Supplementary Document of "Optimal Restart Strategies for Parameter-dependent Optimization Algorithms"

Lisa Schönenberger
lisa.schoenenberger@uni-ulm.de
Ulm University, Institute of Theoretical Computer Science
Ulm, Germany
Vorarlberg University of Applied Sciences
Research Center Business Informatics
Dornbirn, Austria

Hans-Georg Beyer
hans-georg.beyer@fhv.at
Vorarlberg University of Applied Sciences
Research Center Business Informatics
Dornbirn, Austria

This document provides the supplementary material to the paper "Optimal Restart Strategies for Parameter-dependent Optimization Algorithms". It contains all the detailed calculations and additional information regarding the experimental setup.

1 Proof of Theorem 3.2

Proof. The number of required restarts at $\hat{\lambda} = \lambda_k{+}1$ is $\hat{k}(\lambda_k{+}1){=}$ $k + 1$. It holds that $\lambda_k = \lambda_0 + k\nu$ and therefore

$$\begin{aligned}
\mathcal{L}^+(\lambda_k+1,\nu) &= \sum_{j=0}^{k+1}\lambda_j - \lambda_k - 1 = \sum_{j=0}^{k+1}(\lambda_0 + j\nu) - \lambda_0 - k\nu - 1\\
&= (k+2)\,\lambda_0 + \nu\sum_{j=0}^{k+1} j - \lambda_0 - k\nu - 1\\
&= (k+2)\,\lambda_0 + \frac{\nu}{2}\,(k+2)\,(k+1) - \lambda_0 - k\nu - 1\\
&= (k+1)\,\lambda_0 + \frac{\nu}{2}\left(k^2+3k+2\right) - k\nu - 1\\
&= (k+1)\,\lambda_0 + \frac{\nu}{2}\left(k^2+k+2\right) - 1\\
&= (k+1)\,\lambda_0 + \frac{\nu}{2}\left(k^2+k\right) + \nu - 1. \qquad (1)
\end{aligned}$$

Because

$$\lambda_k = \lambda_0 + k\nu \Leftrightarrow k = \frac{\lambda_k - \lambda_0}{\nu}, \qquad (2)$$

it follows that for all $k \geq 0$

$$\begin{aligned}
&\mathcal{L}^+(\lambda_k+1,\nu)\\
&= \left(\frac{\lambda_k-\lambda_0}{\nu}+1\right)\lambda_0 + \frac{\nu}{2}\left(\left(\frac{\lambda_k-\lambda_0}{\nu}\right)^2 + \frac{\lambda_k-\lambda_0}{\nu}\right) + \nu - 1\\
&= \frac{\lambda_0}{\nu}\,(\lambda_k-\lambda_0) + \lambda_0 + \frac{\nu}{2}\,(\lambda_k-\lambda_0)\left(\frac{\lambda_k-\lambda_0}{\nu^2}+\frac{1}{\nu}\right) + \nu - 1\\
&= (\lambda_k-\lambda_0)\left(\frac{\lambda_0}{\nu} + \frac{\lambda_k-\lambda_0}{2\nu} + \frac{1}{2}\right) + \lambda_0 + \nu - 1\\
&= (\lambda_k-\lambda_0)\left(\frac{2\lambda_0+\lambda_k-\lambda_0}{2\nu} + \frac{1}{2}\right) + \lambda_0 + \nu - 1\\
&= \frac{1}{2}\,(\lambda_k-\lambda_0)\left(\frac{\lambda_k+\lambda_0}{\nu}+1\right) + \lambda_0 + \nu - 1\\
&= \frac{1}{2}\,(\lambda_k+1-\lambda_0-1)\left(\frac{\lambda_k+1+\lambda_0-1}{\nu}+1\right) + \lambda_0 + \nu - 1\\
&= \mathcal{L}^+_{\text{up}}(\lambda_k+1,\nu). \qquad (3)
\end{aligned}$$

For $\hat{\lambda} = \lambda_0$ it holds that

$$\begin{aligned}
\mathcal{L}^+_{\text{up}}(\lambda_0,\nu) &= -\frac{1}{2}\left(\frac{2\lambda_0-1}{\nu}+1\right) + \lambda_0 + \nu - 1\\
&= \lambda_0\left(1-\frac{1}{\nu}\right) + \nu + \frac{1}{2\nu} - \frac{3}{2}\\
&\geq \nu + \frac{1}{2\nu} - \frac{3}{2}\\
&\geq 0 = \mathcal{L}^+(\lambda_0,\nu) \qquad (4)
\end{aligned}$$

for $\nu \geq 1$. In (15) from the main paper it was shown that $\mathcal{L}^+(\hat{\lambda},\nu)$ decreases between $\lambda_k + 1$ and λ_{k+1}. Because $\mathcal{L}^+_{\text{up}}$ is an increasing function of $\hat{\lambda}$ it holds for all $\hat{\lambda} \geq \lambda_0$ that

$$\mathcal{L}^+_{\text{up}}(\hat{\lambda},\nu) \geq \mathcal{L}^+(\hat{\lambda},\nu). \qquad (5)$$

□

2 Proof of Theorem 3.4

Proof. The number of required restarts at $\hat{\lambda} = \lambda_k$ is $\hat{k}(\lambda_k) = k$. Using $\lambda_0\rho^k + 1 > \lambda_k = \lceil \lambda_0\rho^k \rceil \geq \lambda_0\rho^k$, then

$$\mathcal{L}^\times(\lambda_k,\rho) = \sum_{j=0}^{k}\lambda_j - \lambda_k = \sum_{j=0}^{k-1}\lambda_j \geq \sum_{j=0}^{k-1}\lambda_0\rho^j = \lambda_0\frac{\rho^k-1}{\rho-1}. \qquad (6)$$

Additionally it holds that

$$\lambda_0\rho^k + 1 > \lambda_k \Leftrightarrow \rho^k > \frac{\lambda_k - 1}{\lambda_0}. \qquad (7)$$

Inserting this into (6), then it follows that

$$\mathcal{L}^\times(\lambda_k,\rho) > \lambda_0\frac{\frac{\lambda_k-1}{\lambda_0}-1}{\rho-1} = \frac{\lambda_k-1-\lambda_0}{\rho-1}. \qquad (8)$$

for all $k \geq 0$. In (15) from the main paper it was shown that $\mathcal{L}^\times(\hat{\lambda},\rho)$ decreases between $\lambda_{k-1}+1$ and λ_k. Therefore, it follows for $\lambda_{k-1}+1 \leq \hat{\lambda} < \lambda_k$ by using (8)

$$\mathcal{L}^\times(\hat{\lambda},\rho) > \mathcal{L}^\times(\lambda_k,\rho) > \frac{\lambda_k-1-\lambda_0}{\rho-1} > \frac{\hat{\lambda}-1-\lambda_0}{\rho-1} = \mathcal{L}^\times_{\text{low}}(\hat{\lambda},\rho), \qquad (9)$$

which holds for all $k \geq 1$. Because (8) includes the case λ_0, it follows for all $\hat{\lambda} \geq \lambda_0$ that

$$\mathcal{L}^\times(\hat{\lambda},\rho) > \mathcal{L}^\times_{\text{low}}(\hat{\lambda},\rho). \qquad (10)$$

□

3 Additional steps for Proof of Lemma 3.5

Assume that the condition $\mathcal{F}_{\text{up}}(k) > \mathcal{F}(k)$ holds for k, then it follows by induction

$$\begin{aligned}
\mathcal{F}_{\text{up}}(k+1) &= \lambda_0\rho + k + 1 + (\lambda_{k+1} - \lambda_0)\left(\rho + \frac{1}{\rho-1}\right) \\
&= \lambda_0\rho + k + (\lambda_k - \lambda_0)\left(\rho + \frac{1}{\rho-1}\right) \\
&- \lambda_k\left(\rho + \frac{1}{\rho-1}\right) + \lambda_{k+1}\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&= \mathcal{F}_{\text{up}}(k) + (\lambda_{k+1} - \lambda_k)\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&> \mathcal{F}(k) + (\lambda_{k+1} - \lambda_k)\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&= \sum_{j=0}^{k+1}\lambda_j - \lambda_k - 1 + (\lambda_{k+1} - \lambda_k)\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&= \sum_{j=0}^{k+2}\lambda_j - \lambda_{k+1} - 1 - \lambda_{k+2} + \lambda_{k+1} + 1 - \lambda_k - 1 \\
&+ (\lambda_{k+1} - \lambda_k)\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&= \mathcal{F}(k+1) - \lambda_{k+2} + \lambda_{k+1} - \lambda_k \\
&+ (\lambda_{k+1} - \lambda_k)\left(\rho + \frac{1}{\rho-1}\right) + 1 \\
&= \mathcal{F}(k+1) - \lambda_{k+2} + \lambda_{k+1} - \lambda_k + \lambda_{k+1}\rho - \lambda_k\rho \\
&+ \frac{\lambda_{k+1} - \lambda_k}{\rho-1} + 1 \\
&> \mathcal{F}(k+1) - \lambda_{k+1}\rho - 1 + \lambda_{k+1} - \lambda_k + \lambda_{k+1}\rho - \lambda_k\rho \\
&+ \frac{\lambda_{k+1} - \lambda_k}{\rho-1} + 1 \\
&\geq \mathcal{F}(k+1) + \lambda_k\rho - \lambda_k - \lambda_k\rho + \frac{\lambda_k\rho - \lambda_k}{\rho-1} \\
&= \mathcal{F}(k+1),
\end{aligned} \tag{11}$$

where (29) from main paper was used for the last two inequalities.

4 Additional steps for Proof of Lemma 3.7

Assume that the condition $\mathcal{F}_{\text{low}}(k) < \mathcal{F}(k)$ holds for k, then it follows by induction that

$$\begin{aligned}
\mathcal{F}_{\text{low}}(k+1) &= \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_0 - k - 1) \\
&= \frac{1}{\rho-1}(\lambda_k - \lambda_0 - k) + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&= \mathcal{F}_{\text{low}}(k) + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&< \mathcal{F}(k) + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&= \sum_{j=0}^{k}\lambda_j - \lambda_k + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&= \sum_{j=0}^{k+1}\lambda_j - \lambda_{k+1} - \lambda_k + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&= \mathcal{F}(k+1) - \lambda_k + \frac{1}{\rho-1}(\lambda_{k+1} - \lambda_k - 1) \\
&< \mathcal{F}(k+1) - \lambda_k + \frac{1}{\rho-1}(\lambda_k\rho + 1 - \lambda_k - 1) \\
&= \mathcal{F}(k+1).
\end{aligned} \tag{12}$$

where (29) from main paper was used for the last inequality.

5 Proof of Theorem 3.10

Proof. The loss function jumps at $\lambda_k + 1$ and the number of restarts is $\hat{k}(\lambda_k + 1) = k + 1$. It holds that

$$\lambda_0(k+1)^\alpha \leq \lceil\lambda_0(k+1)^\alpha\rceil = \lambda_k < \lambda_0(k+1)^\alpha + 1. \tag{13}$$

For the calculations to come the following estimation comes in handy

$$\sum_{j=1}^{k} j^\alpha < \int_0^k (x+1)^\alpha \mathrm{d}x = \frac{(k+1)^{\alpha+1}}{\alpha+1} - \frac{1}{\alpha+1} \tag{14}$$

as visualized in the right plot of Fig. 4 from main paper. Therefore, it holds for all $k \geq 0$ that

$$\begin{aligned}
\mathcal{L}^\#(\lambda_k + 1, \alpha) &= \sum_{j=0}^{k+1}\lambda_j - \lambda_k - 1 = \sum_{j=0}^{k+1}\lceil\lambda_0(j+1)^\alpha\rceil - \lambda_k - 1 \\
&< \sum_{j=0}^{k+1}\left(\lambda_0(j+1)^\alpha + 1\right) - \lambda_0(k+1)^\alpha - 1 \\
&= \lambda_0\sum_{j=0}^{k+1}(j+1)^\alpha + k + 2 - \lambda_0(k+1)^\alpha - 1 \\
&= \lambda_0\sum_{i=1}^{k+2} i^\alpha + k + 1 - \lambda_0(k+1)^\alpha \\
&< \lambda_0\frac{(k+3)^{\alpha+1}}{\alpha+1} - \frac{\lambda_0}{\alpha+1} + k + 1 - \lambda_0(k+1)^\alpha.
\end{aligned} \tag{15}$$

It follows from (13) that for $\lambda_k > \lambda_0$

$$(k+1)^\alpha \leq \frac{\lambda_k}{\lambda_0} \Leftrightarrow k \leq \sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} - 1 \tag{16}$$

$$(k+1)^\alpha > \frac{\lambda_k - 1}{\lambda_0} \Leftrightarrow k > \sqrt[\alpha]{\frac{\lambda_k - 1}{\lambda_0}} - 1. \tag{17}$$

Inserting this into (15), then it follows for $k \geq 0$ and $\lambda_k \geq 1$

$$
\begin{aligned}
&\mathcal{L}^{\#}(\lambda_k + 1, \alpha) \\
&< \frac{\lambda_0}{\alpha+1}\left(\sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} + 2\right)^{\alpha+1} - \frac{\lambda_0}{\alpha+1} + \sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} - \lambda_0 \frac{\lambda_k - 1}{\lambda_0} \\
&= \frac{\lambda_0}{\alpha+1}\left(\sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} + 2\right)^{\alpha+1} + \sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} - \frac{\lambda_0}{\alpha+1} - \lambda_k + 1 \\
&\leq \frac{\lambda_0}{\alpha+1}\left(\sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} + 2\right)^{\alpha+1} + \sqrt[\alpha]{\frac{\lambda_k}{\lambda_0}} - \frac{\lambda_0}{\alpha+1} - 0 \\
&= \mathcal{L}^{\#}_{\mathrm{up}}(\lambda_k + 1, \alpha).
\end{aligned} \tag{18}
$$

For $\hat{\lambda} = \lambda_0$ it holds that

$$
\begin{aligned}
\mathcal{L}^{\#}_{\mathrm{up}}(\lambda_0, \alpha) &= \frac{\lambda_0}{\alpha+1}\left(\sqrt[\alpha]{\frac{\lambda_0 - 1}{\lambda_0}} + 2\right)^{\alpha+1} + \sqrt[\alpha]{\frac{\lambda_0 - 1}{\lambda_0}} - \frac{\lambda_0}{\alpha+1} \\
&= \frac{\lambda_0}{\alpha+1}\left(\left(\sqrt[\alpha]{\frac{\lambda_0 - 1}{\lambda_0}} + 2\right)^{\alpha+1} - 1\right) + \sqrt[\alpha]{\frac{\lambda_0 - 1}{\lambda_0}} \\
&\geq \frac{\lambda_0}{\alpha+1}\left(2^{\alpha+1} - 1\right) + \sqrt[\alpha]{\frac{\lambda_0 - 1}{\lambda_0}} \\
&> 0 = \mathcal{L}^{\#}(\lambda_0, \alpha).
\end{aligned} \tag{19}
$$

In (15) from main paper it was shown that $\mathcal{L}^{\#}(\hat{\lambda}, \alpha)$ decreases between $\lambda_k + 1$ and λ_{k+1}. Because $\mathcal{L}^{\#}_{\mathrm{up}}$ is an increasing function of $\hat{\lambda}$ it holds for all $\hat{\lambda} \geq \lambda_0$ that

$$
\mathcal{L}^{\#}_{\mathrm{up}}(\hat{\lambda}, \alpha) > \mathcal{L}^{\#}(\hat{\lambda}, \alpha). \tag{20}
$$

□

6 Experimental Setup

The experimental results in the right plot of Fig. 8 from main paper are the results from 40 different Rastrigin and Ackley landscapes. The details are given in Tables 1 and 2.

Table 1: List of different Rastrigin landscapes (see Eq. (63) from main paper), the maximum number of generations for a single run $g_{\max}$, and the optimal value of λ used for the experiments in the right plot of Fig. 8.

A	α	N	$g_{\max}$	$\hat{\lambda}$	A	α	N	$g_{\max}$	$\hat{\lambda}$
1	2π	100	500	138	10	2π	30	500	530
1	2π	200	1000	232	10	2π	50	750	610
1	2π	500	1000	452	10	2π	100	1000	1084
4	2π	30	300	168	1	2.5π	30	300	74
4	2π	50	500	264	1	2.5π	50	300	132
4	2π	100	500	466	1	2.5π	100	500	200
4	2π	200	1000	800	1	4π	30	500	166
6	2π	30	500	248	1	4π	50	500	260
6	2π	50	500	388	1	4π	100	500	460
6	2π	100	750	680	1	6π	30	500	556
8	2π	30	500	344	1	6π	50	500	1074
8	2π	50	500	504	1	6π	100	750	1006
8	2π	100	750	872					

Table 2: List of different Ackley landscapes (see Eq. (64) from main paper) with initial start distance R_{init}, initial normalized mutation strength $\sigma^*_{\mathrm{init}} = 3$, the maximum number of generations for a single run $g_{\max}$, and the optimal value of λ used for the experiments in the right plot of Fig. 8.

N	R_{init}	$g_{\max}$	$\hat{\lambda}$	N	R_{init}	$g_{\max}$	$\hat{\lambda}$
100	$10\sqrt{N}$	500	28	50	$35\sqrt{N}$	1000	144
200	$10\sqrt{N}$	1000	34	100	$35\sqrt{N}$	1000	124
300	$10\sqrt{N}$	2000	42	200	$35\sqrt{N}$	1000	136
500	$10\sqrt{N}$	5000	70	100	$40\sqrt{N}$	1000	586
30	$30\sqrt{N}$	300	48	200	$40\sqrt{N}$	2000	444
50	$30\sqrt{N}$	500	42	300	$40\sqrt{N}$	2000	486
100	$30\sqrt{N}$	750	44	100	$42\sqrt{N}$	1000	2800
30	$35\sqrt{N}$	300	296				