A Theoretical Review of Evolutionary Strategies: Applications in Reinforcement Learning - Monday, 25.07.2022

Bachelor Thesis

Michael Van Huffel

Computational **S**cience and **E**ngineering **Lab**
ETH Zürich

With:

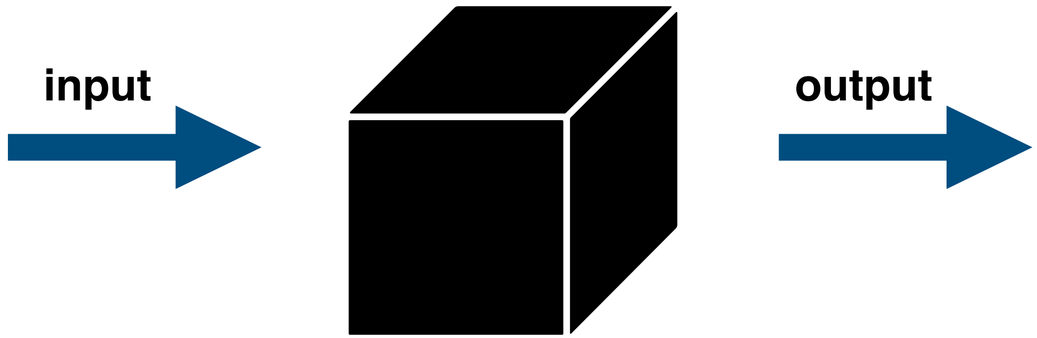
Supervisor: Prof. Dr. Petros Koumoutsakos
Dr. Georgios Arampatzis
Daniel Wälchli

Problem Statement

Given a black-box function $f : D \subset \mathbb{R}^{N_x} \rightarrow \mathbb{R}$ we are interested in the optimization problem

$$\mathbf{z}^\star := \arg\max_{\mathbf{z}\in\Omega} f(\mathbf{z})$$

with $\Omega \subset \mathbb{R}^{N_x}$.



Blackbox function

Instead of solving this objective, we can find the solution to the objective, [1]

$$\boldsymbol{\theta}^\star := \arg\max_{\boldsymbol{\theta}\in\Omega} \mathbb{E}_{p(\mathbf{z}|\boldsymbol{\theta})} f(\mathbf{z}),$$

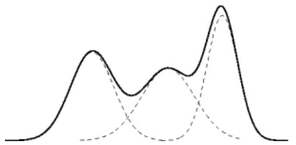
where $p(\mathbf{z} \mid \boldsymbol{\theta})$ is a distribution on D parameterized by the vector $\boldsymbol{\theta} \in \Omega$.

Our approach

$$\boldsymbol{\theta}^\star := \arg\max_{\boldsymbol{\theta}\in\Omega} \mathbb{E}_{p(\mathbf{z}|\boldsymbol{\theta})} f(\mathbf{z})$$

Iterative Maximization

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \log p(\mathbf{z}_i \mid \boldsymbol{\theta}), \quad \mathbf{z}_i \sim p(\mathbf{z} \mid \boldsymbol{\theta}^{(k)})$$



$$p(\mathbf{z} \mid \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{N} w_i \log \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\mu}^{(k)}, \Sigma^{(k)})$$

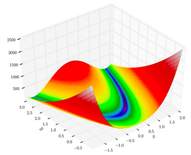
Maximum likelihood approach

- Maximum likelihood estimator
- Maximum a posteriori

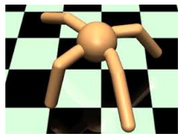
Gradient based approach

- First order method: SGA
- Second order methods: NGA, SNM

Benchmarks



MuJoCo tasks



Expectation maximization

The optimization problem can be rewritten as, [1]

$$\boldsymbol{\theta}^{\star} := \arg\max_{\boldsymbol{\theta}} \log \mathbb{E}_{p(\mathbf{z}|\boldsymbol{\theta})} f(\mathbf{z}) = \arg\max_{\boldsymbol{\theta}} \log \mathbb{E}_{q(\mathbf{z}|\boldsymbol{\theta}')} f(\mathbf{z}) \frac{p(\mathbf{z} \mid \boldsymbol{\theta})}{q(\mathbf{z} \mid \boldsymbol{\theta}')}, \quad f(\mathbf{z}) > 0 \quad \forall \mathbf{z} \sim p(\mathbf{z} \mid \boldsymbol{\theta})$$

where $q(\mathbf{z})$ is any probability density that depends on a fixed parameter $\boldsymbol{\theta}'$.

Using Jensen's inequality and introducing the KL divergence we have, [1]

$$\begin{aligned}
\log \mathbb{E}_{p(\mathbf{z}|\boldsymbol{\theta})} f(\mathbf{z}) &\geq \mathbb{E}_{q(\mathbf{z} \mid \boldsymbol{\theta}')}[\log f(\mathbf{z}) p(\mathbf{z} \mid \boldsymbol{\theta})] - \mathbb{E}_{q(\mathbf{z} \mid \boldsymbol{\theta}')} \left[\log q\left(\mathbf{z} \mid \boldsymbol{\theta}'\right)\right] \\
&= \mathbb{E}_{q(\mathbf{z} \mid \boldsymbol{\theta}')}[\log f(\mathbf{z}) p(\mathbf{z} \mid \boldsymbol{\theta})] + H\left[q\left(\mathbf{z} \mid \boldsymbol{\theta}'\right)\right] \\
&= F(q, \boldsymbol{\theta})
\end{aligned}$$

where $\tilde{p}(\mathbf{z} \mid \boldsymbol{\theta}) = \dfrac{p(\mathbf{z} \mid \boldsymbol{\theta}) f(\mathbf{z})}{\int_D p(\mathbf{z} \mid \boldsymbol{\theta}) f(\mathbf{z}) d\mathbf{z}}$ is the unknown exact posterior.

Expectation maximization

Expectation step:

$$q^{(k+1)}(\mathbf{z} \mid \boldsymbol{\theta}^{(k)}) = \arg\max_{q} \; F(q, \boldsymbol{\theta}^{(k)}) = \arg\min_{q} \; D_{KL}\left(q\left(\mathbf{z} \mid \boldsymbol{\theta}^{(k)}\right) \quad \tilde{p}(\mathbf{z} \mid \boldsymbol{\theta}^{(k)})\right)$$

Maximization step:

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} F(q^{(k+1)}, \boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{q^{(k+1)}(\mathbf{z}|\boldsymbol{\theta}^{(k)})}[\log(p(\mathbf{z} \mid \boldsymbol{\theta}) f(\mathbf{z}))]$$

If we chose $\boldsymbol{\theta}' = \boldsymbol{\theta}^{(k)}$ at each iteration and the approximate posterior $q^{(k+1)}(\mathbf{z} \mid \boldsymbol{\theta}^{(k)})$ to be equal to the mixture of N weighted particles, the maximization step is equal to

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \; \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \log p(\mathbf{z}_i \mid \boldsymbol{\theta}) \,, \quad \mathbf{z}_i \sim p(\mathbf{z}_i \mid \boldsymbol{\theta}^{(k)}) \,,$$

with $\hat{f}(\mathbf{z}_i) = f(\mathbf{z}_i) / \sum_{i=1}^{N} f(\mathbf{z}_i)$.

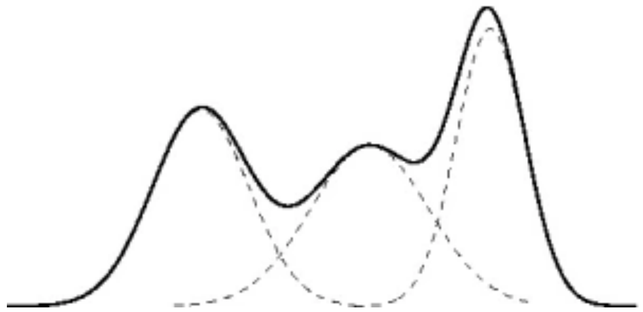
A Gaussian approach

We consider the special case where the search model $p(\mathbf{z} \mid \boldsymbol{\theta})$ is the weighted multivariate Gaussian distribution, hence

$$\mathscr{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} w_i \log \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\mu}, \Sigma), \quad \mathbf{z}_i \sim \mathcal{N}(\,\cdot \mid \boldsymbol{\mu}, \Sigma)\,,$$

where $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\mu} \\ \mathrm{vec}(\Sigma) \end{bmatrix}$, $w_i \geq 0$ and $\sum_{i=1}^{N} w_i = 1$.

$\mathrm{vec}(\,\cdot\,)$: vectorization operator



Exact Maximum Likelihood Solution

Maximum likelihood solution

Recall:

$$\mathscr{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} w_i \log \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\mu}, \Sigma)\,, \quad \mathbf{z}_i \sim \mathcal{N}(\,\cdot \mid \boldsymbol{\mu}, \Sigma)\,,$$

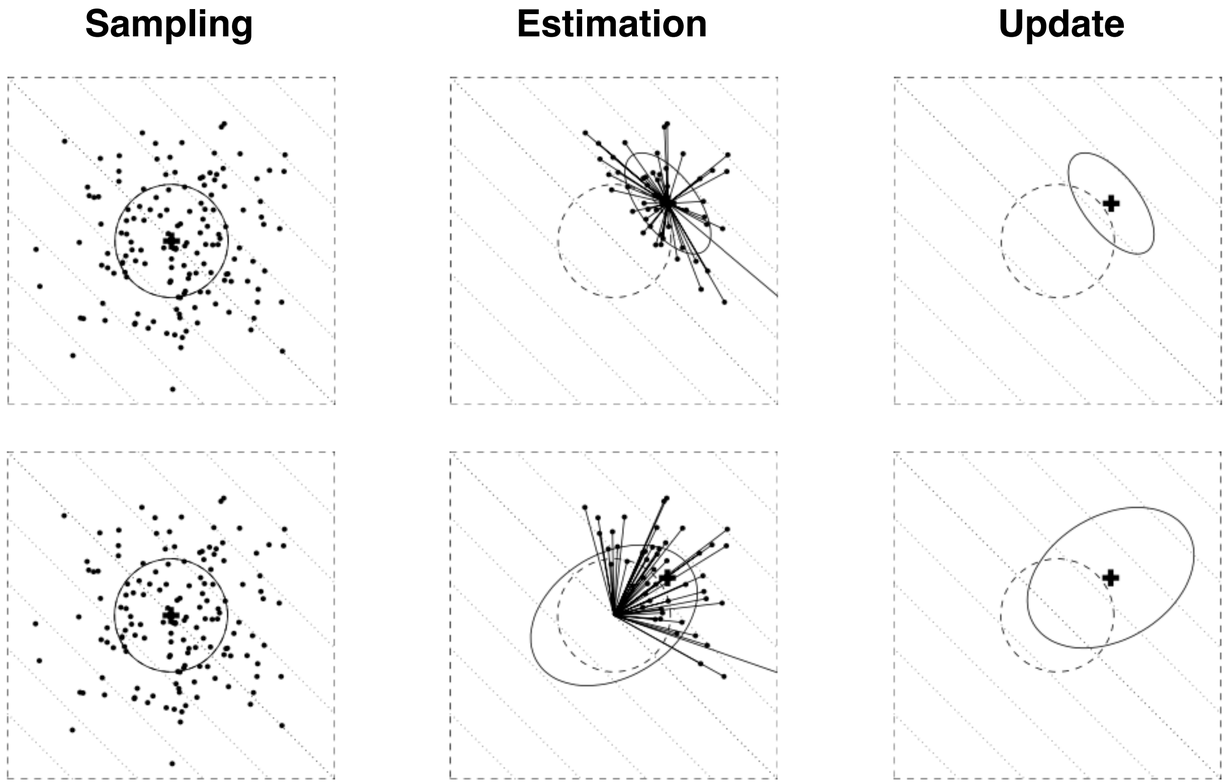
$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \log p(\mathbf{z}_i \mid \boldsymbol{\theta})\,, \quad \mathbf{z}_i \sim p(\mathbf{z}_i \mid \boldsymbol{\theta}^{(k)})\,.$$

When $p(\mathbf{z} \mid \boldsymbol{\theta})$ is chosen as $\mathscr{L}(\boldsymbol{\theta})$, the optimization problem can be solved analytically

$$\boldsymbol{\mu}^{(k+1)} \leftarrow \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\mathbf{z}_i\,,$$

$$\Sigma^{(k+1)} \leftarrow \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k+1)}\right)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k+1)}\right)^{\top}.$$

Premature shrinkage of the covariance matrix



$$\Sigma^{(k+1)} \leftarrow \sum_{i=1}^{N} \hat{f}\left(\mathbf{z}_i\right) \left(z_i - \boldsymbol{\mu}^{(k+1)}\right) \left(z_i - \boldsymbol{\mu}^{(k+1)}\right)^{\mathsf{T}}$$

$$\Sigma^{(k+1)} \leftarrow \sum_{i=1}^{N} \hat{f}\left(\mathbf{z}_i\right) \left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right) \left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)^{\mathsf{T}}$$

Note: the second update rule is similar to the CMA-ES of [2] without “rank-μ” and evolutionary path.

Figure adapted from [2].

Maximum a posteriori solution

The posterior distribution is defined as

$$p(\boldsymbol{\theta} \mid \mathbf{z}) \propto p(\mathbf{z} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{\lambda}), \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \lambda_2)$$

where $p(\boldsymbol{\theta} \mid \boldsymbol{\lambda})$ is the conjugate prior probability function.

The update scheme can be derived analytically and is given by

$$\boldsymbol{\mu}^{(k+1)} \leftarrow (1-\gamma)\boldsymbol{\mu}^{(k)} + \gamma \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\mathbf{z}_i \,,$$

$$\Sigma^{(k+1)} \leftarrow (1-\gamma)\Sigma^{(k)} + \gamma \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)^{\top} + (2\gamma-1)\boldsymbol{\mu}^{(k)}\boldsymbol{\mu}^{(k)^{\top}} + 2\boldsymbol{\mu}^{(k+1)}\boldsymbol{\mu}^{(k)^{\top}} - \boldsymbol{\mu}^{(k+1)}\boldsymbol{\mu}^{(k+1)^{\top}},$$

where $\gamma \in [0,1]$ is a smoothing parameter.

“CMA-ES-like”

$$\boldsymbol{\mu}^{(k+1)} \leftarrow (1-\gamma)\boldsymbol{\mu}^{(k)} + \gamma \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\mathbf{z}_i$$

$$\boldsymbol{\Sigma}^{(k+1)} \leftarrow (1-\gamma)\boldsymbol{\Sigma}^{(k)} + \gamma \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)\left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)^{\top}$$

Gradient Based Approach

Constrained optimization problem

Recall: we seek to solve the following constrained optimization problem

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \log \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\theta}), \quad \text{with} \quad \Sigma \succcurlyeq 0, \quad \mathbf{z}_i \sim \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\theta}^{(k)}).$$

To guarantee the positive semi-definite constrain of the covariance matrix we choose here the parameterization

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{s} \end{pmatrix},$$

where $\mathbf{s}$ is the half vectorized Cholesky decomposition of the covariance matrix $\Sigma = SS^T$.

Half vectorization:

$$S = \begin{bmatrix} a & & & \\ b & c & & \\ d & e & f & \\ g & h & i & j \end{bmatrix} \xrightarrow{\text{vech}(\cdot)} \begin{bmatrix} a \\ b \\ c \\ \cdot\cdot \\ j \end{bmatrix}$$

Euclidean gradient approach

The general update of the euclidean gradient is given by

$$\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} + \alpha^{(k)} \nabla_{\boldsymbol{\theta}} \mathscr{L}\left(\boldsymbol{\theta}^{(k)}\right).$$

Substituting the derivatives of the weighted Gaussian distribution $\mathscr{L}(\boldsymbol{\theta})$ w.r.t the parameter $\boldsymbol{\theta}$ we obtain,

$$\boldsymbol{\mu}^{(k+1)} \leftarrow \boldsymbol{\mu}^{(k)} + \alpha^{(k)} \sum_{i=1}^{N} \Sigma^{-1} \left(\hat{f}(\mathbf{z}_i)\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right),$$

$$\boldsymbol{s}^{(k+1)} \leftarrow \boldsymbol{s}^{(k)} + \frac{\alpha^{(k)}}{2} \operatorname{vech}\left(S^{-\top}\left(\sum_{i=1}^{N} \hat{f}(\mathbf{z}_i)\boldsymbol{u}_i\boldsymbol{u}_i^{\top} - \mathbb{I}_n\right)\right),$$

where $\boldsymbol{u}_i = S^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})$.

Natural gradient approach

The general update of the natural gradient approach is given by

$$\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} + \alpha^{(k)} \mathscr{I}_{\boldsymbol{\theta}}^{-1} \nabla_{\boldsymbol{\theta}} \mathscr{L}\left(\boldsymbol{\theta}^{(k)}\right).$$

Substituting the derivatives of the weighted Gaussian distribution $\mathscr{L}(\boldsymbol{\theta})$ w.r.t the parameter $\boldsymbol{\theta}$ and the inverse of the Fisher matrix $\mathscr{I}_{\boldsymbol{\theta}}$ we obtain

$$\boldsymbol{\mu}^{(k+1)} \leftarrow \boldsymbol{\mu}^{(k)} + \alpha^{(k)} \left(\sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \mathbf{z}_i - \boldsymbol{\mu}^{(k)} \right),$$

$$\boldsymbol{s}^{(k+1)} \leftarrow \boldsymbol{s}^{(k)} + \frac{\alpha^{(k)}}{2} L(\mathbb{I}_n \otimes S^{(k)}) L^{\top} (LNL^{\top})^{-1} L(\mathbb{I}_n \otimes S^{(k)^{\top}}) L^{\top} \cdot \operatorname{vech}\left(S^{-\top} \left(\sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \boldsymbol{u}_i \boldsymbol{u}_i^{\top} - \mathbb{I}_n \right) \right).$$

L: elimination matrix
$N = \left(K + \mathbb{I}_{n^2}\right)/2, K$: duplication matrix

Modified Newton-Raphson method update

The general update scheme of the Newton-Raphson method has the form

$$\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} - \mathscr{H}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}^{(k)}\right) \nabla_{\boldsymbol{\theta}} \mathscr{J}\left(\boldsymbol{\theta}^{(k)}\right) ,$$

where $\mathscr{J}(\boldsymbol{\theta})$ is the negative weighted Gaussian distribution.

$$\boldsymbol{\mu}^{(k+1)} \leftarrow \sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \mathbf{z}_i ,$$

$$\boldsymbol{s}^{(k+1)} \leftarrow \boldsymbol{s}^{(k)} + \frac{1}{2} \underbrace{\left(R^{(k)} + Q^{(k)} \frac{\overbrace{\partial^2 \mathscr{J}}^{\rightarrow \mathscr{J}_{\Sigma}}}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}^T} Q^{(k)\top} \right)^{-1}}_{= \tilde{\mathscr{H}}_{\mathbf{s}}^{(k)^{-1}}} Q^{(k)} \cdot \operatorname{vec}\left(\Sigma^{(k)^{-1}} \left(\sum_{i=1}^{N} \hat{f}(\mathbf{z}_i) \left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right) \left(\mathbf{z}_i - \boldsymbol{\mu}^{(k)}\right)^{\top} - \Sigma^{(k)} \right) \Sigma^{(k)^{-1}} \right) D .$$

D: duplication matrix, $\boldsymbol{\sigma} = \operatorname{vech}(\Sigma)$

Note: Given the positive semi-definiteness of the inverse of the modified Hessian $\tilde{\mathscr{H}}_{\mathbf{s}}^{(k)^{-1}}$ of the Gaussian distribution, this update ensures that $\Sigma^{(k)}$ remains positive semi-definite.

Benchmark results

Benchmark results

- **Benchmark functions:** Hypersphere, Ackley, Rosenbrock, Griewank, and Rastrigin
- **Dimensions n:** 2, 8, 32, 128, 512

	$2D$	$8D$	$32D$	$128D$	$512D$
Full covariance	7	58	730	-	-
Diagonal covariance	5	21	83	340	1350

- **Population:** $pop \approx 1.3 \cdot pop_{min}$ for more stable convergence

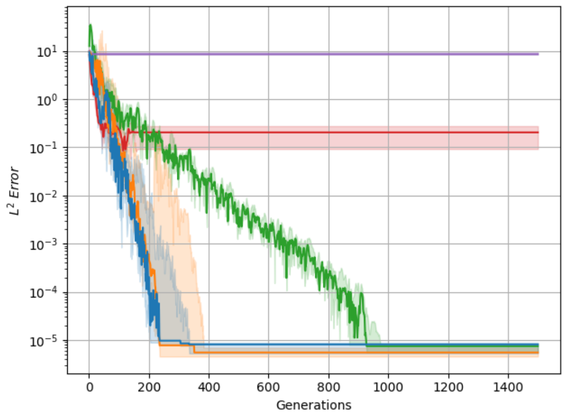
- **Number of runs:** 5
- **Maximum number of generations:** 1500
- **Fixed initial L^2 error:** 10
- **Fixed initial covariance:** $\Sigma^{(0)} = \mathbb{I}_n$
- **Hessian correction method for SNM:** EVdelta

Hypersphere function

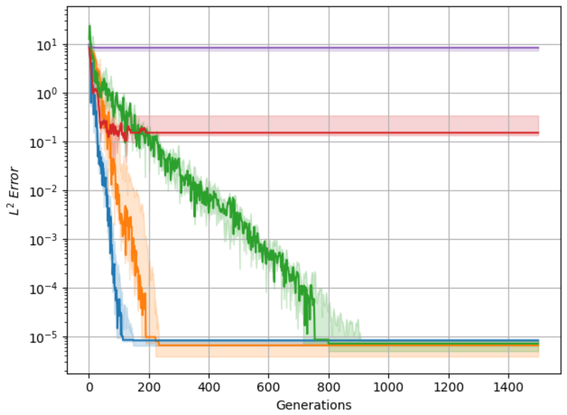
$$f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$$

2D

Diagonal covariance

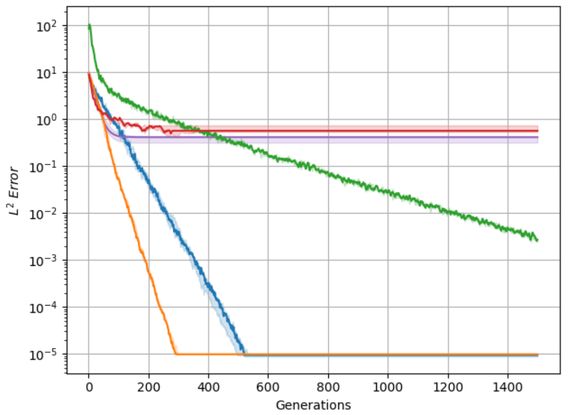


Full covariance

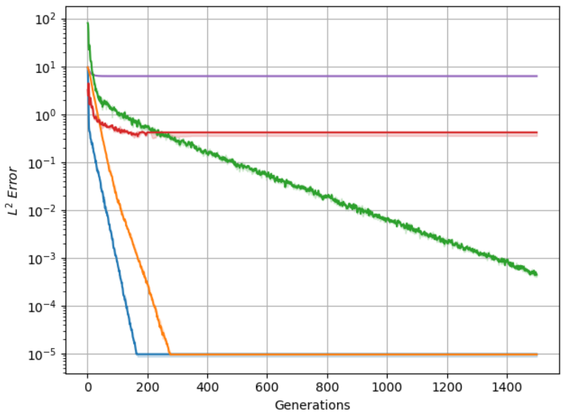


32D

Diagonal covariance

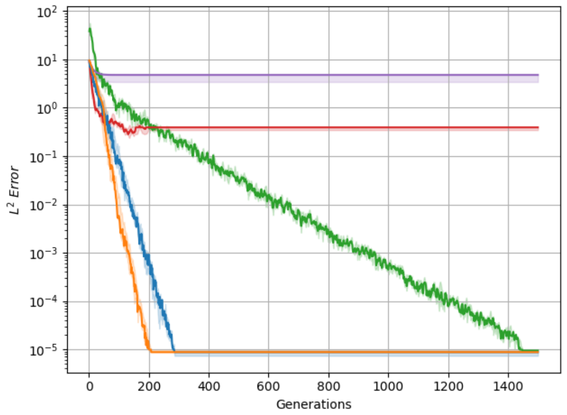


Full covariance

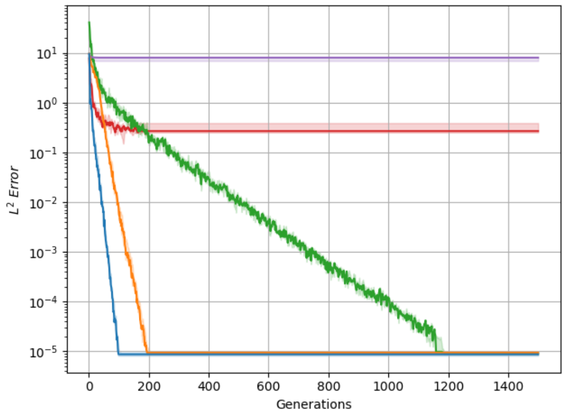


8D

Diagonal covariance

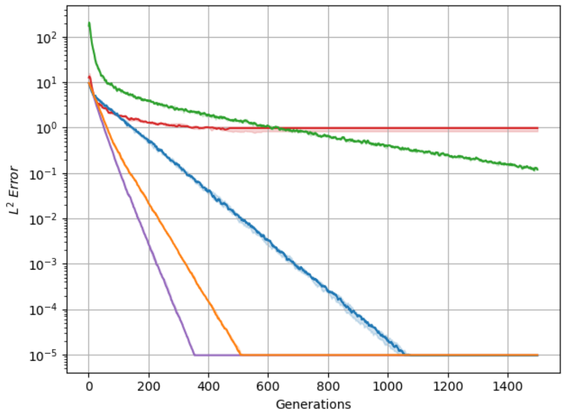


Full covariance



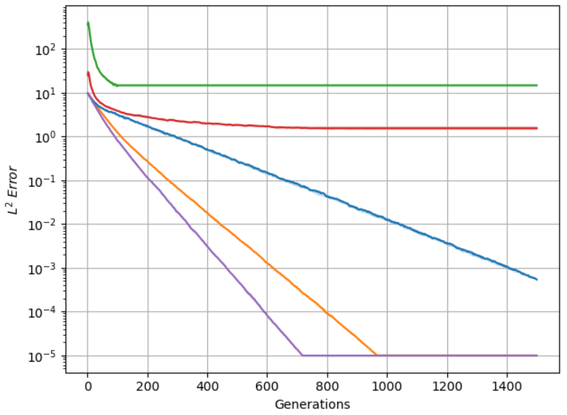
128D

Diagonal covariance



512D

Diagonal covariance



The lines correspond to MAP (—), SNM (—), NGA (—), SGA(—), and EM (—)

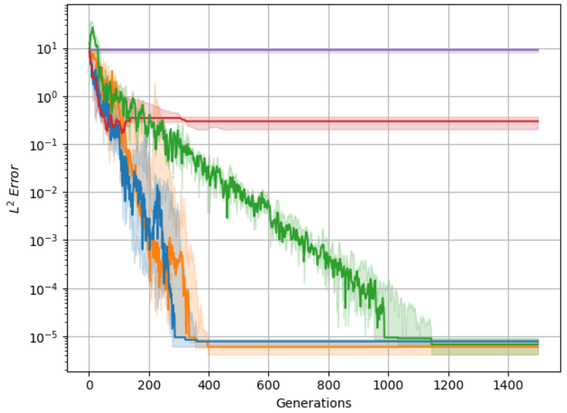
Ackley function

$$f(\boldsymbol{x}) = -a\exp\left(-b\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(x_i c)\right) + a + e$$

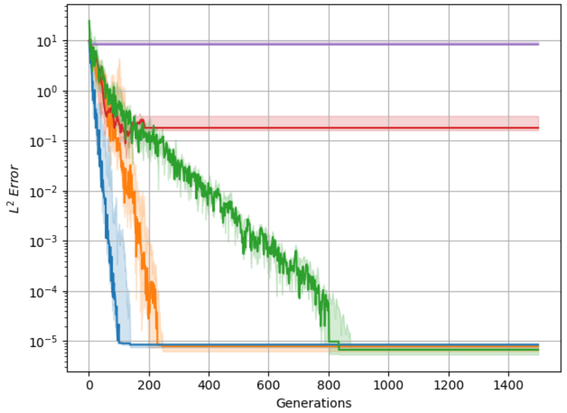
a, b, e constant

2D

Diagonal covariance

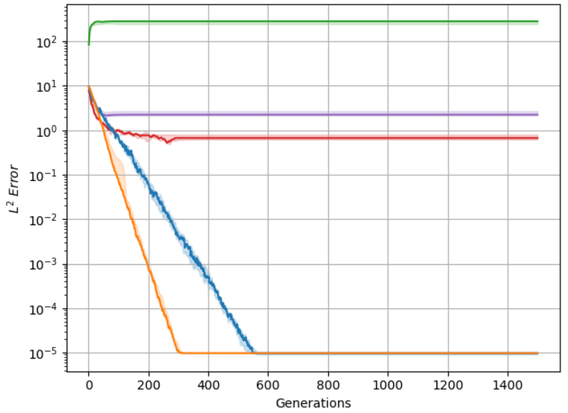


Full covariance

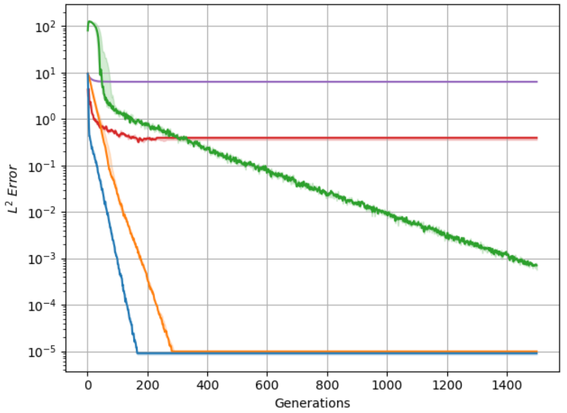


32D

Diagonal covariance

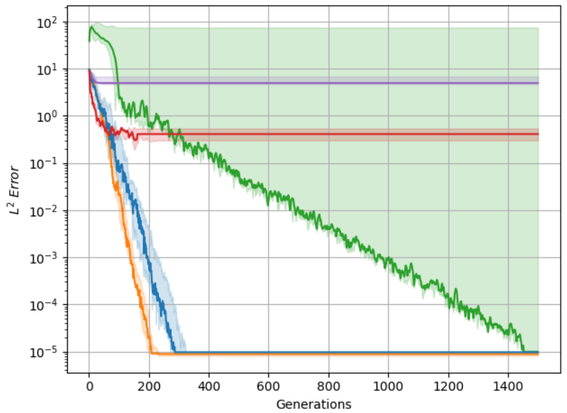


Full covariance

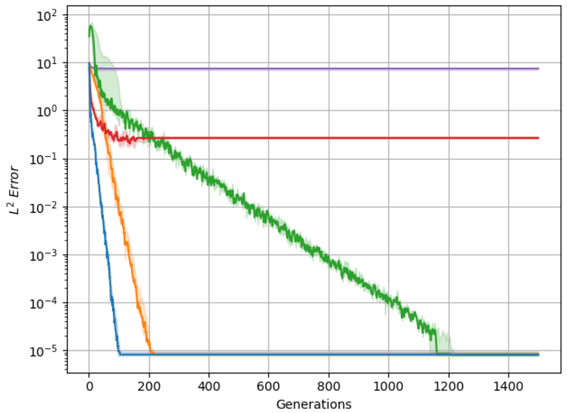


8D

Diagonal covariance

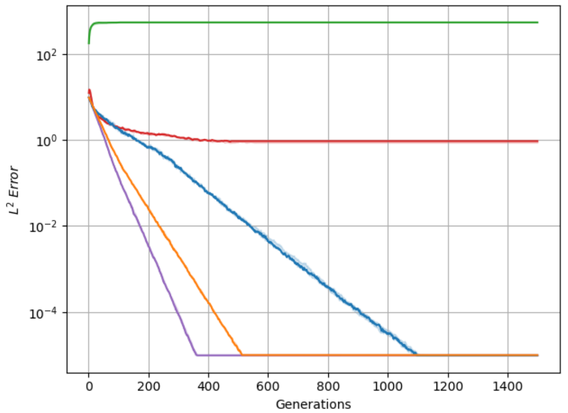


Full covariance



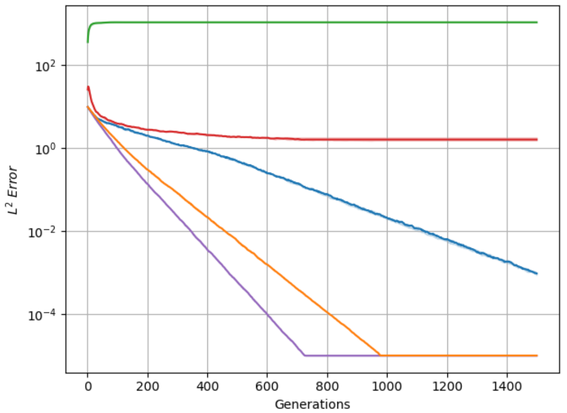
128D

Diagonal covariance



512D

Diagonal covariance



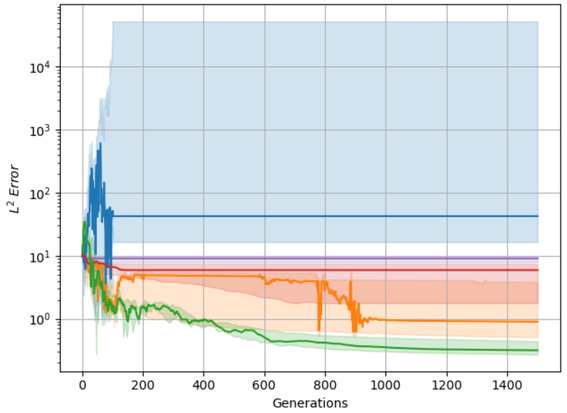
The lines correspond to MAP (—), SNM (—), NGA (—), SGA(—), and EM (—)

Rosenbrock function

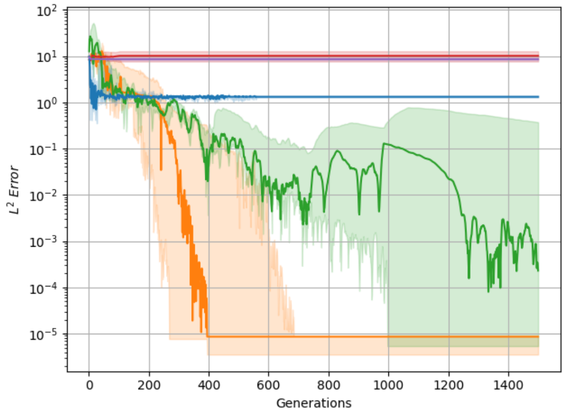
$$f(\boldsymbol{x}) = \sum_{i=1}^{n-1} \left[100 \left(x_{i+1} - x_i^2 \right)^2 + \left(1 - x_i \right)^2 \right]$$

2D

Diagonal covariance

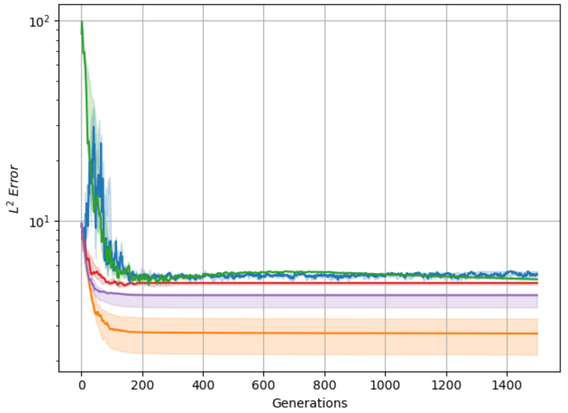


Full covariance

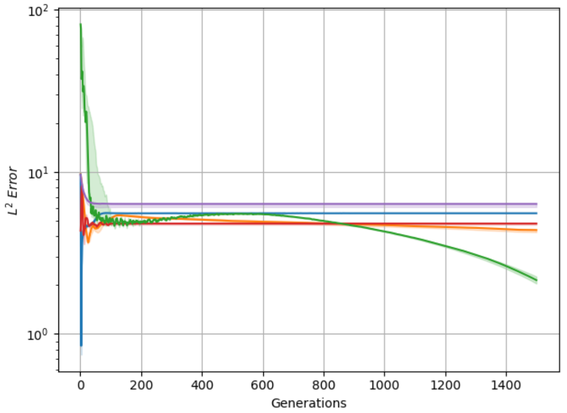


32D

Diagonal covariance

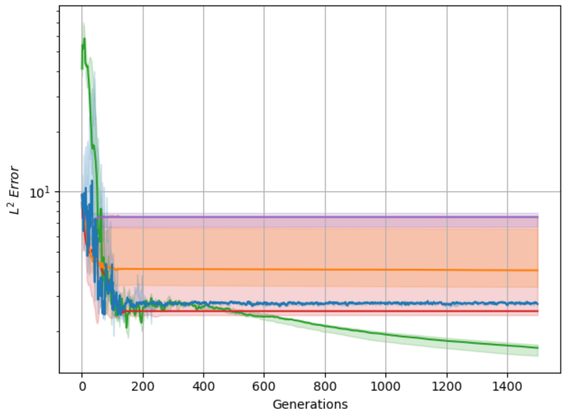


Full covariance

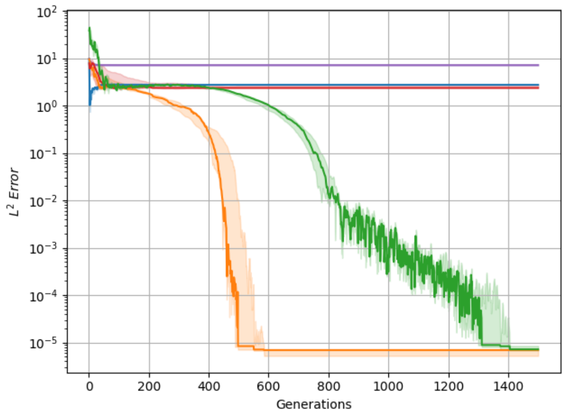


8D

Diagonal covariance

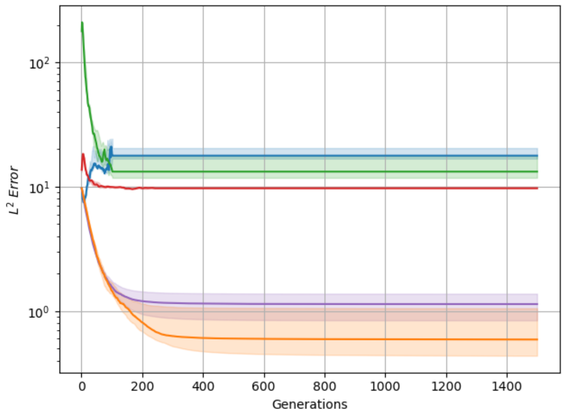


Full covariance



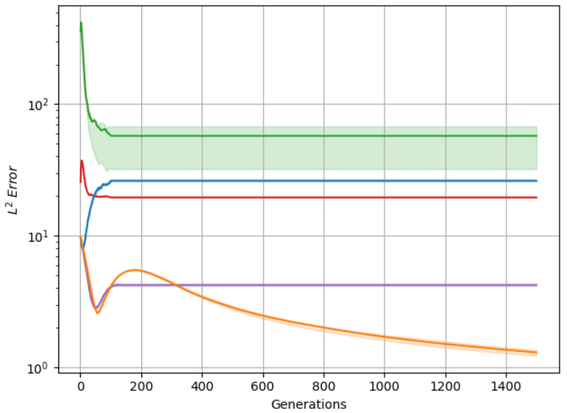
128D

Diagonal covariance



512D

Diagonal covariance



The lines correspond to MAP (—), SNM (—), NGA (—), SGA(—), and EM (—)

Evolution Strategies in Reinforcement Learning

Reinforcement learning optimization problem

In policy-based reinforcement learning, the policy $\pi_{\boldsymbol{\theta}}(\cdot \quad s_t)$ is parameterized by the vector $\boldsymbol{\theta} \in \mathbb{R}^d$ that computes the actions of the learning agent as a function of the states, [6], [7].

In terms of mathematical optimization, we seek to solve the following problem,

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^d} \mathbb{E}_{a_t \sim \pi_{\boldsymbol{\theta}}(\cdot \ s_t)} \left[\sum_{t=0}^{T-1} \gamma^t r(s_t, a_t) \,\middle|\, s_0 = s \right] ,$$

where γ is a discounting factor and $r(s_t, a_t)$ is the reward obtained in state s_t after taking action a_t.

MuJoCo results

Direct policy search to control continuous problems

Benchmark locomotion tasks: InvertedPendulum, Swimmer, Hopper (all three with full covariance), Walker and HalfCheetah locomotion tasks (these last two with diagonal covariance).

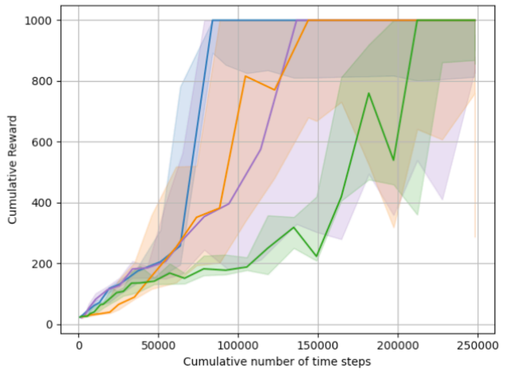
- **Linear policies:** the mapping between state $s_t \in \mathbb{R}^{N_S}$ and action $a_t \in \mathbb{R}^{N_A}$ at step t have the form

  $$a_t = s_t \cdot M_t(\boldsymbol{\theta}) \, .$$

  Here, M_t is a $N_S \times N_A$ matrix, N_S is the dimension of the states and N_A is the dimension of the actions.
- Discount factor of $\gamma = 1$, $T = 1000$ steps for each environment
- **Batch normalization of the states,** allowing a more isotropic exploration of the search domain, [7].
- **Individuals selection:** a weight of ≈ 0 has been assigned to the worst $25\,\%$ of the population at each generation.
- Normalized gradient ascent procedure instead of Adam method in NGA, [8]

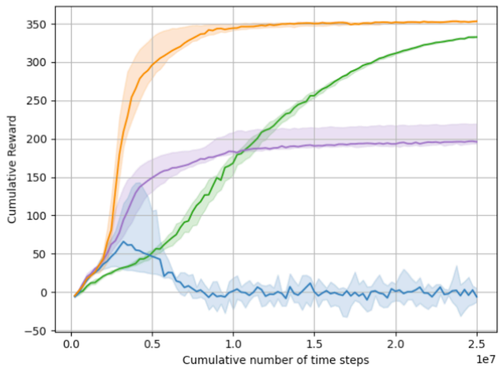
MuJoCo plots and summary of results

InvertedPendulum (4D)



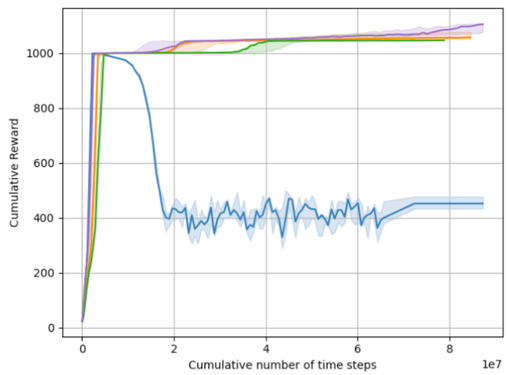
Full covariance

Swimmer (16D)



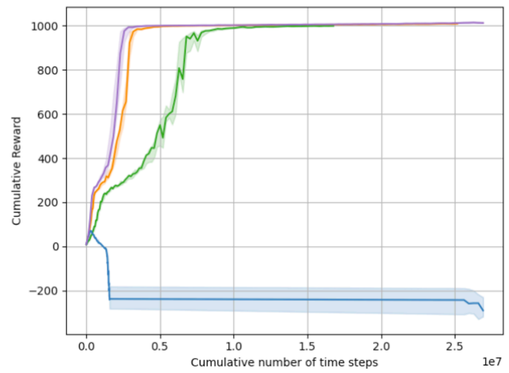
Full covariance

Hopper (33D)



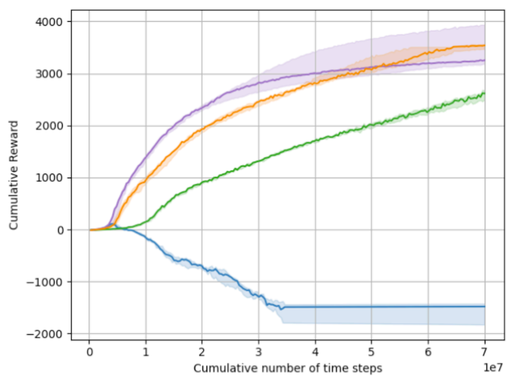
Full covariance

Walker (102D)



Diagonal covariance

HalfCheetah (102D)



Diagonal covariance

The lines correspond to MAP (—), SNM (—), NGA, (—), and EM (—)

	Average number of timesteps needed to reach reward threshold						
Environment	Threshold	TRPO	ES-Sal	EM	MAP	NGA	SNM
InvertedPendulum	1000	5.17e+05	4.55e+05	1.37e+05	8.40e+04	2.12e+05	1.44e+05
Swimmer	128.25	4.59e+06	1.39e+06	4.00e+06	N/A[1]	8.5e+06	2.75e+06
Hopper	3403.46	4.56e+06	3.16e+07	N/A	N/A	N/A	N/A
Walker2D	3830.03	4.81e+06	3.79e+07	N/A	N/A	N/A	N/A
HalfCheetah	2385.79	5.00e+06	2.88e+06	2.13e+07	N/A	6.58e+07	2.97e+07

	Maximum reward reached			
Environment	EM	MAP	NGA	SNM
InvertedPendulum	1000	1000	1000	1000
Swimmer	219	33	143	353
Hopper	1109	476	1046	1077
Walker2D	1017	74	999	1011
HalfCheetah	3937	133	2676	3561

Recap

- We derived different evolution strategies based on theory using gradient descent and maximum likelihood approach.

- We demonstrate that these algorithms can optimize benchmark functions and achieve state-of-the-art sample efficiency on control problems using linear policies.

- We found that some practical problems appear when using a strict-to-theory approach in the derivation of the algorithms.

Gitlab repository: https://gitlab.ethz.ch/michavan/bachelor-thesis-es

Bibliography

[1] D. H. Brookes, A. Busia, C. Fannjiang, K. Murphy, and J. List- garten, *A view of Estimation of Distribution Algorithms through the lens of Expectation-Maximization*, arXiv:1905.10474 [cs, stat], (2020).

[2] N. Hansen, *The CMA Evolution Strategy: A Tutorial*, arXiv:1604.00772 [cs, stat], (2016).

[3] L. S. L. Tan, *Analytic natural gradient updates for cholesky factor in gaussian variational approximation*, 2021.

[4] J. P. G. D. R. Haley and W. S. Levine, *Efficient maximum likelihood identification of a positive semi-definite covariance of initial population statistics*, American Control Conference, (1984), pp. 1085–1089.

[5] S. J. W. Jorge Nocedal, *Numerical Optimization*, 2, Springer New York, NY, 2006.

[6] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever, *Evolution Strategies as a Scalable Alternative to Reinforcement Learning*, arXiv:1703.03864 [cs, stat], (2017). arXiv: 1703.03864.

[7] H. Mania, A. Guy, and B. Recht, *Simple random search provides a competitive approach to reinforcement learning*, arXiv:1803.07055 [cs, math, stat], (2018). arXiv: 1803.07055.

[8] A. Rajeswaran, K. Lowrey, E. Todorov, and S. Kakade, *Towards generalization and simplicity in continuous control*, 2017.

[9] Salimbeni, Hugh & Eleftheriadis, Stefanos & Hensman, James, *Natural Gradients in Practice: Non-Conjugate Variational Inference in Gaussian Process Models*, 2018.

Appendix

Steepest direction in parameter space: intuition

Recall: we seek to solve, [1]

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} \Big(\underbrace{\log \mathbb{E}_{q^{(k+1)}(\mathbf{z}|\boldsymbol{\theta}^{(k)})} f(\mathbf{z}) p(\mathbf{z} \mid \boldsymbol{\theta})}_{:= \mathcal{Q}(\boldsymbol{\theta})} - D_{KL}\left(q^{(k+1)}\left(\mathbf{z} \mid \boldsymbol{\theta}^{(k)}\right) \quad \tilde{p}(\mathbf{z} \mid \boldsymbol{\theta})\right) \Big)$$

Second order Taylor expansion

$$2D_{KL}\left(q\left(\mathbf{z} \mid \boldsymbol{\theta}^{(k)}\right) \quad \tilde{p}(\boldsymbol{\theta} \mid \mathbf{z})\right) \approx \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}\right)^{\top} \mathcal{I}_{\boldsymbol{\theta}} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}\right) = \left\| \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}\right) \right\|^{2}_{\mathcal{I}_{\boldsymbol{\theta}}}$$

vs squared distance in Euclidean space: $\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}\right)^{\top} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}\right)$

Using Lagrange multiplier we can prove the direction a that maximizes $\mathcal{Q}(\boldsymbol{\theta} + a)$ is given by, [3]

$$a = \epsilon \tilde{g}_{\boldsymbol{\theta}} / \|\tilde{g}_{\boldsymbol{\theta}}\|_{\mathcal{I}_{\boldsymbol{\theta}}} \quad \text{with} \quad \tilde{g}_{\boldsymbol{\theta}} = \mathcal{I}_{\boldsymbol{\theta}}^{-1} \nabla_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}) \quad \text{and} \quad \epsilon = \text{small constant.}$$

Modified Newton-Raphson method approach

The general update scheme of the Newton-Raphson method has the form

$$\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} - \mathscr{H}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}^{(k)}\right) \nabla_{\boldsymbol{\theta}} \mathscr{J}\left(\boldsymbol{\theta}^{(k)}\right) ,$$

where $\mathscr{J}(\boldsymbol{\theta})$ is the negative weighted Gaussian distribution.

Let $\boldsymbol{\sigma} = \operatorname{vech}(\Sigma)$ and consider now only $\boldsymbol{\theta}_2 = \boldsymbol{s}$, [4]

$$\frac{\partial \mathscr{J}}{\partial \boldsymbol{s}} = Q \frac{\partial \mathscr{J}}{\partial \boldsymbol{\sigma}}, \quad \text{with} \quad Q_{ij} = \frac{\partial \boldsymbol{\sigma}_j}{\partial \boldsymbol{s}_i} .$$

The hessian of the Cholesky decomposition of the covariance matrix is given by

$$\frac{\partial^2 \mathscr{J}}{\partial \boldsymbol{s} \partial \boldsymbol{s}^T} = R + Q \underbrace{\frac{\partial^2 \mathscr{J}}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}^T}}_{\to \mathscr{J}_\Sigma} Q^T, \quad \text{with} \quad R_{ij} = \sum_{p=1}^{N} \frac{\partial^2 \boldsymbol{\sigma}_p}{\partial \boldsymbol{s}_i \partial \boldsymbol{s}_j^T} \frac{\partial \mathscr{J}}{\partial \boldsymbol{\sigma}_p}$$

Hessian correction in Newton-Raphson method

It can happen that the Hessian matrix may not be positive definite and therefore the Newton direction $p^{(k)} \equiv -\mathscr{H}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}^{(k)})\,\nabla_{\boldsymbol{\theta}}\mathscr{J}(\boldsymbol{\theta}^{(k)})$ may not be an ascent direction, [5].

We tested the following corrections:

- *Diagonal modification method (DiagSearch)*
- *Eigenvalue inversion method (EVinv)*
- *Eigenvalue substitution with fixed δ (EVdelta)*
- *Eigenvalue modification with minimum Euclidean nom (EVeucl)*
- *Eigenvalue modification with minimum Frobenius norm (EVfrob)*
- *Skipping method (UPskip)*
- *Modified symmetric indefinite factorization method (MSIF)*

Hessian correction in Newton-Raphson method

It can happen that the Hessian matrix may not be positive semi-definite and therefore the Newton direction $p^{(k)} \equiv -\mathscr{H}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}^{(k)})\,\nabla_{\boldsymbol{\theta}}\mathscr{J}(\boldsymbol{\theta}^{(k)})$ may not be an ascent direction, [5].

	Low dimension				High dimension			
	q_{20}	q_{80}	q_{50}	DQ	q_{20}	q_{80}	q_{50}	DQ
EVinv	91	517	244	426	964	2627	1134	1663
EVdelta	107	686	380	579	715	1887	1098	1172
EVfrob	115	536	209	421	915	2922	1261	2007
EVeucl	162	597	247	435	888	3429	1741	2541
UPskip	106	-	497	-	1125	-	1507	-
MSIF	251	735	465	484	440	1956	1305	1516

Benchmark: Rosenbrock function
Tolerance: 10^{-5}
Population: pop_{min}
Number of runs: 5
$\delta = 10^{-50}$

Imposing sparsity

Diagonal covariance: $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

Number unknown: $n(n+1)/2 \to n$, where n is the number of dimensions.

Gaussian distribution reads

$$\mathcal{N}(\mathbf{z} \mid \boldsymbol{\mu}, \Sigma) = \prod_{d=1}^{n} \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{1}{2\sigma_d^2}(z_d - \mu_d)^2 \right).$$

The weighted log Gaussian for one dimension is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \mathcal{N}(z_i \mid \mu, \sigma^2) = -\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N} w_i(z_i - \mu)^2 + \text{const}\,.$$

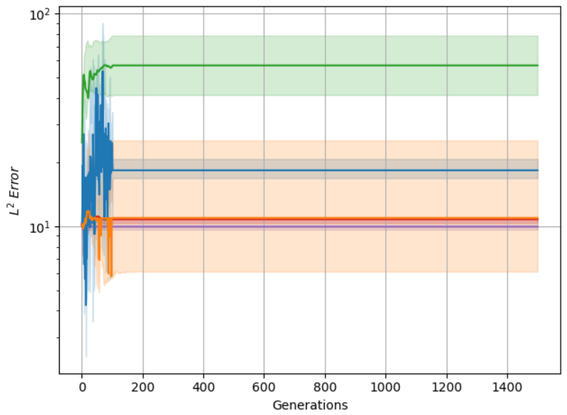
where $\boldsymbol{\theta} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$, $w_i \geq 0$ and $\sum_{i=1}^{N} w_i = 1$.

Griewank function

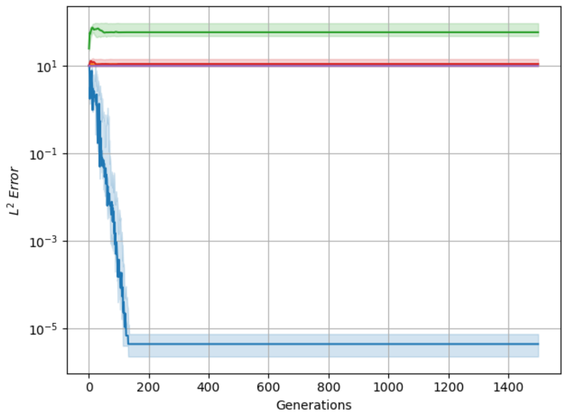
$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{x_i^2}{4000} - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i+1}}\right) + 1$$

2D

Diagonal covariance

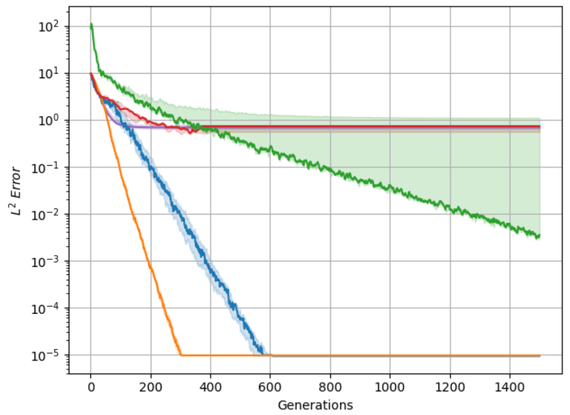


Full covariance

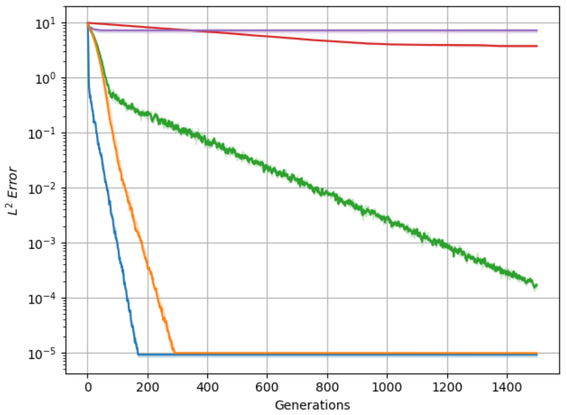


32D

Diagonal covariance

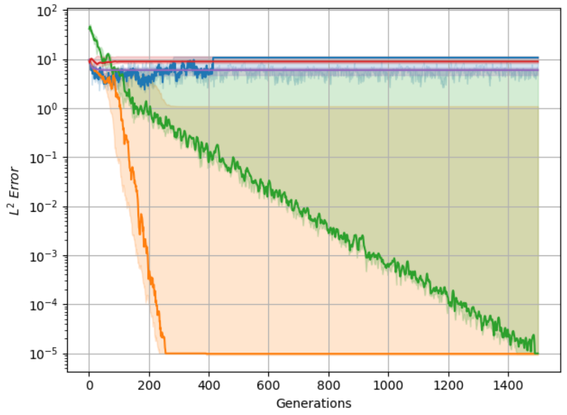


Full covariance

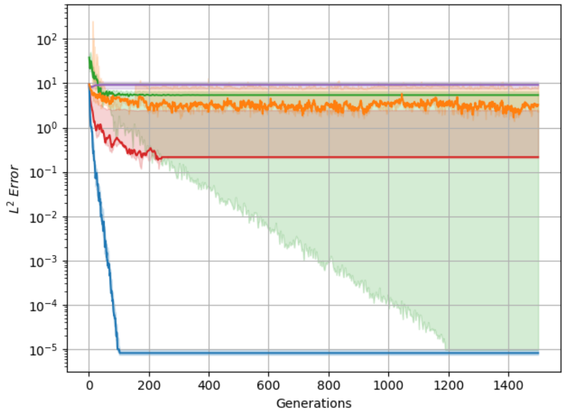


8D

Diagonal covariance

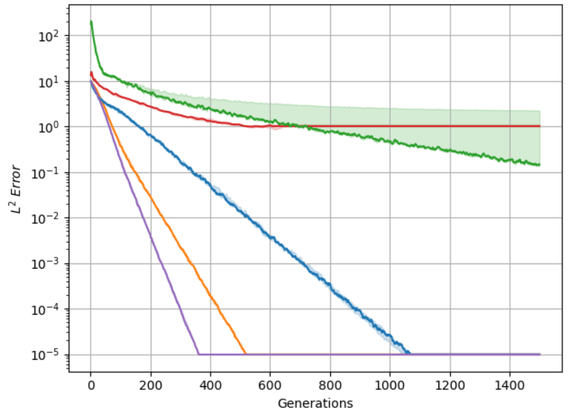


Full covariance



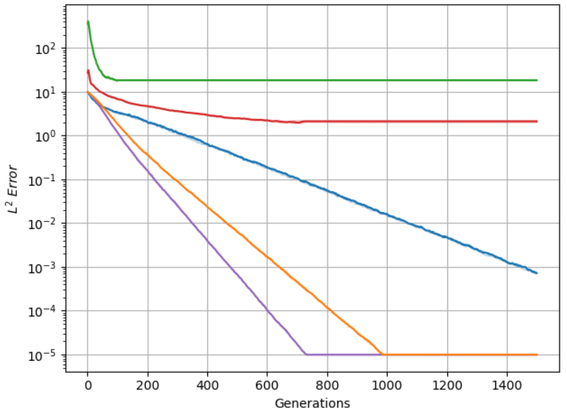
128D

Diagonal covariance



512D

Diagonal covariance



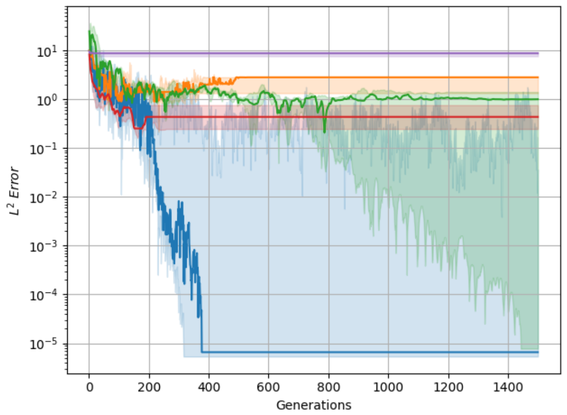
The lines correspond to MAP (—), SNM (—), NGA (—), SGA(—), and EM (—)

Rastrigin function

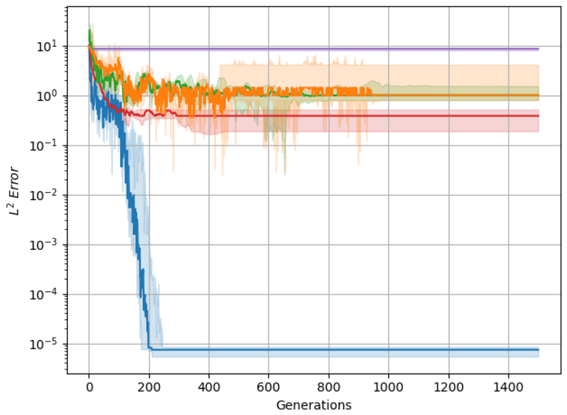
$$f(\mathbf{x}) = 10n + \sum_{i=1}^{n-1} \left[x_i^2 - 10\cos(2\pi x_i) \right]$$

2D

Diagonal covariance

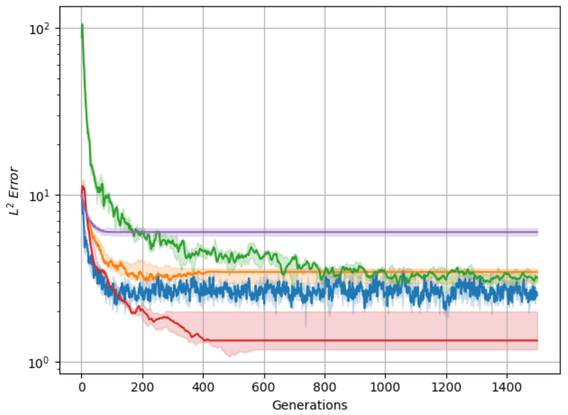


Full covariance

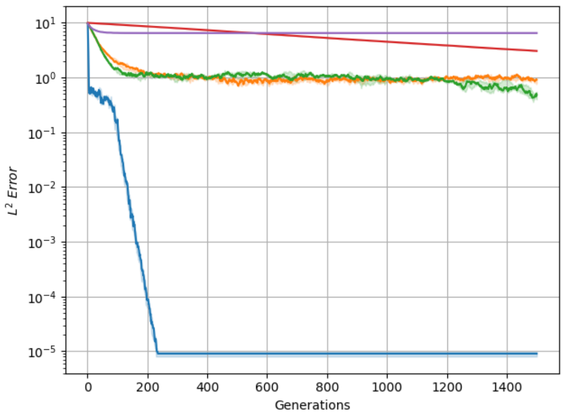


32D

Diagonal covariance

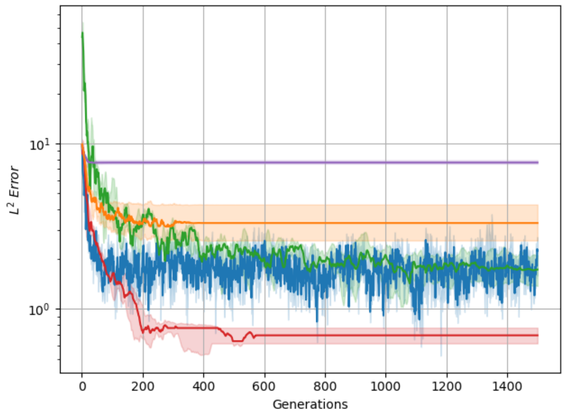


Full covariance

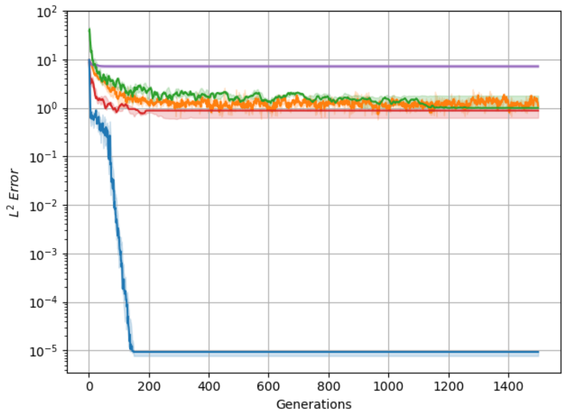


8D

Diagonal covariance

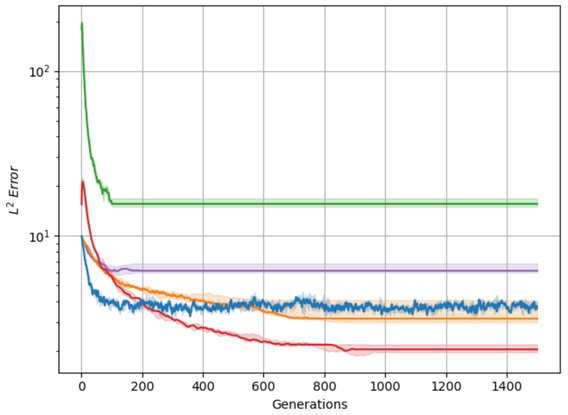


Full covariance



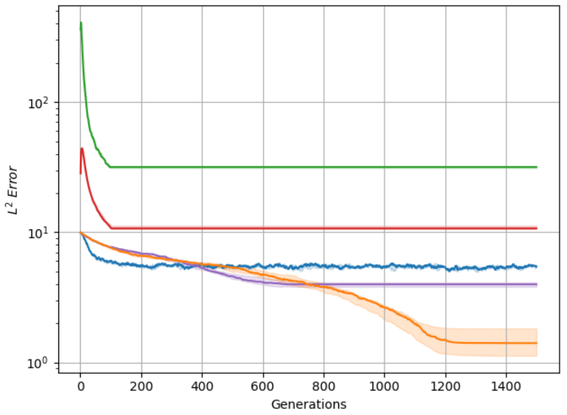
128D

Diagonal covariance



512D

Diagonal covariance



The lines correspond to MAP (—), SNM (—), NGA (—), SGA(—), and EM (—)