Mathematical Optimization with Stochastic Metaheuristics *in a Nutshell*

Black-Box / Combinatorial } Optimization

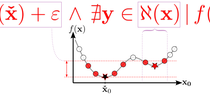
$\mathbf{x} \in \mathbb{R}^n$
$\mathbf{x} \in \mathbb{1}^n$
Parameters

$$\{\{\check{\mathbf{x}}_1 \dots \check{\mathbf{x}}_m\} \dots \{\check{\mathbf{x}}_1 \dots \check{\mathbf{x}}_m\}\} = \underset{\mathbf{x} \in X}{\arg\min}\ E_r(f(\mathbf{x})) + g(\mathbf{x})$$

Multi-objective optimization [LJT09]
Multi-modal optimization [Gri+21]
Noisy optimization / auto-ML [Bis+23]
Constrained optimization

f defines a well-ordered set
$f: f(\mathbf{x}) \leq f(\mathbf{y}) \wedge f(\mathbf{y}) \leq f(\mathbf{z})$
$\implies f(\mathbf{x}) \leq f(\mathbf{z}) \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X$
(Same for f+g)

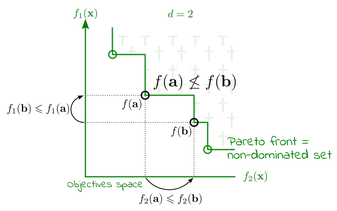
$$\{\mathbf{x} \mid f(\mathbf{x}) \leqslant f(\check{\mathbf{x}}) + \varepsilon \wedge \nexists \mathbf{y} \in \aleph(\mathbf{x}) \mid f(\mathbf{y}) < f(\mathbf{x})\}$$



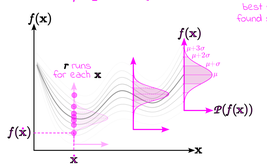
Example: $f: X \to \mathbb{R}^d$
Example: $g: X \to]\hat{\varphi}, +\infty]$
one can penalize with infinity
(this would be a vector in a multi-objective setting)
$\hat{\varphi} = \max_{\mathbf{x} \in X} f(\mathbf{x})$
Penalization

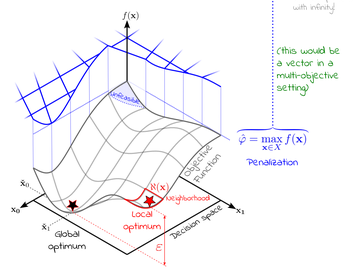
Reuse of notation for the sake of having a modular formulation

$f(\mathbf{x}) \leq f(\mathbf{y}) \iff f_d(\mathbf{x}) \leqslant f_d(\mathbf{y}) \quad \forall d \in \mathbb{N}^+$
Dominates



Example:
$$E_r(f(\mathbf{x})) := \mathop{\mathbb{E}}_{r}\left[\min\left\{0, e^{f(\mathbf{x})} - f(\dot{\mathbf{x}})\right\}\right]$$
best value found so far





Algorithms

Highly simplified, for the sake of understanding the core concepts.

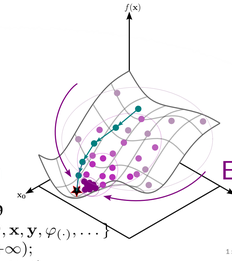
Metaheuristic [Dre+06]

Universe = whatever state you are interested in (e.g. time, iterations, target, parameter...)

Parameters: θ
State: $\mathbb{U} := \{\theta, \mathbf{x}, \mathbf{y}, \varphi_{(\cdot)}, \dots\}$

$(\mathbf{x}, \varphi_x) \leftarrow (\varnothing, +\infty)$;
$(\mathbf{y}, \varphi_y) \leftarrow (\mathcal{U}_X, f(\mathbf{y}))$; initialization
while continue($\mathbb{U}$) **do**
  $(\mathbf{x}, \varphi_x) \leftarrow \mathtt{select}_\theta((\mathbf{x}, \varphi_x), (\mathbf{y}, \varphi_y))$;
  $\theta \leftarrow \mathtt{update}_\theta(\mathbb{U})$;
  $\mathbf{y} \leftarrow \mathtt{variate}_\theta(\mathbf{x})$;
  $\varphi_y \leftarrow f(\mathbf{y})$;

Local Search



Parameters: θ
State: $\mathbb{U} := \{\theta, \mathbf{x}, \mathbf{y}, \varphi_{(\cdot)}, \dots\}$

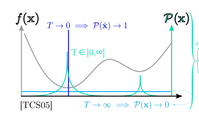
$(\mathbf{x}, \varphi_x) \leftarrow (\varnothing, +\infty)$;
$(\mathbf{y}, \varphi_y) \leftarrow (\mathcal{U}_X, f(\mathbf{y}))$;
while continue($\mathbb{U}$) **do**
  $(\mathbf{x}, \varphi_x) \leftarrow \mathtt{select}_\theta((\mathbf{x}, \varphi_x), (\mathbf{y}, \varphi_y))$;
  $\theta \leftarrow \mathtt{update}_\theta(\mathbb{U})$;
  $\mathbf{y} \sim \aleph(\mathbf{x})$;
  $\varphi_y \leftarrow f(\mathbf{x})$;

Evolutionary Algorithm

Parameters: θ
State: $\mathbb{U} := \{\theta, P, Q, \Phi_{(\cdot)}, \dots\}$

$(P, \Phi_P) \leftarrow (\varnothing, +\infty)$;
$(Q, \Phi_Q) \leftarrow (\mathcal{U}_X, f(Q))$;
while continue($\mathbb{U}$) **do**
  $(P, \Phi_P) \leftarrow \mathtt{select}_\theta((P, \Phi_P), (Q, \Phi_Q))$;
  $\theta \leftarrow \mathtt{update}_\theta(\mathbb{U})$;
  $Q \leftarrow \mathtt{variate}_\theta(P)$;
  $\Phi_Q \leftarrow f(P)$;

Simulated Annealing



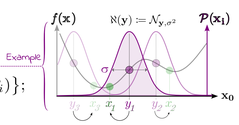
Function select($\mathbf{x}, \mathbf{y}$): Metropolis-Hastings
  State: T (temperature)
  if $f(\mathbf{x}) < f(\mathbf{y}) \vee \mathcal{U}_{0,1} \leqslant e^{\frac{f(\mathbf{y}) - f(\mathbf{x})}{-T}}$ **then**
    return $\mathbf{x}, f(\mathbf{x})$;
  else
    return $\mathbf{y}, f(\mathbf{y})$;

Function update(T):
  Parameter: $\beta \in]0, 1[$
  return $T \leftarrow \beta T$;

Evolution Strategies

Function variate(P):
  Parameter: $\aleph$
  return $\{\mathbf{x}_1 \dots \mathbf{x}_{|\mathbf{P}|} \mid \mathbf{x}_i \sim \aleph(P_i)\}$;

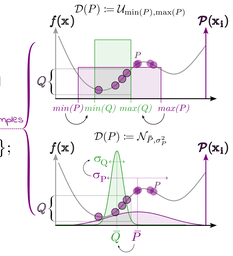
Example:
$\aleph(\mathbf{y}) := \mathcal{N}_{\mathbf{y}, \sigma^2}$



Estimation of Distribution

Function variate(P):
  Parameter: $\mathcal{D}$
  return $\{\mathbf{x}_1 \dots \mathbf{x}_{|\mathbf{P}|} \mid \mathbf{x}_i \sim \mathcal{D}(P)\}$;

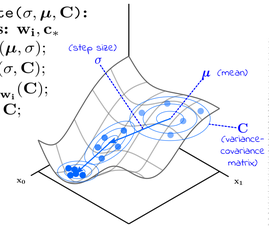
Examples:
$\mathcal{D}(P) := \mathcal{U}_{\min(P), \max(P)}$
$\mathcal{D}(P) := \mathcal{N}_{\hat{P}, \sigma_P^2}$

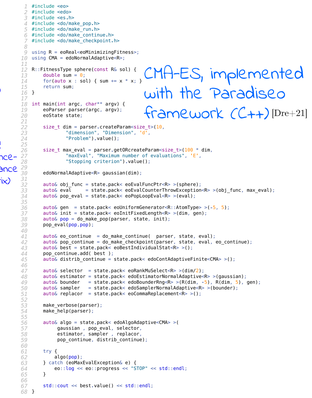


Covariance Matrix Adaptation ES

Function variate(P):
  State: $\sigma, \boldsymbol{\mu}, \mathbf{C}$
  return $\{\mathbf{x}_1 \dots \mathbf{x}_{|\mathbf{P}|} \mid \mathbf{x}_i \leftarrow \boldsymbol{\mu} + \sigma \mathbf{y}_i \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{C})\}$;

Function update($\sigma, \boldsymbol{\mu}, \mathbf{C}$):
  Parameters: $\mathbf{w}_i, \mathbf{c}_*$
  $\boldsymbol{\mu} \leftarrow \mathrm{upm}_{\mathbf{w}_i}(\boldsymbol{\mu}, \sigma)$;
  $\sigma \leftarrow \mathrm{ups}_{\mathbf{c}_\sigma}(\sigma, \mathbf{C})$;
  $\mathbf{C} \leftarrow \mathrm{upC}_{\mathbf{c}_+, \mathbf{w}_i}(\mathbf{C})$;
  return $\sigma, \boldsymbol{\mu}, \mathbf{C}$;



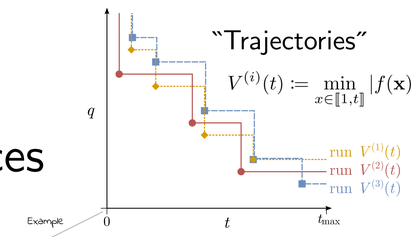


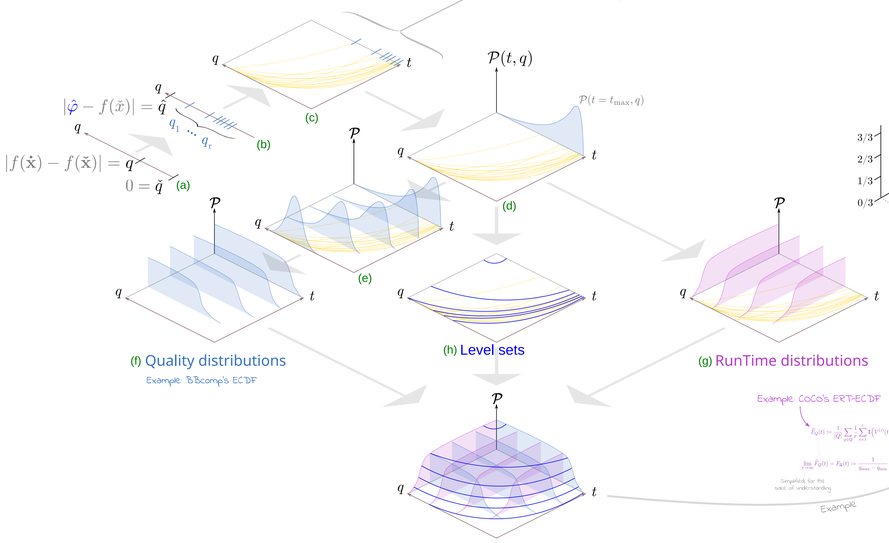
CMA-ES, implemented with the Paradiseo framework (C++) [Dre+21]

Performances

"Trajectories"

$$V^{(i)}(t) := \min_{x \in [\![1, t]\!]} |f(\mathbf{x}) - f(\check{\mathbf{x}})|$$





(f) Quality distributions
Example: BBcomp's ECDF

(h) Level sets

(g) RunTime distributions

Example: COCO's ERT-ECDF

$$\hat{F}_0(t) := \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \sum_{i=1}^{r} \mathbb{1}\left(V^{(i)}(t) < q\right)$$

$$\lim_{r \to \infty} \hat{F}_0(t) = F_0(t) := \frac{1}{q_{\max} - q_{\min}} \int_{q_{\min}}^{q_{\max}} \hat{a}(t, q)\, \mathrm{d}q$$

Simplified, for the sake of understanding

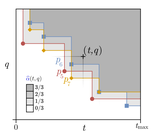
Example

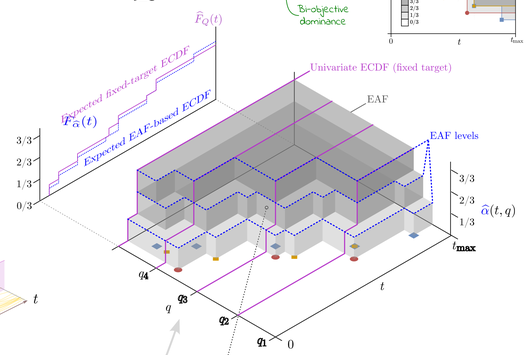
(i) Bivariate Quality/Time distribution = cumulative distribution of trajectories [Dre09]

Empirical Attainment Function [López+24]

$$\hat{a}(t, q) := \frac{1}{r} \sum_{i=1}^{r} \mathbb{1}\left(\exists \mathbf{p} \in V^{(i)} : \mathbf{p} \leq (t, q)\right)$$

Bi-objective dominance

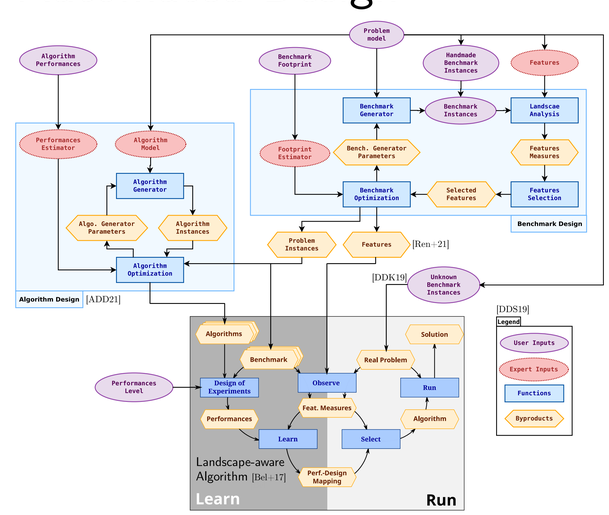




$$\mathrm{AUC}(\hat{a}(t, q)) = \int_{(1, -\infty)}^{(t_{\max}, t_{\min})} \hat{a}(t, q)\, \mathrm{d}t\, \mathrm{d}q$$

Also known as the "hypervolume" $= \frac{1}{r} \sum_{i=1}^{r} \sum_{t=1}^{t_{\max}-1} \max\left\{0, \hat{q} - V^{(i)}(t)\right\}$

Automated Design



References

[López+24] Manuel López-Ibáñez, Diederick Vermetten, Johann Dreo, and Carola Doerr. "Using the Empirical Attainment Function for Analyzing Single-objective Black-box Optimization Algorithms". In: *IEEE Transactions on Evolutionary Computation* (2024).

[Bis+23] Bernd Bischl, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, Difan Deng, and Marius Lindauer. "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges". In: *WIREs Data Mining and Knowledge Discovery* 13.2 (Jan. 2023). ISSN: 1942-4795.

[ADD21] Amine Aziz-Alaoui, Carola Doerr, and Johann Dreo. "Towards Large Scale Automated Algorithm Design by Integrating Modular Benchmarking Frameworks". In: *Proc. Companion of the Genetic and Evolutionary Computation Conference, Lille, France*. 2021.

[Dre+21] Johann Dreo, Arnaud Liefooghe, Sébastien Verel, Marc Schoenauer, Juan J. Merelo, Alexandre Quemy, Benjamin Bouvier, and Jan Gmys. "Paradiseo: from a modular framework for evolutionary computation to the automated design of metaheuristics: 22 years of Paradiseo". In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO '21. ACM, July 2021.

[Gri+21] Christian Grimme, Pascal Kerschke, Pelin Aspar, Heike Trautmann, Mike Preuss, André H. Deutz, Hao Wang, and Michael Emmerich. "Peeking beyond peaks: Challenges and research potentials of continuous multimodal multi-objective optimization". In: *Computers and Operations Research* 136 (Dec. 2021), p. 105489. ISSN: 0305-0548.

[Ren+21] Quentin Renau, Johann Dreo, Carola Doerr, and Benjamin Doerr. "Towards Explainable Exploratory Landscape Analysis: Extreme Feature Selection for Classifying BBOB Functions". In: *24th international conference on evolutionary computation (EvoStar)*. 2021.

[DDK19] Carola Doerr, Johann Dreo, and Pascal Kerschke. "Making a case for (Hyper-)parameter tuning as benchmark problems". In: *Proc. Companion of the Genetic and Evolutionary Computation Conference, GECCO, Prague, Czech Republic, July 13–17*. ACM, 2019, pp. 1755–1764.

[DDS19] Johann Dreo, Carola Doerr, and Yann Semet. "Coupling the design of benchmark with algorithm in landscape-aware solver design". In: *Proc. Companion of the Genetic and Evolutionary Computation Conference, Prague, Czech Republic, Jul. 13–17*. ACM, 2019, pp. 1419–1420.

[Bel+17] Naeiu Belkhir, Johann Dreo, Pierre Savéant, and Marc Schoenauer. "Per instance algorithm configuration of CMA-ES with limited budget". In: *Proc. of the Genetic and Evolutionary Comp. Conf., Berlin, Jul. 15–19*. ACM, 2017, pp. 681–688.

[Dre09] Johann Dreo. "Using performance fronts for parameter setting of stochastic metaheuristics". In: *Proc. Companion of the Genetic and Evolutionary Comp. Conf., Montreal, Québec, Canada, Jul. 8–12*. ACM, 2009, pp. 2197–2200.

[LJT09] Arnaud Liefooghe, Laetitia Jourdan, and El-Ghazali Talbi. "A unified model for evolutionary multi-objective optimization and its implementation in a general purpose software framework". In: *2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*. IEEE, Mar. 2009.

[Dre+06] Johann Dreo, Alain Pétrowski, Patrick Siarry, and Éric Taillard. "Metaheuristics for hard optimization: methods and case studies". Springer Science, 2006. ISBN: 978-3-642-06194-3.

[TCS05] E. Triki, Y. Collette, and P. Siarry. "A theoretical study on the behavior of simulated annealing leading to a new cooling schedule". In: *European Journal of Operational Research* 166.1 (Oct. 2005), pp. 77–92. ISSN: 0377-2217.

© 2024 Johann Dreo <johann.dreo@pasteur.fr>