Blackbox optimization: Part 3/4: Applications

Sébastien Le Digabel

GERAD
GROUPE D’ÉTUDES ET DE RECHERCHE EN ANALYSE DES DÉCISIONS

POLYTECHNIQUE MONTRÉAL

ETICS 2023
Presentation outline

Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References
Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References
Aircraft takeoff trajectories

- [Torres et al., 2011]


- Biobjective optimization problem
Definition of the optimization problem

- Concept: Optimization of vertical flight path based on procedures designed to reduce noise emission at departure to protect airport vicinity.

- Minimization of environmental and economical impact: Noise and fuel consumption.

- Variables define the NADP (Noise Abatement Departure Procedure): During departure phase, the aircraft will target its climb configuration:
  - Increase the speed up to climb speed (acceleration phase)
  - Reduce the engine rate to climb thrust (reduction phase)
  - Gain altitude
Parametric Trajectory: 5 optimization variables (*)

Acceleration and thrust reduction can occur in any order.
The blackbox: Multi-Criteria Departure Procedure

DATA
Scenario configuration
(Aircraft, weather, airport, etc.)

NADP Trajectory

Multi-Criteria Departure Procedure

→ Noise
→ NOx Emissions
→ Consumption
→ Constraints

One evaluation \( \approx 2 \) seconds
Special features

- Must execute on different platforms including some old Solaris distributions

- The best trajectory parameters are returned to the pilot who enters them in the aircraft system manually → the less decimals the better

- Finite precision on optimization parameters: Discretization of optimization variables → granular variables [Audet et al., 2019]
Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References
Characterization of objects from radiographs - LANL

We want to identify an unknown object inside a box, using a x-ray source that gives an image on a detector.

In this work, the unknown object is supposed to be spherical.
A radiograph is the observed image on the detector. For example:
Description of the problem

► The problem consists to **identify the unknown object** with sufficient precision so that the object can be classified as dangerous or not.

► Must work **rapidly**

► Must work for radiographs **not created on a well-controlled experimental environment**

► Must **not crash** for unreasonable user inputs
Definition of the optimization problem

- **Variables:**
  - They represent a **spherical object**
  - **Meta variables:** Number of layers and type of material of each layer
  - **Continuous variables:** Radius of each layer
  - The **number of variables can change** depending on the number of layers

- **Objective function:**
  - A score associated to the difference between the observed image on the detector, and a simulated image obtained from the candidate object (**inverse problem**)
  - A numerical code – **the blackbox** – produces this simulated radiograph, using raytracing
Motivations for MADS and NOMAD

- A blackbox is involved
- Presence of meta variables
- Robustness of the code regarding the uncertainty and noise in the data
Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References
HPO with HyperNOMAD

- PhD project of Dounia Lakhmiri
- Published in TOMS [Lakhmiri et al., 2021]
- We focus on the HPO of deep neural networks

Our advantages:

- **Blackbox optimization problem:**
  
  *One blackbox call = Training + validation + test, for a fixed set of hyperparameters*

- Presence of categorical variables (ex.: number of layers)

- Existing methods are mostly heuristics
  
  *(grid search, random search, GAs, etc.)*

- Based on the NOMAD implementation of MADS
Principle

Blackbox

- Construct the network
- Network training, validation, testing

HyperNOMAD optimizer

- Hyperparameters
- Block structure
- Neighbors
- NOMAD

Initialization: dataset, starting point, evaluation budget

New point

Accuracy
## Hyperparameters for the architecture \((5n_1 + n_2 + 4)\)

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Type</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of convolutional layers ((n_1))</td>
<td>Meta</td>
<td>[0;20]</td>
</tr>
<tr>
<td>Number of output channels</td>
<td>Integer</td>
<td>[0;50]</td>
</tr>
<tr>
<td>Kernel size</td>
<td>Integer</td>
<td>[0;10]</td>
</tr>
<tr>
<td>Stride</td>
<td>Integer</td>
<td>[1;3]</td>
</tr>
<tr>
<td>Padding</td>
<td>Integer</td>
<td>[0;2]</td>
</tr>
<tr>
<td>Do a pooling</td>
<td>Boolean</td>
<td>0 or 1</td>
</tr>
<tr>
<td>Number of full layers ((n_2))</td>
<td>Meta</td>
<td>[0;30]</td>
</tr>
<tr>
<td>Size of the full layer</td>
<td>Integer</td>
<td>[0;500]</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td>Activation function</td>
<td>Categorical</td>
<td>ReLU, Sigmoid, Tanh</td>
</tr>
</tbody>
</table>
## Hyperparameters for the optimizer (5)

<table>
<thead>
<tr>
<th>Optimizer</th>
<th>Hyperparameter</th>
<th>Type</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic Gradient Descent (SGD)</td>
<td>Learning rate</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Momentum</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Dampening</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Weight decay</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td>Adam</td>
<td>Learning rate</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Weight decay</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td>Adagrad</td>
<td>Learning rate</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Learning rate decay</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Initial accumulator</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Weight decay</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td>RMSProp</td>
<td>Learning rate</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Momentum</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
<tr>
<td></td>
<td>Weight decay</td>
<td>Real</td>
<td>[0;1]</td>
</tr>
</tbody>
</table>
Results on CIFAR-10 (vs Hyperopt)

- Training with 40,000 images, validation/test on 10,000 images
- One evaluation (training+test) \( \approx 2 \) hours

\[(i7-6700@3.4 \text{ GHz, GeForce GTX 1070})\]

(a) Default starting point

(b) From a VGG architecture
Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References
CSP tower plant with molten salt thermal energy storage

- A large number of mirrors (heliostats) reflects solar radiation on a receiver at the top of a tower.
- The heat collected from the concentrated solar flux is removed from the receiver by a stream of molten salt.
- Hot molten salt is then used to feed thermal power to a conventional power block.
- The photo shows the Thémis CSP power plant, the first built with this design.

Source: https://commons.wikimedia.org/wiki/File:Themis_2.jpg
System dynamics

Heliostats

Receiver

Hot storage

Pump

Steam generator

Cold storage

Pump

Turbine

Condenser

Pump

Heat rejection

Electrical power

Molten salt cycle

Steam cycle

BBO: Applications

References

23/30
## Ten instances

<table>
<thead>
<tr>
<th>Instance</th>
<th># of variables</th>
<th># of obj.</th>
<th># of constraints</th>
<th># of stoch. outputs</th>
<th>Static surrogate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cont. discr. (cat.)</td>
<td>n</td>
<td>simu. a priori (lin.)</td>
<td>m</td>
<td>obj. or constr.</td>
</tr>
<tr>
<td>solar1</td>
<td>8 1 (0)</td>
<td>9</td>
<td>2 3 (2)</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>solar2</td>
<td>12 2 (0)</td>
<td>14</td>
<td>1 9 4 (2)</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>solar3</td>
<td>17 3 (1)</td>
<td>20</td>
<td>1 8 5 (3)</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>solar4</td>
<td>22 7 (1)</td>
<td>29</td>
<td>1 9 7 (5)</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>solar5</td>
<td>14 6 (1)</td>
<td>20</td>
<td>1 8 4 (3)</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>solar6</td>
<td>5 0 (0)</td>
<td>5</td>
<td>1 6 0 (0)</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>solar7</td>
<td>6 1 (0)</td>
<td>7</td>
<td>1 4 2 (1)</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>solar8</td>
<td>11 2 (0)</td>
<td>13</td>
<td>2 4 5 (3)</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>solar9</td>
<td>22 7 (1)</td>
<td>29</td>
<td>2 10 7 (5)</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>solar10</td>
<td>5 0 (0)</td>
<td>5</td>
<td>1 0 0 (0)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

^1 analytic objective  
^2 unconstrained
Features for BBO benchmarking

- Several numerical methods: real-world blackbox
- Reproducibility across all platforms
- Continuous and discrete variables
- Different types of constraints (quantifiable, relaxable, a priori, hidden)
- Stochastic and deterministic outputs
- Static surrogates with variable fidelity
- Number of replications is controllable
## Feasibility with sampling and NOMAD

<table>
<thead>
<tr>
<th>Instance</th>
<th>LH search (10k points)</th>
<th>NOMAD3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>satisf. ap constr. feas. pts</td>
<td>satisf. ap constr. feas. pts</td>
</tr>
<tr>
<td>solar1</td>
<td>30% 0.35%</td>
<td>96% 74%</td>
</tr>
<tr>
<td>solar2</td>
<td>0% 0%</td>
<td>97% 0%</td>
</tr>
<tr>
<td>solar3</td>
<td>0.49% 0%</td>
<td>99% 9%</td>
</tr>
<tr>
<td>solar4</td>
<td>0% 0%</td>
<td>83% 0%</td>
</tr>
<tr>
<td>solar5</td>
<td>0% 0%</td>
<td>83% 59%</td>
</tr>
<tr>
<td>solar6</td>
<td>90% 5%</td>
<td>99% 0%</td>
</tr>
<tr>
<td>solar7</td>
<td>2% 1%</td>
<td>74% 72%</td>
</tr>
<tr>
<td>solar8</td>
<td>1% 0.03%</td>
<td></td>
</tr>
<tr>
<td>solar9</td>
<td>1% 0%</td>
<td></td>
</tr>
</tbody>
</table>

There has been no violation of hidden constraints during the construction of this table.
Optimization on solar1
Biobjective optimization (by L. Salomon)

Pareto front approximations for solar8 (left) and solar9 (right) with different solvers with a budget of 5K evaluations. Taken from [Bigeon et al., 2022]
Example 1: Aircraft takeoff trajectories

Example 2: Characterization of objects from radiographs

Example 3: Hyperparameters Optimization

Example 4: Solar thermal power plant

References

