Optimization for quantitative decisions: a versatile multi-tasker or an utopia?

Rodolphe Le Riche*

* CNRS at LIMOS (Mines Saint Etienne, UCA) France

25 July 2022
Cotonou, Benin, IA summer school

Acknowledgements : Vallet foundation, Bernard Guy

Abstract

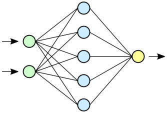
The study of optimization algorithms started at the end of World War II and has since then experienced a constantly growing interest, fueled by needs in engineering, computational simulation and Machine Learning. In this talk, we look into the history of the new scientific objects that are the optimizers. The point-of-view is historical and philosophical rather than mathematical. It is explained how the conditions for the emergence of the optimizers as new scientific objects correspond to a state of sufficient knowledge decomposition. The versatility of optimizers is illustrated through examples. Although they are just tools, optimizers are a source a fascination because they guide through a space of representations in a process that resembles learning. But rationality is bounded and, in that sense, optimization problems are utopias: they are an over-simplification of decisions which are actually rooted in human relationships; they often cannot be solved because the computational capacities are limited. The presentation finishes with some contemporary challenges in optimization.

The advent of optimization algorithms I

[Bach, 2020]

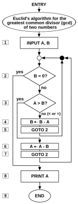
"Intelligence" $=$

models $+$

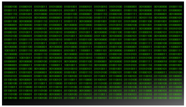


algorithms

$+$



data $+$



computing power

The advent of optimization algorithms II

There are numerous conferences* (and journals) where new optimization algorithms are presented, whether in

- applied mathematics. Ex: SIAM Conf. on Optimization, FGS (French-German-Swiss) workshops series, PGMO Days, ...
- computer science (including machine learning). Ex: LION (Learning and Intelligent OptimizatioN conf.), EURO (EURopean Operational research soc.) conf., Optimization days, NEURIPS, ...
- or in application fields. Ex: SIAM Conf. on Uncertainty Quantification (statistics), AIAA/MAO (Multidisciplinary Analysis and Optimization) or WCSMO (World Congress on Structural and Multidisciplinary Optimization) conferences (engineering), ECCOMAS (aeronautics), ICASP (civil engineering), ...

(* : arbitrarily listing some of the meetings I participated to)

The advent of optimization algorithms III

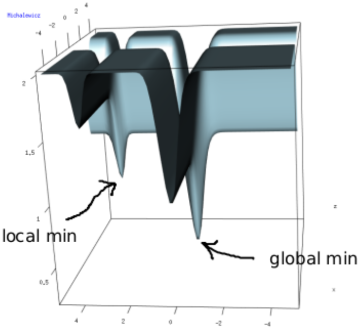
Some recent algorithms are highly cited, i.e., are seen as key technological components :

- NAG ([Nesterov, 1983], 1983, $>$ 5500 citations),
- EGO ([Jones et al., 1998], 1998, $>$ 7100 citations),
- CMA-ES ([Hansen and Ostermeier, 2001], 2001, $>$ 4100 citations),
- NSGA-II ([Deb et al., 2002], 2002, $>$ 41000 citations),
- ADAGRAD ([Duchi et al., 2011], 2011, $>$ 10400 citations),
- RMSprop ([Tieleman et al., 2012], 2012, $>$ 6000 citations),
- Adam ([Kingma and Ba, 2014], 2014, $>$ 113000 citations),
- ...

Optimization algorithms?

What are we talking about?

The fundamental optimization problem I



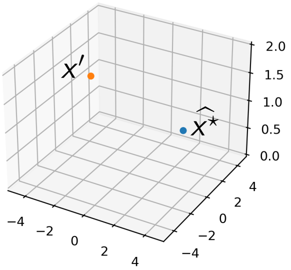
(2D continuous example)

Find the lowest point of a function,

$$\min_{x \in \mathcal{S}} f(x)$$

Looks easy on this drawing.
But the function is known pointwise, each evaluation costs computer time, and $\mathcal{S}$ may be complex (high-dimensional, non-continuous, constrained ...)

The fundamental optimization problem II



(what the random optimizer sees)

Random search

Require: $x^{\text{LB}},\ x^{\text{UB}}, t^{\max}$

$t \leftarrow 0,\ \widehat{f^\star} \leftarrow +\infty$

while $t < t^{\max}$ **do**

$\quad x' \leftarrow \mathcal{U}[x^{\text{LB}},\ x^{\text{UB}}]$ $\quad$ {uniform law}

$\quad$ calculate $f(x')$, $t \leftarrow t + 1$

$\quad$ **if** $f(x') < \widehat{f^\star}$ **then**

$\quad\quad \widehat{x^\star} \leftarrow x' \quad , \quad \widehat{f^\star} \leftarrow f(x')$

$\quad$ **end if**

end while

return $\widehat{x^\star}, \widehat{f^\star}$

Optimization = a quantitative formulation of decision I

Optimization is a way of mathematically modeling decision.

$$\min_{x \in \mathcal{S}} f(x)$$

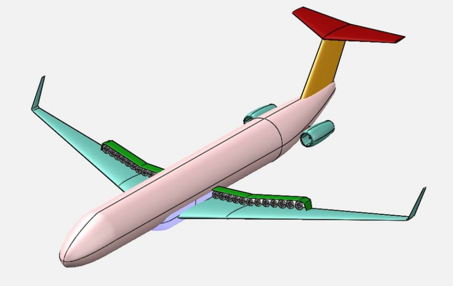
- x vector of decision parameters (variables) : dimensions, investment, tuning of a machine / program, ...
- $f(x)$: decision cost, minus $\times$ performance, ...
- $\mathcal{S}$: set of possible values for x, search space

Optimization = a quantitative formulation of decision II

A versatile approach to problem solving.

Examples :

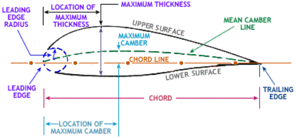
Optimization example: aircraft global pre-design

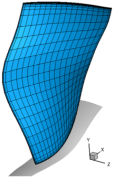
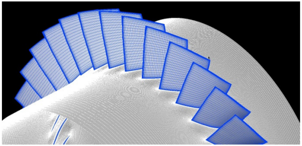


(from [Sgueglia et al., 2018])

x = aircraft parameters (here distributed electrical propulsion)
$f()$ = $-1\times$ performance metric (aggregation of $-1\times$ range, cost, take-off length, ...)
At the minimum, the design is "optimal".

Optimization example: blade detailed design



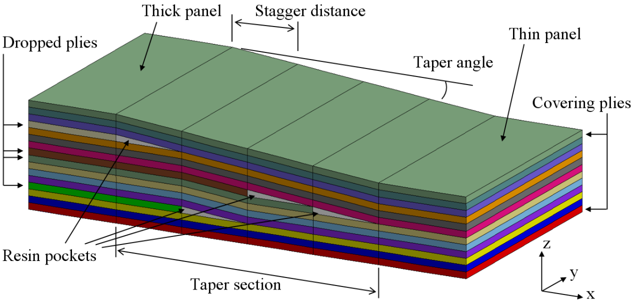



x: The blades are described by 4 cross-sections for a total of 20 design parameters.

$f()$: 5 constraints about the inlet and outlet relative flow angles, the flow speed reduction, excessive loading and the Mach number of the blade tips. The objective function is the polytropic (compressor) efficiency. From [Roustant et al., 2021].

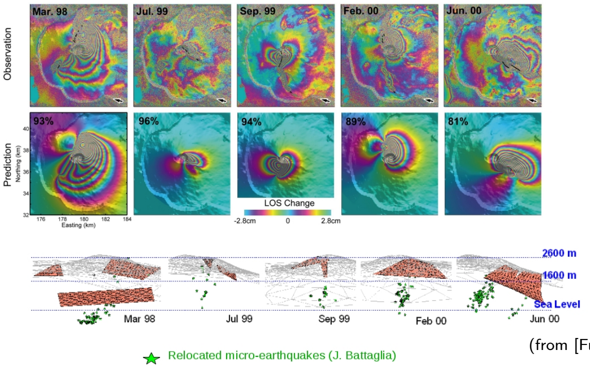
Favorable application: non-intuitive design and a small gain makes a large difference.

Optimization example: composite structure design



x is the orientation of the fibers within the plies of a composite laminate and the location where the plies are dropped.
$f()$ mechanical performance (stiffness, low mass, strength, stability, ...)
Many arrangements of the x's have almost equivalent performances, leading to local optima (from [Irisarri et al., 2014]).

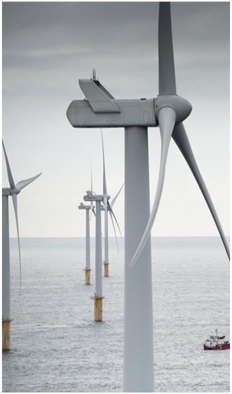
Optimization example: model identification

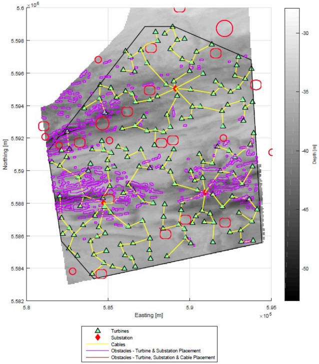


(from [Fukushima et al., 2010])

x = dike position, geometry, internal pressure
$f()$ = distance between measures (from RADARSAT-1 satellite) and model (boundary elements, non trivial computation)
At the minimum, the model best matches measurements and should correspond to the underground phenomenon.

Optimization example: wind farm layout



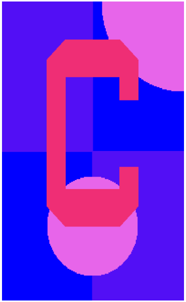


x = position and characteristics of the wind mill, electrical network.
$f()$ = cost aggregated with -1 $\times$ average power production.

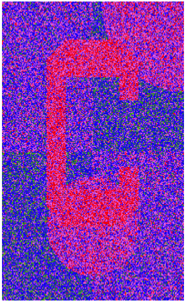
Optimization example: image denoising

$$\min_{x} f(x) \quad , \quad f(x) = \frac{1}{2} \sum_{i=1}^{N_{\text{pixels}}} (y_i - x_i)^2 + \lambda \sum_{i=1}^{N_{\text{pixels}}} \sum_{j \text{ near } i} |x_i - x_j|$$

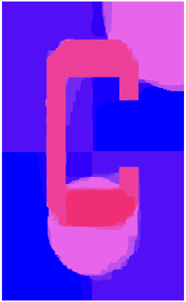
$\lambda \geq 0$ regularization constant



target image



noisy (observed) $= y_i$'s



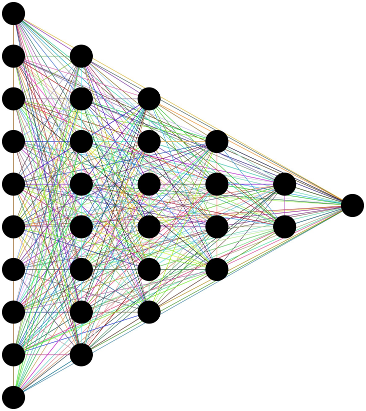
denoised (optimized) $= x^\star$

(from [Ravikumar and Singh, 2017])

Optimization example: neural net learning

x = neural network (NN) weights and biases

$f()$ = an error of the NN predictions w.r.t. data (a loss function)



from techxplore.com/news/2020-11-neural-network.html, CC0 public domain

An utopia ? Humans in the loop

Optimization as a mathematical model for decision

$$\min_{x \in \mathcal{S}} f(x)$$

Don't forget the human in the loop !

- In [Tsoukiàs, 2008], broader framework (model of the human rationality) $\Rightarrow$ decision aiding theory.
- Multi-Disciplinary Optimization [Brévault et al., 2020] as an attempt to model interacting disciplines.
- Human systems are complex, not easy to model.

Here, modest and practical goal : optimization as a (fascinating) tool.

Historical and epistemological hints

Emergence conditions

Optimization algorithms have become a scientific subject of its own when the problem to solve has been decomposed into

- a real object to optimize,
- a space of representations $\mathcal{S}$ where x is chosen,
- a model of the object valid for each representation, $m(x)$,
- a performance measure $f(m(x))$,
- a motion principle (the optimization algorithm).

Ex. in engineering: an object to design, x CAD parameters, $m(x)$ a finite elements model, $f()$ a compromise between -strength and cost.

Ex. in machine learning: a neural network to learn, x the weights, $m(x)$ the responses of the network over a certain data base, $f()$ the loss function.

The decomposition principle

Elements of scientific knowledge appear when a problem is sufficiently decomposed.

- Close to Descartes' 2nd Rule in Discourse on the Method : "Divide each difficulty into as many parts as is feasible and necessary to resolve it" (but applied to a scientific domain).

In the case of optimizers:

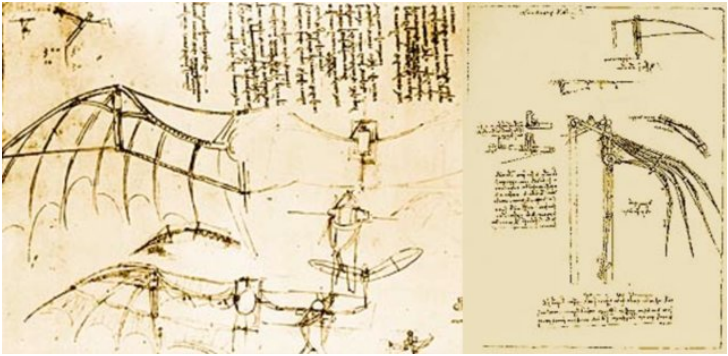
- A space-time generalized decomposition in the sense of Bernard Guy [Guy, 2011], where time is related to the motion of the optimizer.
- When applied to machine learning, optimizers are part of a constructivist system [Sarkar, 2016]: learnings occurs by the interaction between perception (ideas, = simulated actions) and action, new experience, in a trial and error process guided by the optimizer.

The emergence of optimization algorithms occured after World War II.
Optimization problems were solved before, but in a unified manner.

Examples:

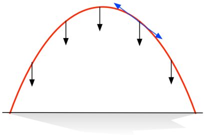
da Vinci's flying machine

Design for a flying machine, Leonardo da Vinci, 1488.



Hooke's arch

Optimization of an arch.
Hooke's theorem (1675) :
"As hangs the flexible chain, so but inverted will stand the rigid arch."

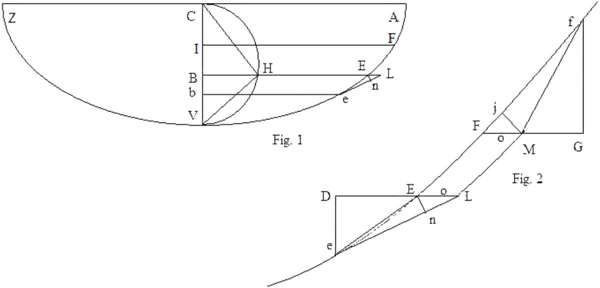


Robert Hooke holding a chain that forms a catenary curve. Image by Rita Greer. Licensed under Free Art License 1.3, via Wikimedia Commons.

The Brachistochrone curve

Minimize time to go from a point to a lower point with a frictionless ball under the action of gravity.

Posed and solved by Johann Bernoulli in 1696.



Millau's viaduct

Michel Virlogeux, main designer:
"The preliminary design of Millau's viaduct was first scribbled on a restaurant tablecloth and it was quite right. Computations then confirmed the ideas and fine tuned the details."

(talk at Maison de la Mécanique, Courbevoie, France, around 2005. Millau's viaduct is currently the tallest bridge in the world.)

The first optimization algorithms

They appeared after World War II, thanks to the first computers and a sufficient scientific maturity (cf. decomposition principle).

- The Steepest Descent Algorithm was formulated in its general form in 1944 by Haskell B. Curry [Curry, 1944] after initial ideas by Augustin Louis Cauchy in 1847 applied to astronomical equations solving [Lemaréchal, 2012].
- The Simplex Algorithm for solving linear problems was proposed around 1947 by George B. Dantzig [Dantzig et al., 1955].
- The Response Surface Method for approximating f by a quadratic function throughout $\mathcal{S}$ was proposed by George E. P. Box and Kenneth B. Wilson in 1951 [Box and Wilson, 1951].

Today's optimization algorithms I

(the discussion is voluntarily kept as short and non-technical as possible)

- Proven rapid (in polynomial time or super-linear) convergence for convex problems: cf. improved gradient descents such as BFGS (accounting for curvatures) and NAG [Nesterov, 1983].
- Algorithms adapted to stochastic functions (variants of stochastic gradients such as Adam [Kingma and Ba, 2014]).
- Algorithms using stochastic choices (e.g., CMA-ES [Hansen and Ostermeier, 2001]).

$\Rightarrow$ Recognition of the importance of uncertainties, for which randomness is both an ingredient and a cure.

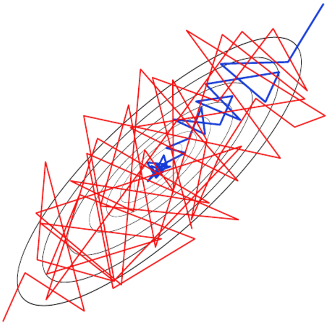
Today's optimization algorithms II

- Algorithms building models of the function (Bayesian optimization, [Mockus, 1975]) or of their validity domain (DFO, [Audet and Hare, 2017]).
- A lot of specialized versions: with/without constraints, for multiple objectives, with discrete and/or continuous variables, depending on the numerical cost of the function and the number of variables, . . .

An utopia ? Computational limitations

The fundamental problem, $\min_{x \in \mathcal{S}} f(x)$, is rarely solved in practice because the optimizer stops its iterations earlier (computational limitation).

In machine learning, when optimizing the loss function with a stochastic gradient, a situation in between the red and the blue convergences occurs. Yet, "it works".



An utopia ? Bounded rationality

(cf. Herbert A. Simon)

what's the point of accurately solving the problem since the formulation will always be imperfect?

- The model, $m()$ is inaccurate.
- Some variables are missing (some dimensions of $\mathcal{S}$).
- Other objectives (utilities, $f_1(), \ldots, f_N()$, N potentially infinite) exist as well.

Some current challenges

Optimization algorithms are a recent research domain. Still a lot to do!

- Properly testing and characterizing optimization algorithms is needed. The relationship between the best optimizer and the problem is complex. Initial work in [Hansen et al., 2021, Bosse and Griewank, 2012, Kerschke and Trautmann, 2019].
- Handling uncertainties in optimization problems is fundamental. An active sub-domain. Recent bibliography in [Pelamatti et al., 2022].
- The curse of dimensionality (with limited computations). Cf. e.g. [Binois and Wycoff, 2022].
- Less common types of variables : mixed (discrete-continuous, e.g.,[Cuesta Ramirez et al., 2022]), functional, or graph variables.
- Massively parallel, asynchronous, fault tolerant optimizers.

References I

Audet, C. and Hare, W. (2017).
Derivative-free and blackbox optimization, volume 2.
Springer.

Bach, F. (2020).
Optimization for large scale machine learning.
Hausdorff School, `https://www.di.ens.fr/~fbach/hausdorff2020_with_video.pdf`.

Binois, M. and Wycoff, N. (2022).
A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization.
working paper or preprint.

Bosse, T. and Griewank, A. (2012).
The relative cost of function and derivative evaluations in the CUTEr test set.
In *Recent Advances in Algorithmic Differentiation*, pages 233–240. Springer.

Box, G. E. and Wilson, K. B. (1951).
On the experimental attainment of optimum conditions.
Journal of the Royal Statistical Society, Series B, 13(1):1–45.

References II

Brévault, L., Balesdent, M., Morio, J., et al. (2020).
Aerospace System Analysis and Optimization in Uncertainty.
Springer.

Cuesta Ramirez, J., Le Riche, R., Roustant, O., Perrin, G., Durantin, C., and Gliere, A. (2022).
A comparison of mixed-variables bayesian optimization approaches.
Advanced Modeling and Simulation in Engineering Sciences, 9(1):1–29.

Curry, H. B. (1944).
The method of steepest descent for non-linear minimization problems.
Quarterly of Applied Mathematics, 2(3):258–261.

Dantzig, G. B., Orden, A., Wolfe, P., et al. (1955).
The generalized simplex method for minimizing a linear form under linear inequality restraints.
Pacific Journal of Mathematics, 5(2):183–195.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002).
A fast and elitist multiobjective genetic algorithm: Nsga-ii.
IEEE transactions on evolutionary computation, 6(2):182–197.

References III

Duchi, J., Hazan, E., and Singer, Y. (2011).
Adaptive subgradient methods for online learning and stochastic optimization.
Journal of machine learning research, 12(7).

Fukushima, Y., Cayol, V., Durand, P., and Massonnet, D. (2010).
Evolution of magma conduits during the 1998–2000 eruptions of piton de la fournaise volcano, réunion island.
Journal of Geophysical Research: Solid Earth, 115(B10).

Guy, B. (2011).
Penser le temps et l'espace.
Philosophia Santiae, 15(3):891–113.
doi=10.4000/philosophiascientiae.684.

Hansen, N., Auger, A., Ros, R., Mersmann, O., Tušar, T., and Brockhoff, D. (2021).
COCO: A platform for comparing continuous optimizers in a black-box setting.
Optimization Methods and Software, 36(1):114–144.

Hansen, N. and Ostermeier, A. (2001).
Completely derandomized self-adaptation in evolution strategies.
Evolutionary computation, 9(2):159–195.

References IV

Irisarri, F.-X., Lasseigne, A., Leroy, F.-H., and Le Riche, R. (2014).
Optimal design of laminated composite structures with ply drops using stacking sequence tables.
Composite Structures, 107:559–569.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global optimization, 13(4):455–492.

Kerschke, P. and Trautmann, H. (2019).
Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning.
Evolutionary computation, 27(1):99–127.

Kingma, D. P. and Ba, J. (2014).
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980.

Lemaréchal, C. (2012).
Cauchy and the gradient method.
Doc Math Extra, 251(254):10.

References V

Mockus, J. (1975).
On Bayesian methods for seeking the extremum.
In *Optimization Techniques IFIP Technical Conference*, pages 400–404. Springer.

Nesterov, Y. (1983).
A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$.
In *Doklady an USSR*, volume 269, pages 543–547.

Pelamatti, J., Le Riche, R., Helbert, C., and Blanchet-Scalliet, C. (2022).
Coupling and selecting constraints in bayesian optimization under uncertainties.
arXiv preprint arXiv:2204.00527.

Ravikumar, P. and Singh, A. (2017).
Convex optimization.
`http://www.cs.cmu.edu/~pradeepr/convexopt/`.

Roustant, O., Le Riche, R., Garnier, J., Ginsbourger, D., Deville, y., Helbert, C., Pronzato, L., Prieur, C., Gamboa, F., Bachoc, F., Rohmer, J., Perrin, G., Marrel, A., Damblin, G., Gliere, A., Sinoquet, D., Richet, y., da Veiga, S., and Huguet, F. (2021).
Chair in applied mathematics OQUAIDO Activity report.
Research report, OQUAIDO consortium.
HAL report `https://hal.archives-ouvertes.fr/hal-03217277`.

References VI

Sarkar, A. (2016).
Constructivist design for interactive machine learning.
In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pages 1467–1475.

Sgueglia, A., Schmollgruber, P., Bartoli, N., Atinault, O., Benard, E., and Morlier, J. (2018).
Exploration and sizing of a large passenger aircraft with distributed ducted electric fans.
In *2018 AIAA Aerospace Sciences Meeting*, page 1745.

Tieleman, T., Hinton, G., et al. (2012).
Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude.
COURSERA: Neural networks for machine learning, 4(2):26–31.

Tsoukiàs, A. (2008).
From decision theory to decision aiding methodology.
European journal of operational research, 187(1):138–161.