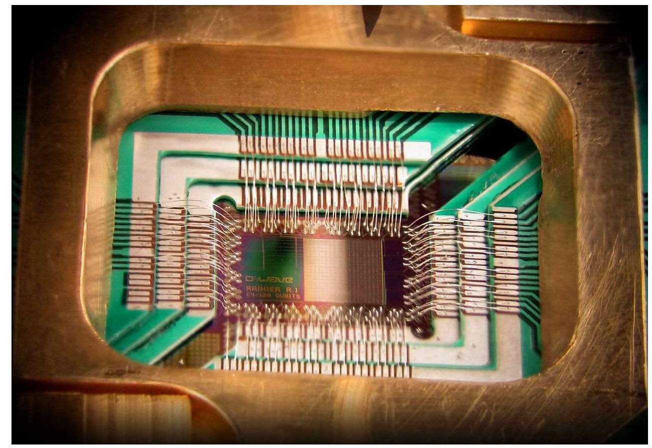
- *Fitness function* $f(x), x \in \mathbb{R}^n$
    - Convex: Local optimizer will work (e.g., gradient descent)
    - **Non-Convex (multi-modal): Local optimizer will not work**
- $D \subset \mathbb{R}^n$: constraint space (Subset of $\mathbb{R}^n$ in which global optimum of $f(x)$ is to be found)
    - Also: "*Search space*", "feasible space"
- From here on, only consider the *minimization* problem



Biology Inspired Global Optimizers

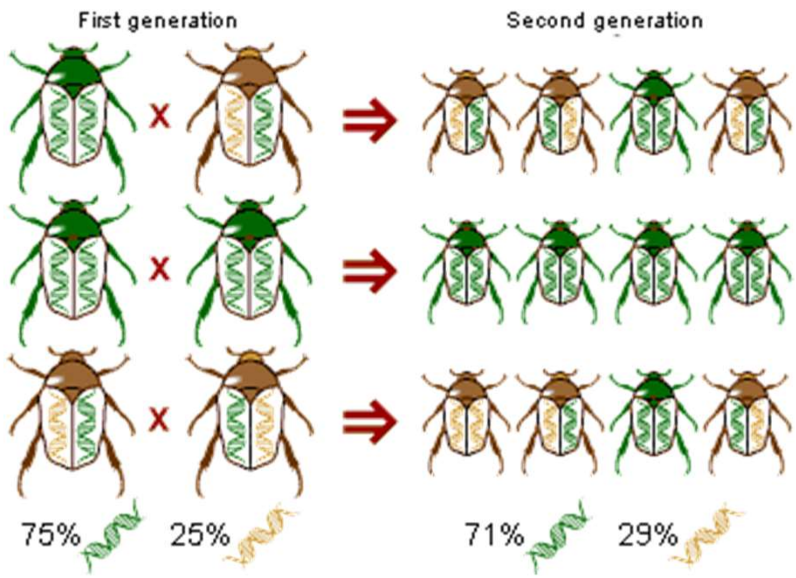
Optimization based on biological systems

- Biological systems seek an optimum living condition ("fitness")
  - Optimization over a high-dimensional parameter space
  - Extremely rugged fitness function ("Fitness landscape")
- Multi-agent + information sharing nature
  - More efficient exploration of search space
  - Exploitation: information sharing leads to recruitment of agents for exploitation
- Simulated Annealing: optimization based on a physical system
  - "Temperature" controls:
    - Extent of random jumps in search space
    - Non-zero acceptance probability for jump to a point with worse fitness
  - Not multi-agent, no information sharing
    - Parallel tempering→ information exchange→better performance



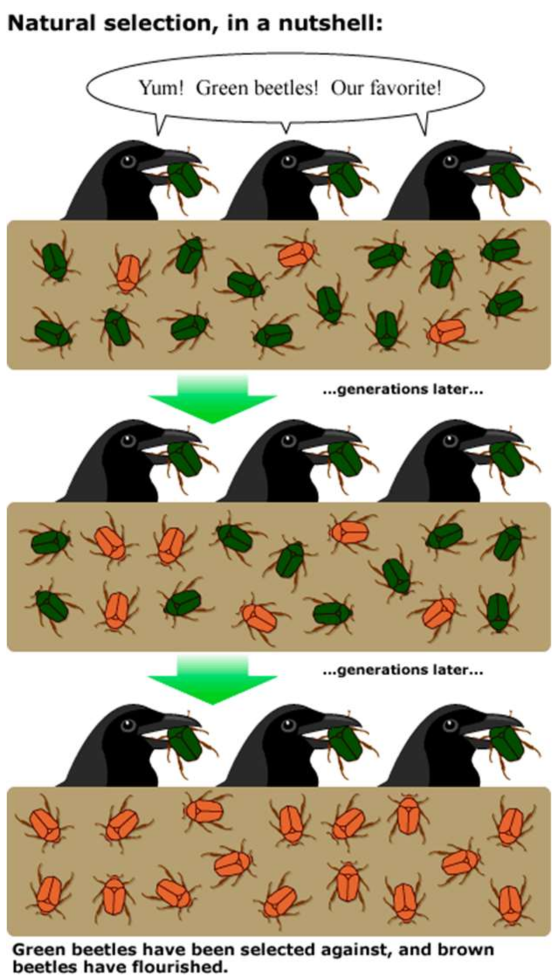
Quantum Annealing? (D-Wave systems)

Genetic Algorithm: Based on evolution

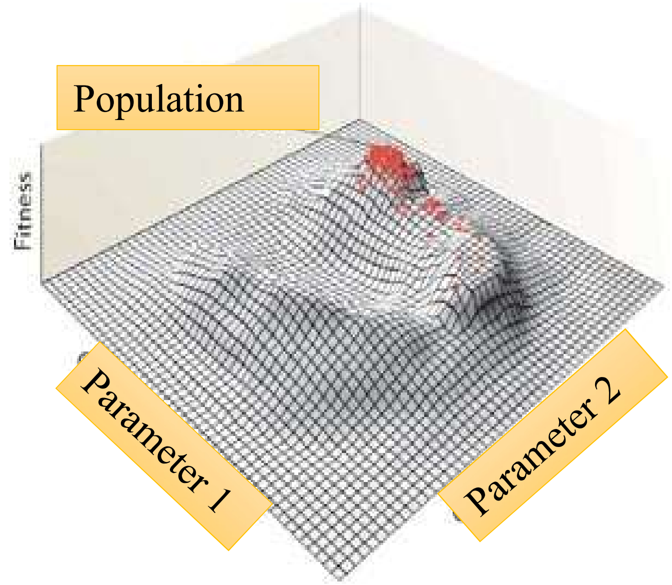


- An individual is described by many parameters
- Individuals in a species: variation in the values of all the parameters
- Natural variation is produced by a variety of mechanisms
- Nature invented sexual reproduction to greatly increase the range of variations
- Mixing of genes from different parents

- Natural selection



http://evolution.berkeley.edu/evolibrary/search/imagedetail.php?id=380&topic_id=&keywords=



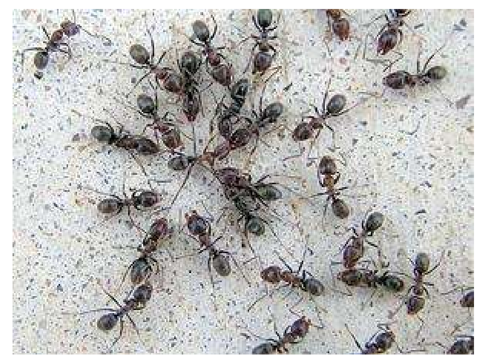
- The probability of survival and reproduction are more for individuals with higher fitness
- Higher fitness individuals leave behind more offspring on the average
- Over many generations, the average fitness of the population increases

Meta-heuristics

- Within the class of GA, there are a large number of methods that differ in the specific choice of cross-over, mutation, selection schemes (e.g., Roulette selection, tournament selection)
- But the basic idea behind all of them is to copy natural evolution
- Here, natural evolution is the Meta-Heuristic and the different methods are realizations of this meta-heuristic

Swarm Intelligence

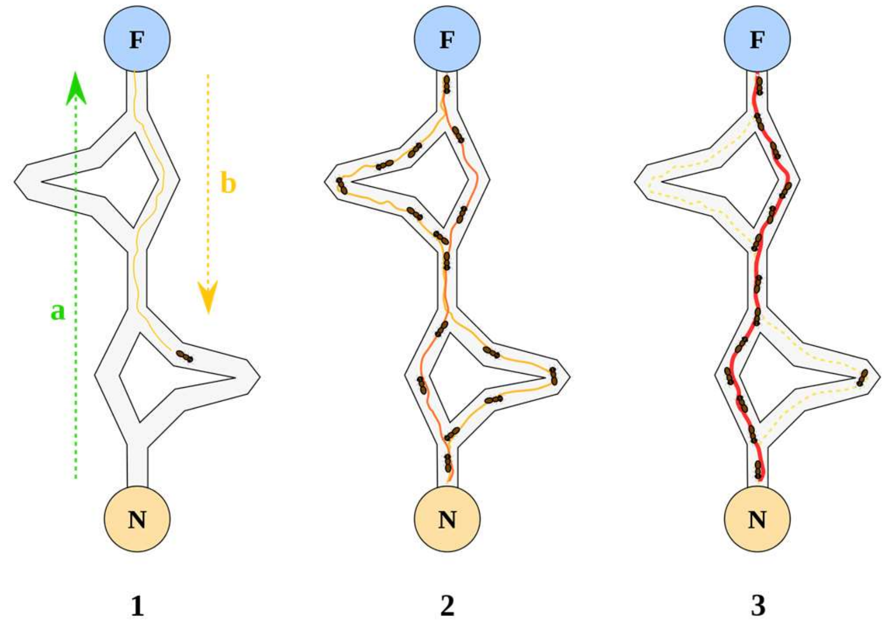
- Multi-agent algorithms with communication of information between agents
  - Individual agents follow simple rules
  - Swarm as a whole has an emergent behavior
- Agents do not die or are born but move around in the search space
- Particle Swarm metaheuristic
  - Based on the behavior of animal groups
  - Example: Bird Flocks
- Ant Colony metaheuristic
  - Based on the foraging behavior of ants
  - Find the shortest path between two points on a graph ("nest" and "food")



http://www.greenbiz.com/sites/default/files/imagecache/wide_large/0909AntsH.jpg

Ant Colony Optimization

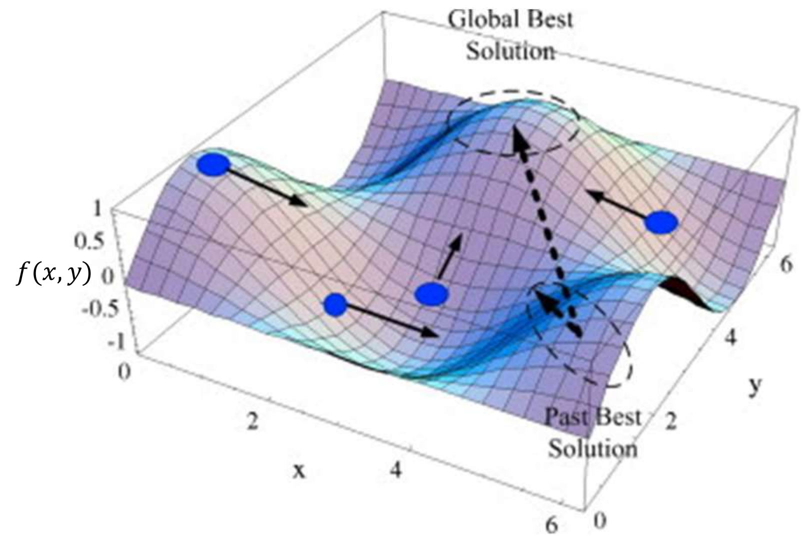
- Based on the foraging behavior of ants
- Each ant leaves a “pheromone” trail
- More pheromone attracts more ants
- Better suited to combinatorial optimization (like GA)
  - Shortest path problems



http://www.sciencedirect.com/science/article/pii/S0142061515005840

Overview of Particle Swarm Optimization

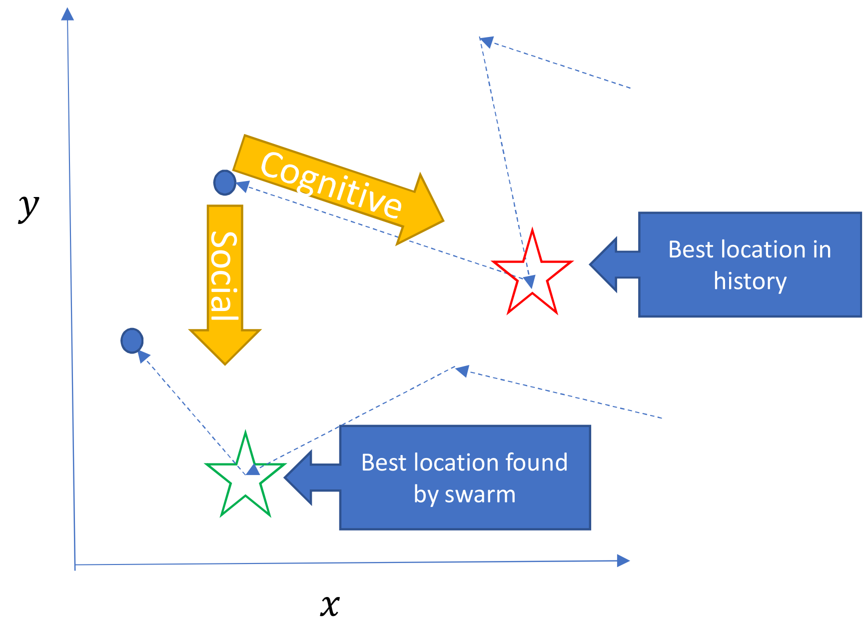
- PSO uses a population ("swarm") of searchers (called "particles")
- Each particle has a "velocity" vector
- The velocity vector tells a particle where to move next
- The velocity vector is updated iteratively
- Many different variations under the same metaheuristic



https://www.mathworks.com/matlabcentral/fileexchange/43541-particle-swarm-optimization--pso-?requestedDomain=www.mathworks.com

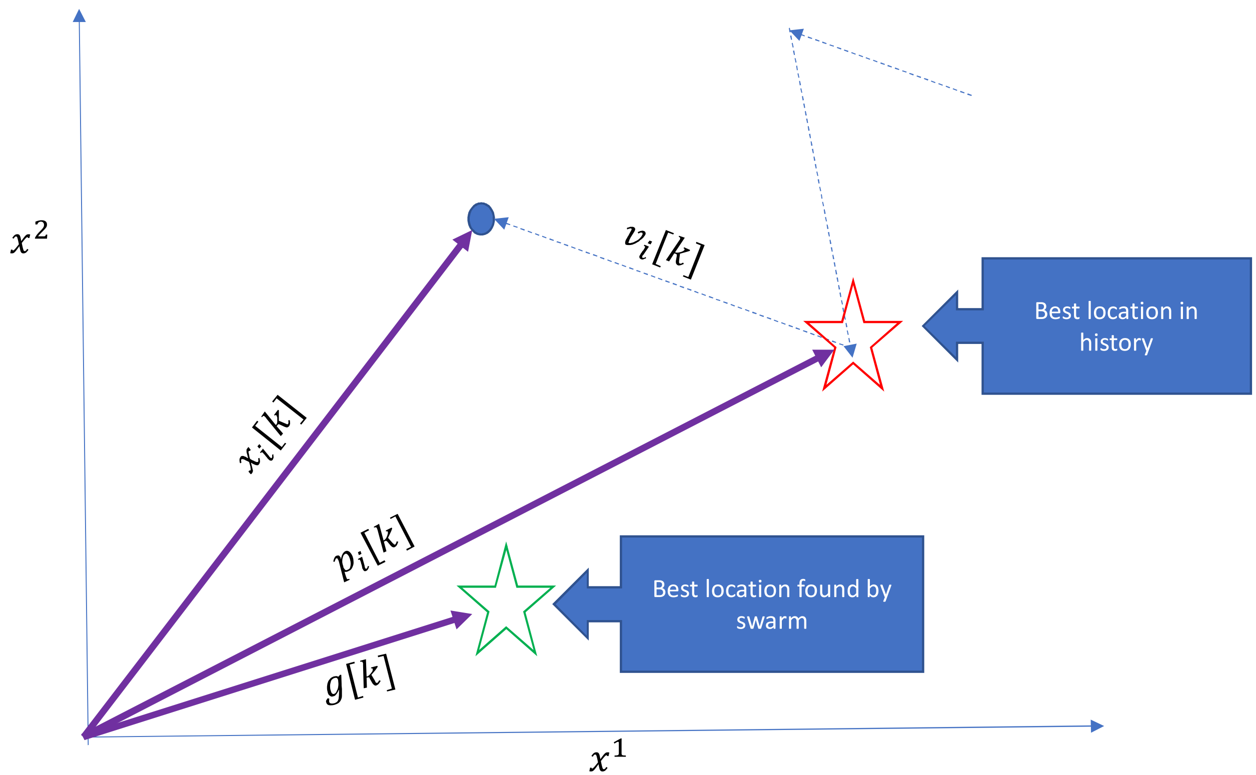
Particle Swarm Optimization

- Exploration trajectory of a particle is guided by three components
    - Inertia: continue along the same trajectory
    - Cognitive: attraction towards the best fitness location it has found so far
    - Social: attraction towards the best fitness location found by its neighbors
- PSO is better suited to optimization in a continuous search space
    - $\min_x f(x)$ where $x \in \mathbb{R}^d$



http://armedwithvisions.com/wp-content/uploads/2011/07/flock-birds.jpg

Notation



x^2
$v_i[k]$
Best location in history
$x_i[k]$
$p_i[k]$
Best location found by swarm
$g[k]$
x^1

Velocity update

- Update velocity:

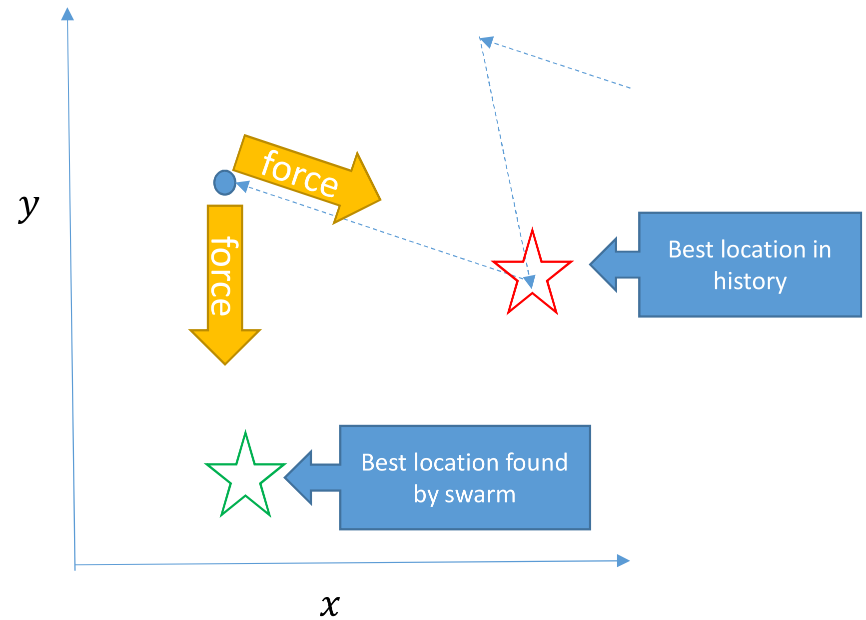
$$v_i^j[k+1] = w\, v_i^j[k] + c_1 r_{1,j}(p_i^j[k] - x_i^j[k]) + c_2 r_{2,j}(g^j[k] - x_i^j[k])$$

- Update position: $x_i^j[k+1] = x_i^j[k] + v_i^j[k+1]$

c_1, c_2 : “acceleration constants”
w : “inertia”
$r_{1,j}, r_{2,j}$: independent random numbers with uniform distribution in $[0,1]$

Velocity update

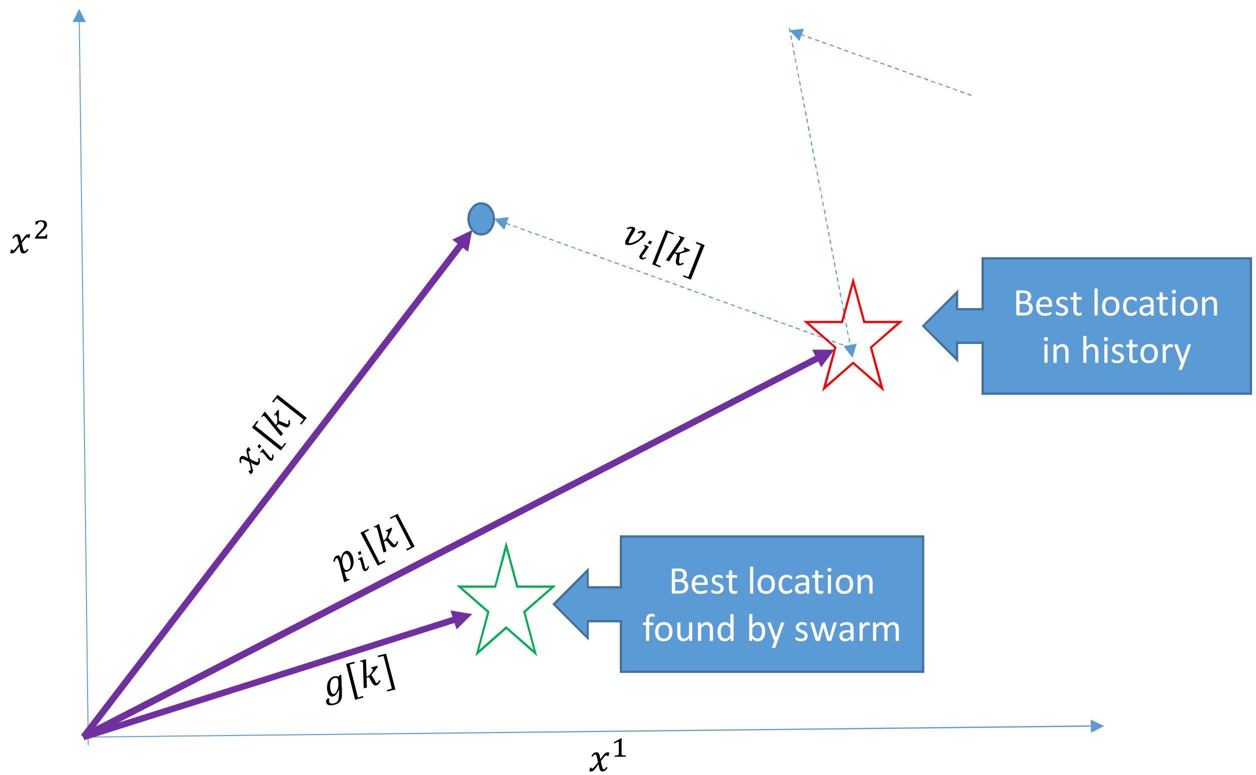
A particle explores the search space randomly but constantly feels an attractive force towards the best location it has found so far and the best location found by the swarm so far



Basic terms

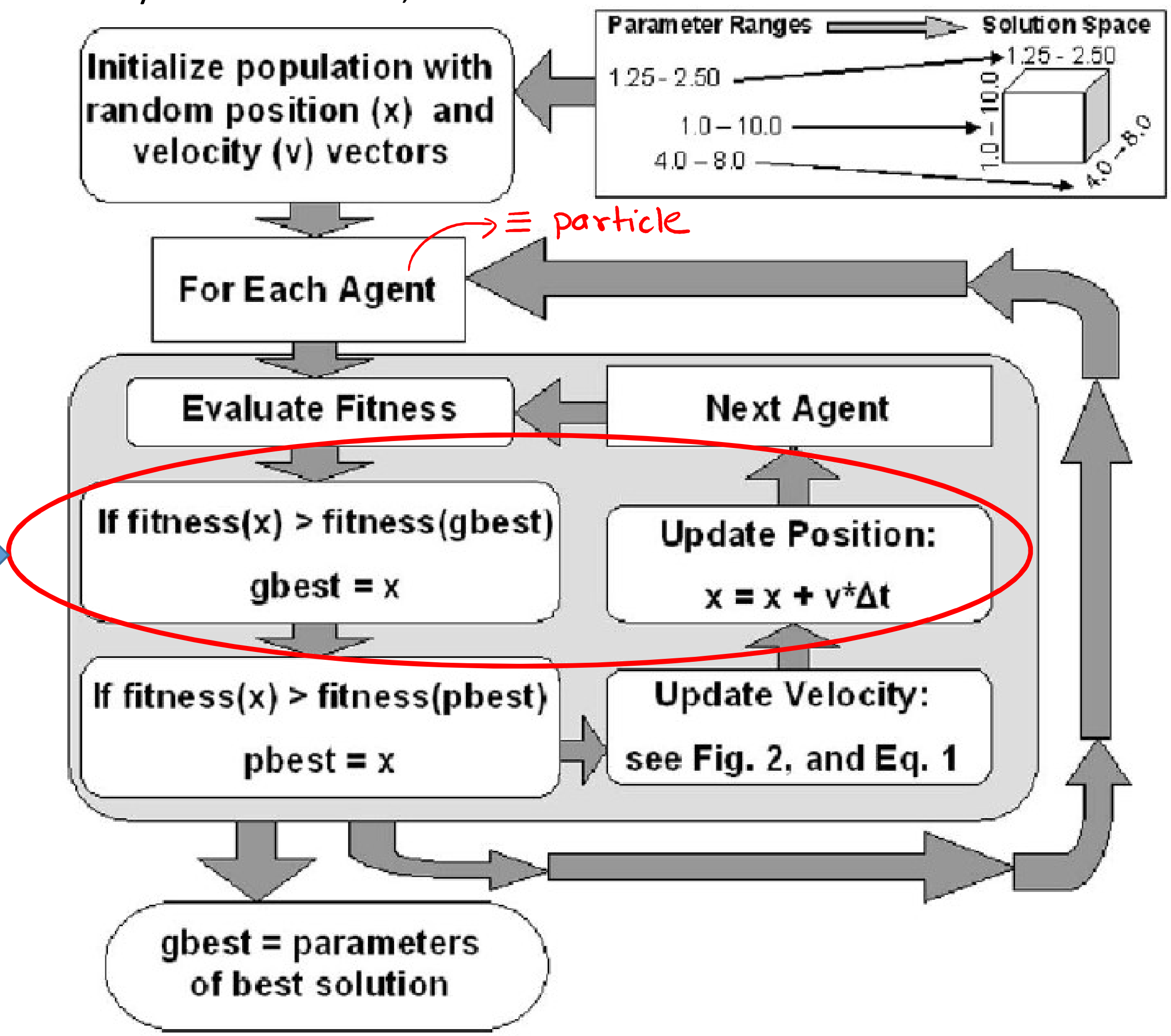
Term	Definition
Particles	Set of locations at which the fitness function is calculated in each iteration
$x_i[k]$	• Position of i^{th} particle in the k^{th} iteration • The components of the position are $x_i^j[k]$
$v_i[k]$	• Velocity of the i^{th} particle in the k^{th} iteration • The components of the velocity are $v_i^j[k]$
pbest : particle best location ($p_i[k]$)	The location of the best fitness value found by the i^{th} particle from the first iteration up to and including the current iteration
gbest : best location found by the swarm ($g[k]$)	The location of the best fitness value found by all the particles from the first iteration up to and including the current iteration
Maximum velocity v_{max} (“Velocity Clamping”)	$v_i^j[k] \in [-v_{max}, v_{max}]$

Notation



x^2
x^1
$x_i[k]$
$v_i[k]$
$p_i[k]$
$g[k]$
Best location in history
Best location found by swarm

Robinson and Yahya Rahmat-Samii, 2004



Synchronous vs asynchronous update: Change gbest after/before evaluating all fitness values

PSO Dynamical Equations

- Update velocity:

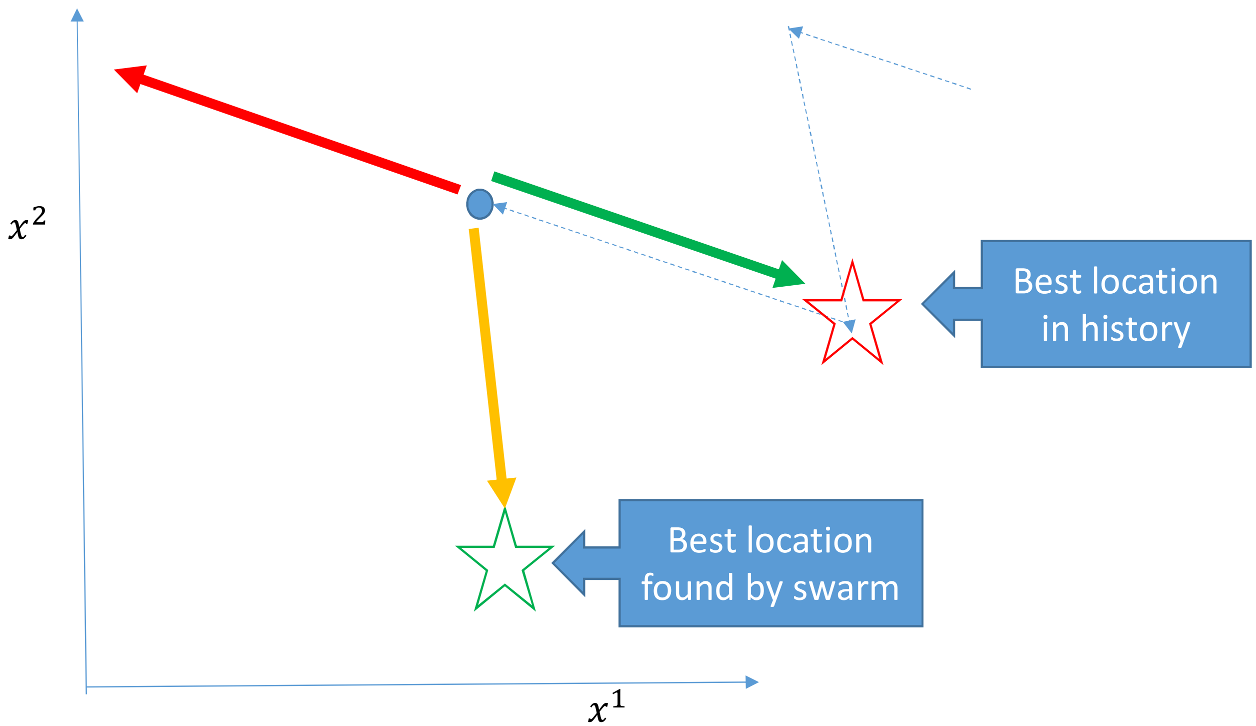
$$v_i^j[k+1] = w\, v_i^j[k] + c_1 r_{1,j}(p_i^j[k] - x_i^j[k]) + c_2 r_{2,j}(g^j[k] - x_i^j[k])$$

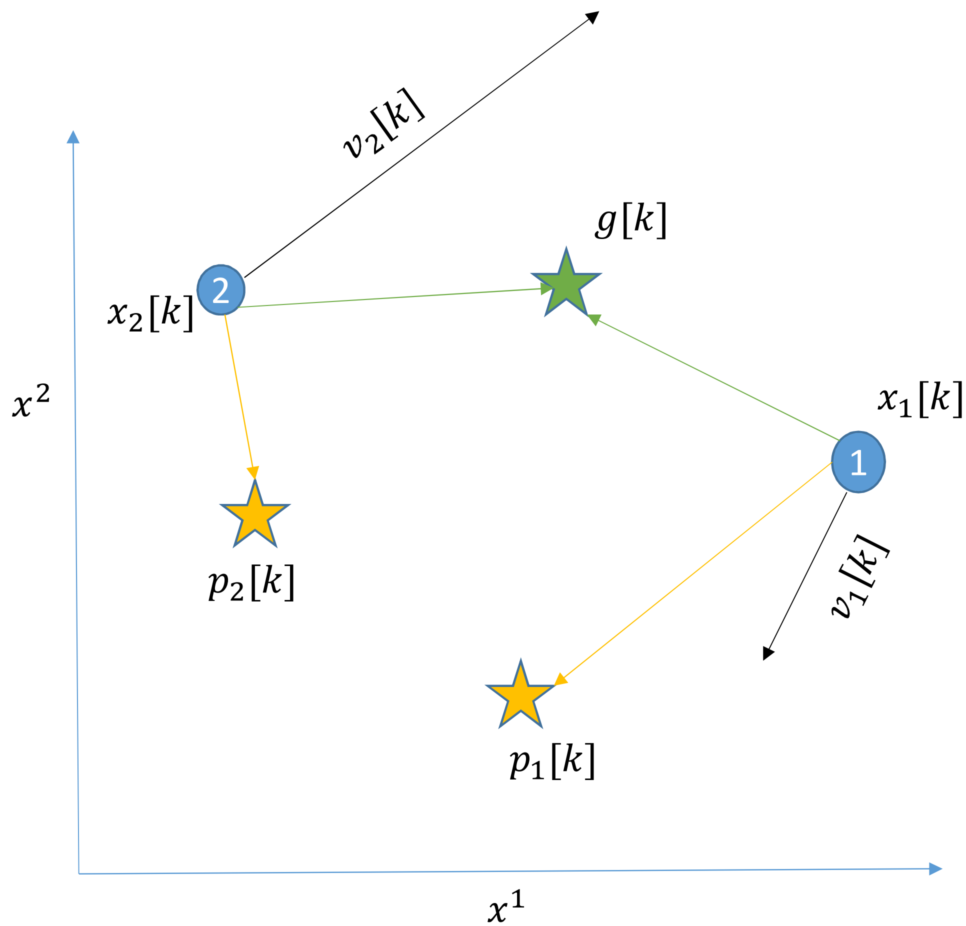
- Update position: $x_i^j[k+1] = x_i^j[k] + v_i^j[k+1]$

c_1, c_2 : “acceleration constants”
w : “inertia”
$r_{1,j}, r_{2,j}$: independent random numbers with uniform distribution in $[0,1]$

Interpretation

$$v_i^j[k+1] = w\, v_i^j[k] + c_1 r_{1,j}(p_i^j[k] - x_i^j[k]) + c_2 r_{2,j}(g^j[k] - x_i^j[k])$$




$v_2[k]$
$g[k]$
$x_2[k]$
2
x^2
$x_1[k]$
1
$v_1[k]$
$p_2[k]$
$p_1[k]$
x^1

$$v_i^j[k+1] = w\, v_i^j[k] + c_1 r_{1,j}(p_i^j[k] - x_i^j[k]) + c_2 r_{2,j}(g^j[k] - x_i^j[k])$$

Simple model of flocking behavior

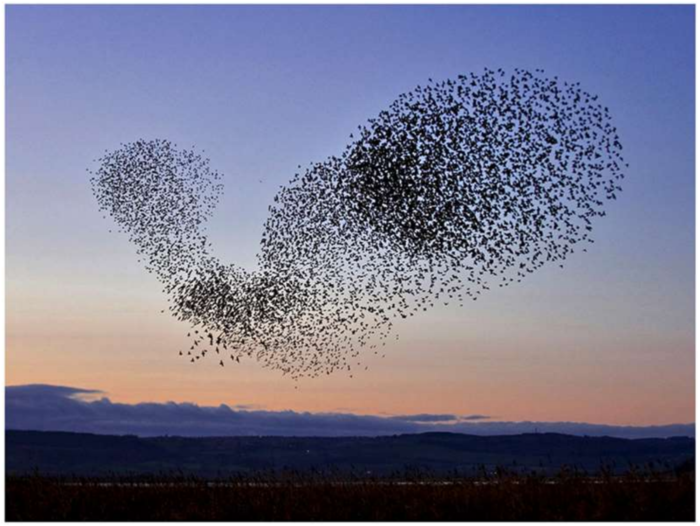
$c_1 r_{1,j}(p_i^j[k] - x_i^j[k])$: “Cognitive Term”

$c_2 r_{2,j}(g[k] - x_i^j[k])$: “Social Term”

$w\, v_i^j[k]$: “Inertia Term”

w is called the inertia weight

c_1, c_2 : acceleration constants (strengths of Cognitive and Social components)



Bird flock
We assume that each bird follows simple rules but the flock has an emergent behavior