Algorithm	Algorithm class	population size	Key idea // how it works	Internal parameters / information
Nelder-Mead	Hill climber	μ > 1	– Initialize simplex of n+1 points – Reflect through centroid to find x_r – Depending on quality of x_r, apply contraction, expansion or shrinking operator	– Centroid point x_0
BFGS	Hill climber	μ = 1	– quasi-Newton method that approximates Hessian matrix – Appr. Hessian is updated through gradient calculations	– Hessian appr. matrix B_k – Step size α_k
SANN	Trajectory (exploring)	μ = 1	– Escape local optima through probabilistic selection – Probability to accept deteriorations decreases over time – 'temperature' determines probability, updated with cooling scheme	– Current temperature T_k
VNS	Trajectory (systematic)	μ = 1	– Apply local search method in the neighbourhood of initial point – If no improvement is made, move to further away neighbourhoods	– Current neighbourhood index k
Tabu search	Trajectory (systematic)	μ = 1	– Store already made moves in tabu list – By prohibiting previously made moves, local optima can be escaped	– Tabu list
(H)MLSL	Trajectory (systematic) HMLSL: Hybrid MLSL+DE	μ > 1	– Create a reduced sample with best candidates from randomly sampled points – Perform local search for best points in cluster (no better points within critical distance r_k)	– Critical distance r_k
DE	Population (classic)	μ ≥ 4	– Apply mutation and recombination to find new candidates – Only accept improvements – Step size is controlled by difference calc. in mutation operator	– n.a.
PSO	Population (classic)	μ > 1	– Create a swarm of particles that move over solution landscape with velocity – particles have access to their all-time best position and the best position found by its neighbours – Thereby, momentum, social forces and cognitive forces influence particle movement	– Previous best position $p_{best,I}$ at p_i – Current best global position in N, $g_{best,i}$ at g_i – Velocity for each particle, v_i
CMA-ES	Population (model-based)	μ > 1	– New candidates are samples from multivariate normal distribution – Covariance matrix is continuously updated with evolution paths – Covariance matrix rotates normal distribution	– C: Covariance matrix – m: Current centre of mass – σ: global step size

Algorithm	Advantages	Disadvantages
Nelder-Mead	✓ Does not require or compute derivatives	✗ Sensitive to scaling ✗ Hard to determine initial simplex ✗ Can converge to non-stationary points
BFGS	✓ Less computational effort since Hessian matrix is not computed, only approximated ✓ Works on non-convex functions	✗ Only applicable for functions where gradients and second order derivatives are available (but also acceptable performance even for non-smooth optimization instances)
SANN	✓ Compared to simple hill climbers, SANN has the ability to escape local optima ✓ Simulated Annealing guarantees convergence upon running sufficiently large (infinite) number of iterations	✗ Difficult to initialize temperature ✗ Difficult to decide on cooling scheme
VNS	✓ Ease of implementation ✓ Can be combined with other search heuristics ✓ Insights into the reasons for good/bad performance	✗ Difficult to define neighbourhood structure
Tabu search	✓ Historic information is used to find good solutions, superior to random processes ✓ Escape local optima ✓ Intensify and diversify search with intermediate and long term memory	✗ Difficult to decide on neighbourhood structure ✗ Find balance between intensification and diversification ✗ Keeping memory with increasing dimensionality and iteration numbers
(H)MLSL	✓ Seems to perform exceptionally well during the last phase of optimization (exploitation)	✗ Only for a moderate size of dimensions ✗ Difficult to decide on initial parameters N and γ
DE	✓ Only a few parameters to tune ✓ Search scales automatically from global to local	✗ Dependence on initial points ✗ No automatic scaling back from local to global ✗ Requires decoding functions for discrete values
PSO	✓ Easy to implement ✓ Able to solve many real-world application problems ✓ competitive for hard multi-dimensional non-linear functions ✓ Deals with multimodality	✗ Without inertia weights or v_{max}, velocities tend to explode ✗ Difficult to conduct theoretical analysis
CMA-ES	✓ For ill-conditioned or rugged functions ✓ Derivative-free, making it feasible on non-smooth, even non-continuous problems, as well as multi-modal and or noisy problems ✓ Works good on high dimensions, $d \geq 10$ ✓ fast (log-linear) convergence and possibly linear scaling with the dimension	✗ Better solutions are available for small dimensions ($d \ll 10$), partly separable problems, problems with cheap gradients, small running time budgets ($f_{eval} < 100n$)

Algorithm	Tuneable parameters	Number tuneable parameters	Implementation available
Nelder-Mead	Reflection coefficient α Expansion coefficient γ Contraction coefficient ρ Shrink coefficient σ	4	
BFGS	Initial selection of x0 und B0	2	
SANN	Initial temperature T_0 Cooling scheme / schedule For exponential cooling: factor α	3	
VNS	Neighbourhood structure // distance measure The maximum index of neighbourhoods, k_{max}	2	
Tabu search	Neighbourhood structure // distance measure Tabu list size	2	
(H)MLSL	Number of points per iteration, N Size of reduced sample, given by gamma γ (The budget of function evaluations for local search, in HMLSL: 10%)	3	No
DE	Crossover probability Cr Population size Np Scaling factor F	3	
PSO	Acceleration factors ϕi If applied: Inertia weight ω	2	
CMA-ES	**Population size lambda λ**; Number of parents mu μ; recombination weights w(i…λ); **Initial step size σ;** cc, decay rate for the evolution path; c1, learning rate for rank-one update of C ; cμ, learning rate for rank-μ update of C; cσ , decay rate of the evolution path; dσ, damping for σ-change	9 // **2**	Yes

Sources / Literature

CMA-ES

Tutorial slides (gecco2013-CMA-ES-tutorial)

Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaption, Hansen & Ostermeier (1996)

Completely Derandomized Self-Adaptation in Evolution Strategies, Hansen & Ostermeier (2001)

Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES), Hansen et. Al (2003)

Wikipedia: https://en.wikipedia.org/wiki/CMA-ES#Example_code_in_MATLAB/Octave

Overview: http://www.cmap.polytechnique.fr/~nikolaus.hansen/cmaesintro.html

PSO

Particle Swarm Optimization, Kennedy & Eberhart (1995)

A quarter century of particle swarm optimization, Cheng et. Al (2018)

Lecture slides, Thomas Bäck

Particle swarm optimization, Poli et. al (2007)

DE

Lecture slides Thomas Bäck

Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, Storn & Price (1997)

Book Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series), Storn & Price (2006)

BFGS

Wikipedia page: https://en.wikipedia.org/wiki/Broyden%E2%80%93Fletcher%E2%80%93Goldfarb%E2%80%93Shanno_algorithm

A modified BFGS method and its global convergence in nonconvex minimization, Li & Fukushima (2001)

On the limited memory BFGS method for large scale optimization, Liu & Nocedal (1989)

The four papers from Broyden, Fletcher, Goldfarb and Shanno (all 1970) - they derived the approx. Hessian update independently from each other in the same year.

Sources / Literature

VNS
Variable Neighborhood Search, Mladenovic & Hansen (1997)
Variable Neighborhood search: Principles and applications, Hansen & Mladenovic (2001)
Variable Neighborhood search: Principles and applications, Hansen & Mladenovic (2010)

(H)MLSL
Stochastic Global Optimization Methods Part II: Multi Level Methods, Kan & Timmer (1987)
Benchmarking a Hybrid Multi Level Single Linkage Algorithm on the BBOB Noiseless Testbed, Pal (2013)

SANN
Lecture slides on simulated annealing
Optimization by Simulated Annealing, Kirkpatrick et. al (1983)
Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm, Cerny (1985)
Nice application on the travelling salesperson problem: https://www.heatonresearch.com/aifh/vol1/tsp_anneal.html

Nelder-Mead
Wikipedia article: https://en.wikipedia.org/wiki/Nelder%E2%80%93Mead_method

Tabu search
Tabu Search - Part I, Glover (1989)
Tabu Search - Part II, Glover (1990)
Tabu Search, Glover & Marti (2006)
Wikipedia article: https://en.wikipedia.org/wiki/Tabu_search