Multilevel Evolution Strategies for Multigrid Problems

Ofer M. Shir
School of Computer Science, Tel-Hai College, and
The Galilee Research Institute - Migal
Upper Galilee, Israel
ofersh@telhai.ac.il

ABSTRACT

We introduce a multilevel mechanism into Evolution Strategies (ESs) to address multigrid problems, which represent real-world applications of extremely high dimensions possessing a multiscale nature (i.e., low-resolution variants provide coarser approximations to the original problem). ESs may obtain fine solutions to the high-scale formulations only within an impractically large number of objective function calls, and we therefore devise a novel multilevel ES framework to efficiently treat such problems. We propose an automated leveling-up scheme to facilitate guided-search over increasingly finer levels of the optimization problem, which terminates after a solution to the ultimate high-scale problem is attained. We instantiate the proposed multilevel self-adaptive ES framework by two specific strategies: the elitist single-child (1+1)-ES and the non-elitist multi-child derandomized (μ_W, λ)-sep-CMA-ES. We show that the proposed approach is suited for targeting a global optimization problem which was heretofore viewed as too complex to address.

Keywords

Evolution strategies; multilevel global optimization; multigrid; scalability; self-adaptation; high-definition control

1. MULTILEVEL EVOLUTION STRATEGY

Multigrid methods have been adjusted to global optimization targets in order to devise multilevel (ML) solvers [2] (not to be confused with ML in the sense of decomposition). Relevant problems adhere to the following **assumptions**: (i) The decision variables are defined on a 1D grid, or otherwise may be arranged on such, (ii) The objective function is well-defined per each grid-scale, and (iii) The model is static in the sense that the objective function does not shift during the course of optimization per each grid-scale.

Let an optimization model $\mathcal{M}$ be formulated over various grid-scales (dimensions), $\{n_\ell\}$, by means of minimization problems $\{\mathcal{P}_\ell : \mathbb{R}^{n_\ell} \rightarrow \mathbb{R}\}$ that are all normalized with a global minimum that reads a zero objective function value. The n_ℓ-dimensional feasible domain is denoted as $\mathcal{X}_\ell \subseteq \mathbb{R}^{n_\ell}$. Consider a self-adaptive ES, minimizing $\mathcal{P}_\ell$ at dimension n_ℓ, employing a set of strategy parameters $\mathcal{S}_\ell$. Note that $\mathcal{S}_\ell$ may comprise in our consideration either a scalar for a zeroth-order ES (the global-step-size $\sigma_\ell \in \mathbb{R}$), or a vector in the case of a first-order ES (individual step-sizes $\vec{d}_\ell \in \mathbb{R}^{n_\ell}$). The ES is deployed on $\mathcal{P}_\ell$ with a seed point $\vec{x}_\ell^{(0)} \in \mathcal{X}_\ell$, aiming to obtain up to a threshold ϵ with regard to the global minimum ("ϵ-solving"). Consequently, this randomized heuristic search outputs a minimizer $\vec{x}_\ell^* \in \mathcal{X}_\ell$ and an adapted strategy $\mathcal{S}_\ell$. We denote such a self-adaptive procedure by `solveES`. The main idea of the proposed ML-ES is to iteratively increase n_ℓ upon ϵ-solving each problem instance $\mathcal{P}_\ell$. Each iteration's output, $\{\vec{x}_\ell^*, \mathcal{S}_\ell\}$, is then leveled-up to the next dimension $n_{\ell+1}$ by means of a dedicated *upscale operator*, except for the global step-size which is reduced by a factor of $\sqrt{n_{\ell+1}/n_\ell}$. The adapted output becomes the following iteration's input, $\left\{\vec{x}_{\ell+1}^{(0)}, \mathcal{S}_{\ell+1}\right\}$, when ϵ-solving $\mathcal{P}_{\ell+1}$.

The proposed approach is summarized as Algorithm 1, whose *leveling-up schedule* is **fixed**, for simplicity, and set to a factor of υ. A straightforward treatment to the upscaling of

input : problemModel $\mathcal{M}$, initialDim N_i, finalDim N_f
output : minimizer $\vec{x}^* \in \mathbb{R}^{N_f}$
$\ell \leftarrow 1$
$n_\ell \leftarrow N_i$
$\vec{x}_\ell^{(0)} \leftarrow$ `randomInit`$(\mathcal{M}, n_\ell)$
$\mathcal{S}_\ell \leftarrow$ `initStrategy`$(\mathcal{M}, n_\ell)$
while $n_\ell \leq N_f$ **do**
  $\mathcal{P}_\ell \longleftarrow$ `formProblem`$(\mathcal{M}, n_\ell)$
  if $\ell > 1$ **then**
    $\vec{x}_\ell^{(0)} \leftarrow$ `upscale`$\left(\vec{x}_{\ell-1}^*, n_\ell\right)$
    $\mathcal{S}_\ell \setminus \{\sigma_\ell\} \leftarrow$ `upscale`$(\mathcal{S}_{\ell-1} \setminus \{\sigma_{\ell-1}\}, n_\ell)$
    $\sigma_\ell \leftarrow \frac{\sigma_{\ell-1}}{\sqrt{n_\ell / n_{\ell-1}}}$
  end
  $\left\{\vec{x}_\ell^*, \mathcal{S}_\ell\right\} \longleftarrow$ `solveES`$\left(\mathcal{S}_\ell, \mathcal{P}_\ell, \vec{x}_\ell^{(0)}, \epsilon\right)$
  if $n_\ell == N_f$ **then return** $\vec{x}_\ell^*$
  else if $\upsilon \cdot n_\ell \leq N_f$ **then** $n_{\ell+1} \leftarrow \upsilon \cdot n_\ell$
  else $n_{\ell+1} \leftarrow N_f$
  $\ell \leftarrow \ell + 1$
end

Algorithm 1: Multilevel ES with a fixed schedule.

the decision variables' vectors is to conduct standard *interpolation*, while fixing the edges, as done in standard image scaling on a 1D grid. We consider the following variants: (U-1) Nearest neighbor: setting the value of the nearest

GECCO '16 July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07.
DOI: http://dx.doi.org/10.1145/2908961.2909013

sample grid point, (U-2) Linear: setting linear interpolants between each pair of grid points, and (U-3) Cubic: setting shape-preserving piecewise cubic interpolants based on the neighboring grid points.

Theoretical results devising the optimal step-size for the (1+1)-ES operating with the so-called *1/5th success-rule* on an n_ℓ-dimensional Sphere, are due to Rechenberg [1]:

$$\sigma^*_{\text{theory}}\left(\vec{x}_\ell\right) \approx 1.224\frac{R\left(\vec{x}_\ell\right)}{n_\ell}, \qquad (1)$$

where $R\left(\vec{x}_\ell\right) = \sqrt{f_{\text{Sphere}}\left(\vec{x}\right)}$. In the current ML perspective, assuming that the decision variables are simply duplicated per each leveling-up between n_ℓ to $n_{\ell+1}$, the quadratic modeling is subject to increasing the *objective function value* by a factor of $n_{\ell+1}/n_\ell$. Since the optimal step-size is proportional to $R\left(\vec{x}_\ell\right)/n_\ell$, the step $\sigma_{\ell+1}$ should be reduced by a factor of $\sqrt{n_{\ell+1}/n_\ell}$ in each leveling-up. Adhering to the broad validity of the 1/5th rule [1], this argumentation justifies our step-size update scheme.

We instantiate the proposed ML approach by two ESs:

(I) **ML-$(1+1)$-ES**: employing σ_ℓ with the 1/5th rule; $\mathcal{S}_\ell^{(1+1)\text{-ES}} = \{\sigma_\ell\}$.

(II) **ML-(μ_W, λ)-sep-CMA-ES**: employing σ_ℓ and $\vec{d}_\ell$; $\mathcal{S}_\ell^{(\mu_W,\lambda)\text{-sepC}} = \left\{\sigma_\ell,\ \vec{d}_\ell\right\}$ (thus resetting the evolution paths each leveling-up). The recommended population sizing is utilized: $\mu_\ell = \lfloor \lambda_\ell/2 \rfloor, \quad \lambda_\ell = 4 + \lfloor 3 \cdot \log(n_\ell) \rfloor$.

2. EXPERIMENTAL OBSERVATION

Arranging a high-dimensional Sphere function on a 1D grid serves as a proof-of-concept for examining the proposed instantiations with $N_i = 10 \rightsquigarrow N_f = 10^4$ (Figure 1). Without ML, the default variants required on average at least $4 \cdot 10^5$ function evaluations on $N_f = 10^4$. The ML variants required on average $3 \sim 4 \cdot 10^4$ evaluations across the different upscaling realizations – obtaining a speed-up by a factor of 10 in evaluations. Also, it is evident that the ML-$(1+1)$-ES' global step-size systematically follows the pattern of the optimal step-size while keeping a steady small gap; the observed gap medians are $3.5 \sim 6 \cdot 10^{-5}$.

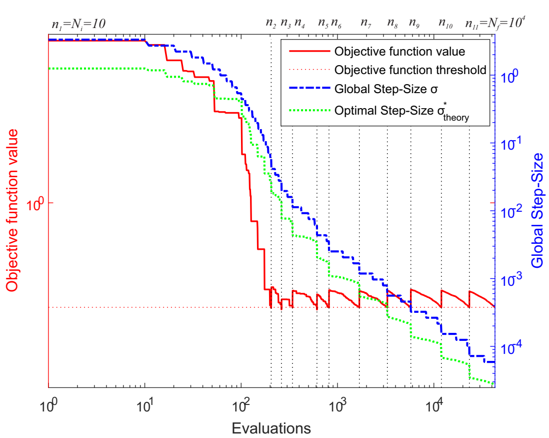

Figure 1: ML-$(1+1)$-ES applied to f_{Sphere} with $N_i = 10 \rightsquigarrow N_f = 10^4$, targeting a threshold of $\epsilon = 0.05$ in a log-log scale, employing (U-3), with vertical dashed lines that represent each leveling. The theoretically-optimal step-size σ^*_{theory} is calculated using Eq. (1).

Next, we targeted a real-world application from non-linear Optics, namely a simulation of two-photon absorption (TPA) processes [3], with 2^{14} variables (Figure 2). ML-$(1+1)$-ES and ML-(μ_W, λ)-sep-CMA-ES performed best on f_{TPA} when operating with (U-1) (on average $3 \sim 4 \cdot 10^3$ evaluations). The default ESs were not run on the high grid-scale due to the excessive computation time; when deployed on 2^{10} variables, they required at least $4 \cdot 10^4$ evaluations. The proposed ML approach successfully tackled grid-scales which have never been handled heretofore and achieved a speed-up by a factor of 10 with respect to the highest-scale treated.

3. REFERENCES

[1] H.-G. Beyer. *The Theory of Evolution Strategies.* Springer, Heidelberg, 2001.

[2] A. Brandt and D. Ron. Multigrid solvers and multilevel optimization strategies. In J. Cong and J. R. Shinnerl, editors, *Multilevel Optimization in VLSICAD*, volume 14 of *Combinatorial Optimization*, pages 1–69. Springer US, 2003.

[3] F. Laforge, J. Roslund, O. M. Shir, and H. Rabitz. Multiobjective adaptive feedback control of two-photon absorption coupled with propagation through a dispersive medium. *Phys. Rev. A*, 84:013401, Jul 2011.

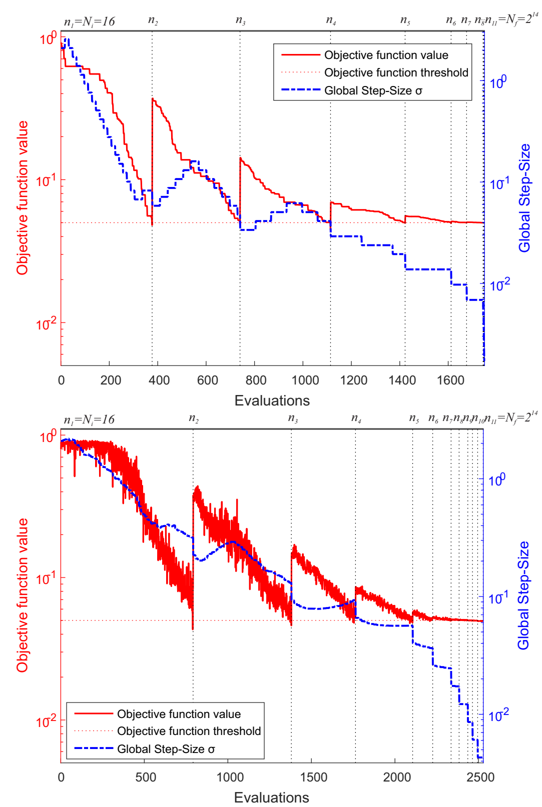

Figure 2: ML-$(1+1)$-ES [top] and ML-(μ_W, λ)-sep-CMA-ES [bottom] applied to f_{TPA} with $N_i = 2^4 \rightsquigarrow N_f = 2^{14}$, targeting a threshold of $\epsilon = 0.05$, both employing (U-1). Vertical dashed lines represent each leveling. The objective function and global step-size values are both log-scaled.