Computational Intelligence in The Natural Sciences: Machine Learning, Optimization, and Heuristic Search

Tel-Hai College

MIGAL Galilee Research Institute

Dr. Ofer M. Shir
School of Computer Science, Tel-Hai College
Migal - The Galilee Research Institute
Upper Galilee, ISRAEL

Background: Scientific Vitae

- 2016- : Senior Lecturer, School of Computer Science, Tel-Hai College
- 2013- : Principal Investigator, Migal - The Galilee Research Institute

- 2010-2013: Research Scientist at IBM-Research, Haifa Lab
  - 2nd Postdoctoral Term: Decision Analytics and Combinatorial Optimization

- 2008-2010: Postdoctoral Research Associate, Princeton University
  - Rabitz Group, Department of Chemistry
  - Main Project: Analytics in Experimental Chemistry

- 2004-2008: Doctoral Candidate, Leiden University
  - Natural Computing Group, Leiden Institute of Advanced CS
  - PhD Project: “An Evolutionary Approach to Many-Parameter Physics”
  - Promoters: Prof. Thomas Bäck and Prof. Marc Vrakking

- 2003-2004: MSc in Computer Science, Leiden University
  - Promoter: Prof. Thomas Bäck
  - Thesis Title: “Niching in Evolution Strategies”

- 2000-2003: BSc in Physics and Computer Science, HUJI

Spheres of Research: Learning/Optimization

Experimental Learning/Optimization in The Sciences
algorithmic design, efficiency,
robustness to uncertainty/noise

Simulation-Based Optimization
motivate refinements, mark limits,
scientific programming, design/architecture

Domain-Specific Mathematical/Statistical Analysis
back-to-basics in a new domain

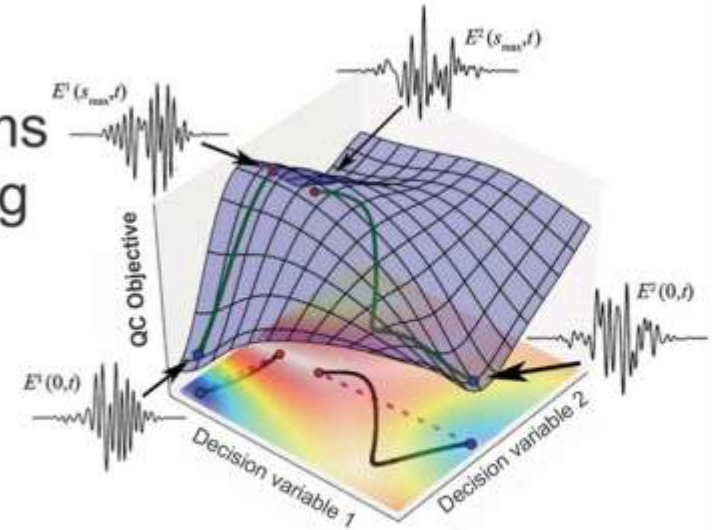
Learning/Optimization: Problem Formulation
motivate research

Theoretical Foundations to Heuristic Search
rigorous theoretical basis

Disprove presumptions

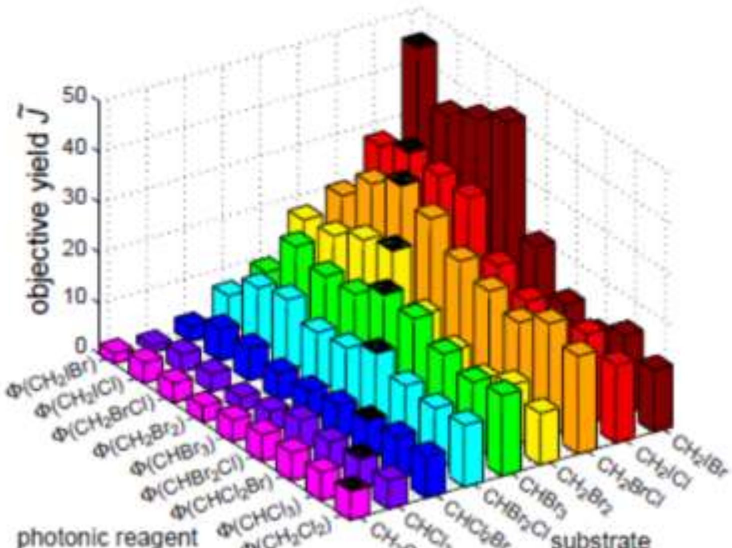
Motivation: Scientific Discovery as a Combinatorial Optimization Problem

- An underlying problem shared by scientists is to achieve optimal behavior of their systems and arrive at new discoveries while searching over a vast array of parameters (decision variables).



- This is commonly visualized in terms of a 'landscape', in which a candidate solution is mapped onto a 'position',its quality onto an 'altitude'.

- The task is translated into efficiently navigating within this search-space, which scales exponentially with the number of variables.



Kell, D.B., Scientific discovery as a combinatorial optimisation problem: How best to navigate the landscape of possible experiments? BioEssays, 2012. **34**(3): p. 236-244.

Effective Landscape Learning

Tel-Hai College

MIGAL Galilee Research Institute

UPMC SORBONNE UNIVERSITÉS

TECHNION Israel Institute of Technology

Goal: Efficient Hessian Learning

- Hessian determination about the optimum is very important:
  - Sensitivity analysis: assessing robustness of attained solutions
  - Reduced dimensional form for the optimal control basis
  - Mechanism investigation

- Is it possible to exploit derandomized search information for reduced-cost Hessian (no derivatives)?

$$\mathbf{H}(f(\vec{x})) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

FOCAL: Modus Operandi

- FOCAL: **F**orced **O**ptimal **C**ovariance **A**daptive **L**earning

- Idea: Employ CMA-ES, force exploration about the global maximum, and invert the attained covariance matrix
  Natural Computing = evolutionary pressure + statistical learning

- Shift the focus on canonical global optimization to landscape learning

- <u>Several modification to the default CMA-ES:</u>

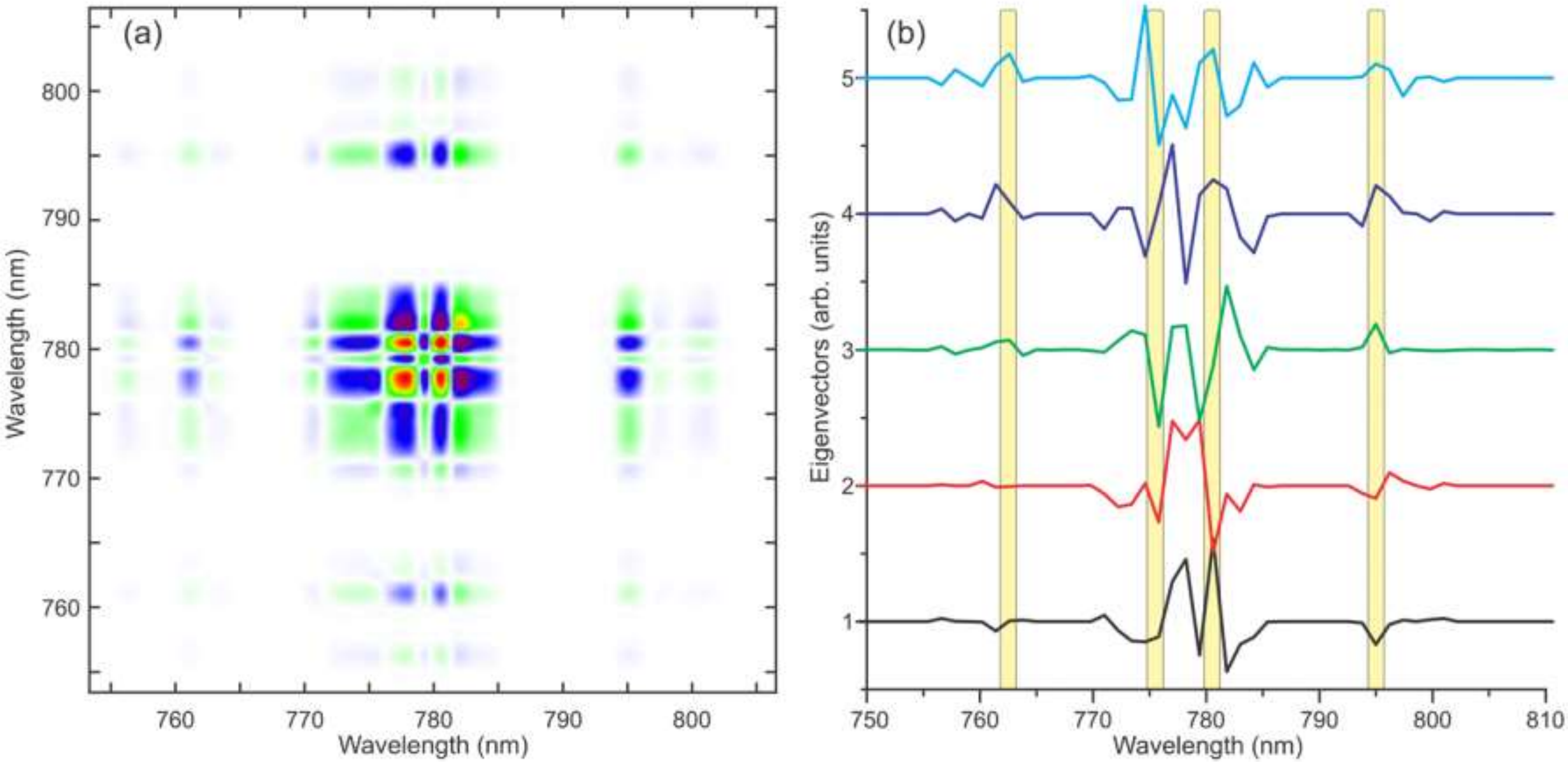
  - Forcing a finite-size step, especially upon convergence to optimum

  - Covariance learning rate is enhanced

  - Greedy selection pressure strives for the least yield declines, and pushes toward learning of the optimal mutation distribution

FOCAL: Experimental Results

(a) Retrieving the Hessian by FOCAL for rank-deficient atomic Rubidium
(b) 5 most important Hessian eigenvectors; Physical form corroborated



Shir, O.M., Roslund, J., Whitley, D., Rabitz, H.: Efficient retrieval of landscape Hessian: Forced optimal covariance adaptive learning. Physical Review E **89**(6) (2014) 063306.

Rigorous Foundations to Hessian Learning

Research question: *What is the relation between the statistically-learned covariance matrix and the landscape Hessian if a single ES winner is selected in each iteration assuming isotropic Gaussian samples?*

We distinguish between the optimization phase that aims to arrive at the optimum and is not discussed here, to the statistical learning of the basin – which lies in the focus of this study.

We assume a quadratic attraction basin.

By evaluating λ offspring in each iteration, the winner is recorded: $\vec{y} = \arg\min\{J(\vec{x})\}$.
ω denotes the objective function value:

$$\omega = J(\vec{y}) = \min\{J_1,\ J_2,\ \ldots,\ J_\lambda\}.$$

$t \leftarrow 0$
$\mathcal{S} \leftarrow \emptyset$
repeat
  for $k \leftarrow 1$ **to** λ **do**
    $\vec{x}_k^{(t+1)} \leftarrow \vec{x}^{\,*} + \vec{z}_k, \quad \vec{z}_k \sim \mathcal{N}\left(\vec{0}, \mathbf{I}\right)$
    $J_k^{(t+1)} \leftarrow \texttt{evaluate}\left(\vec{x}_k^{(t+1)}\right)$
  end
  $m_{t+1} \leftarrow \arg\min\left(\left\{J_i^{(t+1)}\right\}_{i=1}^{\lambda}\right)$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \left\{\vec{x}_{m_{t+1}}^{(t+1)}\right\}$
  $t \leftarrow t+1$
until $t \geq N_{iter}$
output: $\mathcal{C}^{\texttt{stat}} = \texttt{statCovariance}(\mathcal{S})$

Shir, O.M., Roslund J., Yehudayoff A., In: Proceedings of GECCO-2016, ACM Press (2016), 151-152.

Theoretical Results

Let $\mathcal{C}$ denote the covariance matrix of $\vec{y}$, and let $\mathcal{H}$ denote the Hessian about the optimum $\vec{x}^*$.

Theorem: *$\mathcal{C}$ and $\mathcal{H}$ are* ***commuting matrices*** *when the objective function follows the quadratic approximation*, that is, they are *simultaneously diagonalizable* and *share the same eigenvectors.*

Proof sketch:

1. The density of $\vec{y}$ reads: $\texttt{PDF}_{\vec{y}}\left(\vec{x}\right) = \texttt{PDF}_{\omega}\left(J(\vec{x})\right) \cdot \dfrac{\texttt{PDF}_{\vec{z}}\left(\vec{x}\right)}{\texttt{PDF}_{\psi}\left(J(\vec{x})\right)}$

2. Target $\mathcal{C}_{ij} = \int x_i x_j \texttt{PDF}_{\vec{y}}\left(\vec{x}\right) \mathrm{d}\vec{x}$, and apply a change of variables ($\mathcal{U}^{-1}\mathcal{H}\mathcal{U}=\mathcal{D}$):
$\vec{\vartheta}=\mathcal{U}^{-1}\vec{x},\ \mathrm{d}\vec{\vartheta}=\mathrm{d}\vec{x}.$

3. Consider $\mathcal{I}_{ij} = \left(\mathcal{U}^{-1}\mathcal{C}\mathcal{U}\right)_{ij}$ and show that it vanishes *for any i≠j due to symmetry considerations.*

4. Hence, $\mathcal{I}$ is the diagonalized form of $\mathcal{C}$, with $\mathcal{U}$ holding the eigenvectors.

Analytical Approximation

- We seek the density $\mathtt{PDF}_\omega$ to obtain the covariance form:

$$\mathtt{PDF}_\omega(\psi) = \lambda \cdot (1 - \mathtt{CDF}_\psi(\psi))^{\lambda-1} \cdot \mathtt{PDF}_\psi(\psi).$$

- Assuming $\lambda \to \infty$, we consider *minimal generalized extreme value distributions* $(\mathrm{GEVD}_{\mathrm{min}})$, to approximate the density:

$$\mathtt{PDF}_\omega^{\mathrm{GEVD}}\left(\tilde{\psi}\right) = \frac{n}{2}\tilde{\psi}^{\frac{n}{2}-1}\exp\left(-\tilde{\psi}^{\frac{n}{2}}\right)$$

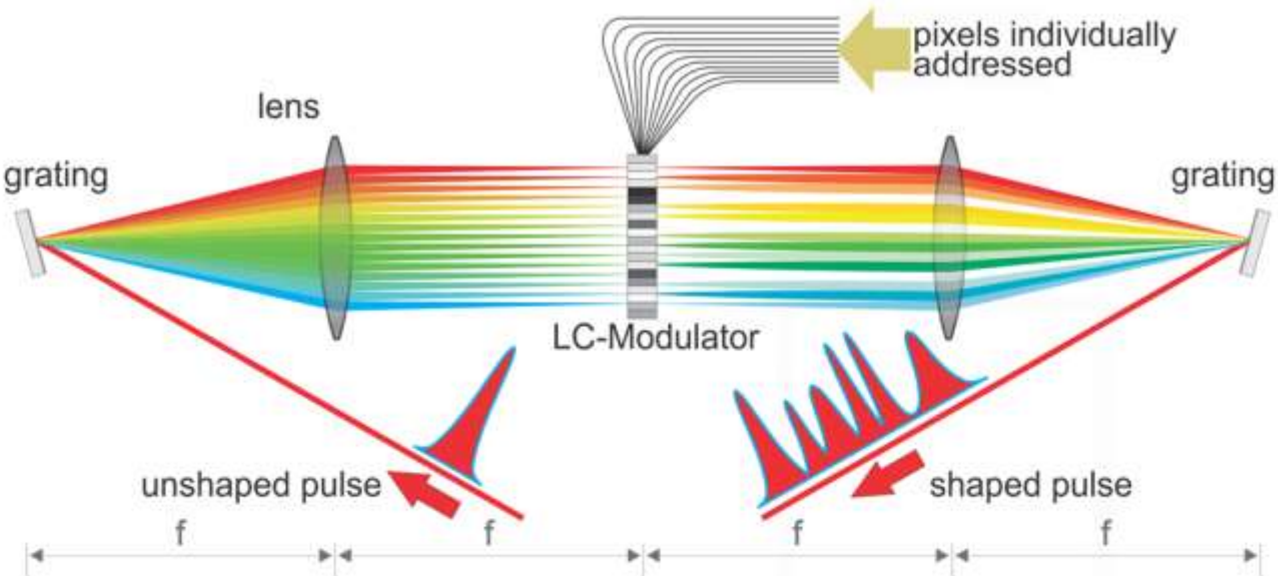
- Upon applying the necessary normalization, one obtains:

$$\mathcal{C}_{ij} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} N_{\mathcal{C}} x_i x_j \left(\vec{x}^T \mathcal{H} \vec{x}\right)^{\frac{n}{2}-\eta} \times$$
$$\times \exp\left[\Upsilon \vec{x}^T \mathcal{H} \vec{x} - \left(\frac{\vec{x}^T \mathcal{H} \vec{x}}{a_\lambda^*}\right)^{\frac{n}{2}} - \frac{1}{2}\vec{x}^T \vec{x}\right] \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n.$$

- For the isotropic case:

$$\mathcal{C}^{(\mathcal{H}=h_0\mathbf{I})} = \frac{\Gamma\left(\frac{n}{2}\right) \cdot \Gamma\left(1 + \frac{2}{n}\right) \cdot c(n) \cdot a_\lambda^*}{2\pi^{n/2}} \cdot \mathcal{H}^{-1}$$

Multilevel Optimization for High-Definition Control



Multiresolution and The Curse of Dimensionality

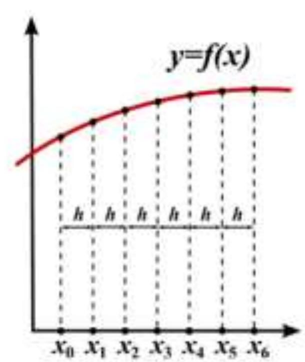
We consider a *practical* class of global optimization problems with thousands of variables whose optimizers may be meshed by a hierarchy of resolutions.

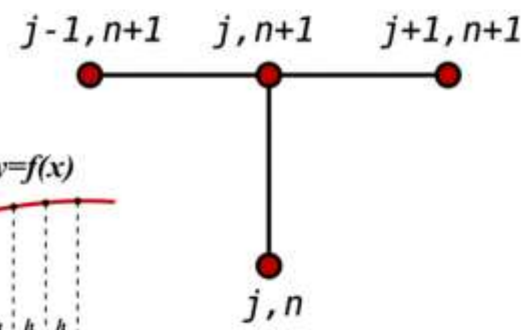
Such problems represent real-world applications of extremely high dimensions, often *discretizing* one or several functions, which possess a *multiscale nature*.

State-of-the-art Evolution Strategies (ESs) obtain fine solutions to the high-scale formulations only within an impractically large number of objective function calls.

We introduce a novel Multilevel ES (ML-ES) framework to efficiently treat such problems, adhering to the following assumptions:

- The decision variables are defined on a one-dimensional grid.
- The objective function is well-defined per each scale of the grid.
- The model is static - the objective function does not shift during optimization.





Shir, O.M., In: Proceedings of GECCO-2016, ACM Press (2016), 33-34.

Introducing: Multilevel Evolution Strategies

We introduce a heuristic to address *multigrid problems*, which represent high-dimensional optimization problems possessing a *multiscale nature:*

- An automated leveling-up scheme
- Search over increasingly finer levels
- Termination after a solution to the ultimate high-scale problem is attained

input : problemModel $\mathcal{M}$, initialDim N_i, finalDim N_f
output : minimizer $\vec{x}^* \in \mathbb{R}^{N_f}$
$\ell \leftarrow 1$
$n_\ell \leftarrow N_i$
$\vec{x}_\ell^{(0)} \leftarrow \texttt{randomInit}(\mathcal{M}, n_\ell)$
$\mathcal{S}_\ell \leftarrow \texttt{initStrategy}(\mathcal{M}, n_\ell)$
while $n_\ell \leq N_f$ **do**
  $\mathcal{P}_\ell \longleftarrow \texttt{formProblem}(\mathcal{M}, n_\ell)$
  if $\ell > 1$ **then**
    $\vec{x}_\ell^{(0)} \leftarrow \texttt{upscale}\left(\vec{x}_{\ell-1}^*, n_\ell\right)$
    $\mathcal{S}_\ell \setminus \{\sigma_\ell\} \leftarrow \texttt{upscale}(\mathcal{S}_{\ell-1} \setminus \{\sigma_{\ell-1}\}, n_\ell)$
    $\sigma_\ell \leftarrow \frac{\sigma_{\ell-1}}{\sqrt{n_\ell / n_{\ell-1}}}$
  end
  $\{\vec{x}_\ell^*, \mathcal{S}_\ell\} \longleftarrow \texttt{solveES}\left(\mathcal{S}_\ell, \mathcal{P}_\ell, \vec{x}_\ell^{(0)}, \epsilon\right)$
  if $n_\ell == N_f$ **then return** $\vec{x}_\ell^*$
  else if $2n_\ell \leq N_f$ **then** $n_{\ell+1} \leftarrow 2n_\ell$
  else $n_{\ell+1} \leftarrow N_f$
  $\ell \leftarrow \ell + 1$
end

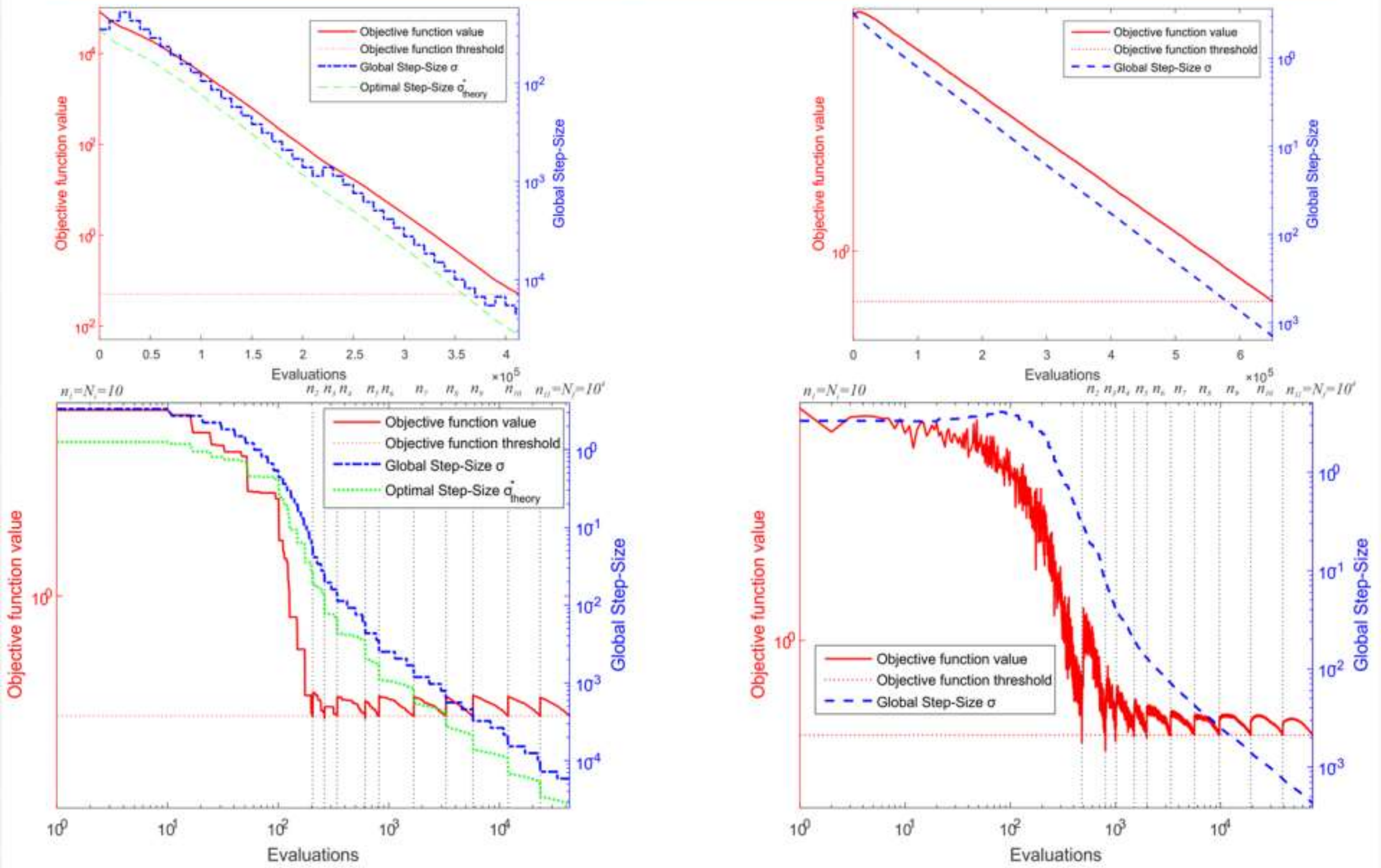
ML-ES *featuring* a fixed schedule with $n_{\ell+1}/n_\ell = 2$.

The ES is run on each problem-instance (level), solving it up to a threshold ϵ, where each level's output is upscaled to become the next level's input.

- The global step-size is reduced by a factor: $$\sigma_\ell \leftarrow \frac{\sigma_{\ell-1}}{\sqrt{n_\ell / n_{\ell-1}}}$$
- The decision vector and the strategy parameters are upscaled with an *Interpolating operator:*

$$\vec{x}_\ell^{(0)} \leftarrow \texttt{upscale}(\vec{x}_{\ell-1}^*, n_\ell)\ , \quad \mathcal{S}_\ell \leftarrow \texttt{upscale}(\mathcal{S}_{\ell-1}, n_\ell)$$

P.o.C.: High-Dimensional Quadratic Model Placed on a Grid



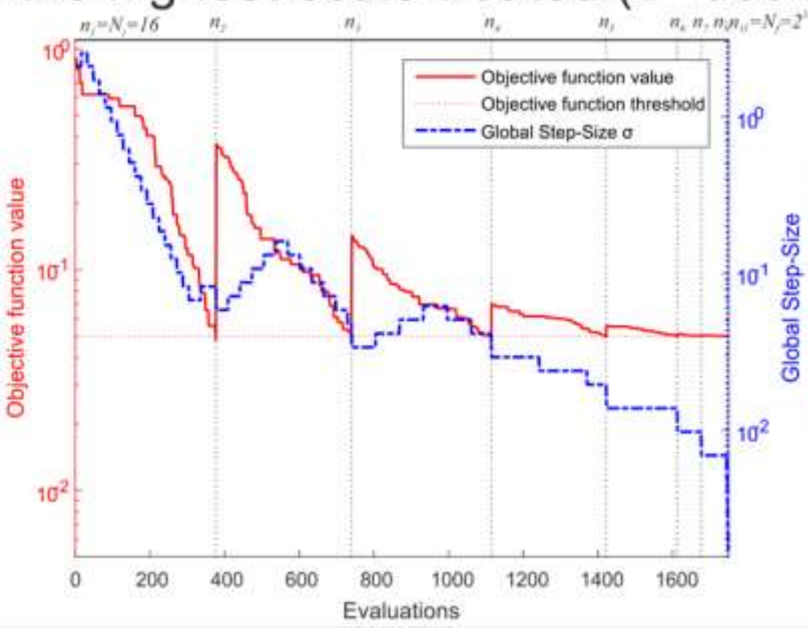
HD QC: Two-Photon-Absorption through Dispersive Media

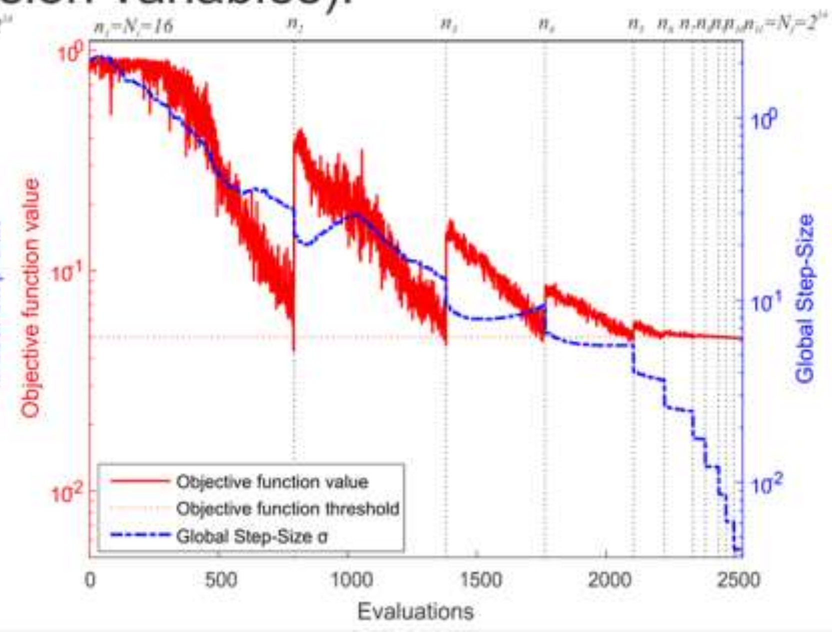
We targeted a simulated TPA system that accounts for effects of linear dispersion on the electric field, with 2^{14} variables.
Experimental optimization of such a system in *dispersive toluene* has been accomplished for a fixed setup ($n=128$) [LaForge et al., 2011].

ML-ES variants performed very well, utilizing $3\sim4\cdot10^3$ function calls.

The default variants were not run on the high-definition problem, due to the excessive computation time.

In sum, ML-ES 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 (2^{10} decision variables).





Collaborations and Acknowledgments

Prof. Amir **Yehudayoff**, **Technion-IIT**

Dr. Jonathan **Roslund**, **UPMC**

Prof. Herschel **Rabitz, Princeton University**
Dr. Xi **Xing, Princeton University**

Prof. Thomas **Bäck, Leiden University**
Dr. Michael **Emmerich**, **Leiden University**