Utilizing the Expected Gradient in Surrogate-assisted Evolutionary Algorithms

Kei NISHIHARA, Masaya NAKATA Yokohama National University, Japan

YNU YOKOHAMA National University

Mail Paper (PDF)

Surrogate-assisted Evolutionary Algorithm (SAEA)

- **SAEAs are an effective approach to addressing expensive optimization problems (EOPs)**
  - Function evaluations (FEs) in EOPs are computationally or financially expensive
  - SAEAs estimate a promising solution among candidates by assessing their quality with surrogates
  - Surrogates usually approximate the objective functions

    Gaussian Process (GP), Radial Basis Function Network (RBFN), etc. …

- **Modern SAEAs alternates global and local search phases**

Global	Phase	Local
Whole landscape	Approximation target	Promising region
Find a new region	Purpose	Intensive search

e.g.) ω = 30 in GORS-SSLPSO [Yu+ 19] and SAHO [Pan+ 21]

- **Many SAEAs set a small number of generations ω** [Cai+ 19]
  - to reduce the runtime?
  - to prevent solutions from being guided to the wrong region?

RQ How to sufficiently optimize the approximate objective function?

Expected Gradient in GP

Objective function $f : \mathbb{R}^D \to \mathbb{R}$

Dataset $\{(x_i, f(x_i))\}_{i=1}^{n}$ ($x_i \in \mathbb{R}^D$)

The approximation of $f(x)$ $\hat{f}(x) = \mu + k_x^T K^{-1}(f - 1\mu),\ \mu = \dfrac{1^T K^{-1} f}{1^T K^{-1} 1}$

Gaussian correlation for the dth dimensional deviation $k_{ij,d}(x_{i,d}, x_{j,d}) = \exp(-\theta_d \|x_{i,d} - x_{j,d}\|^2)$

Correlation function matrix whose elements K (size: $n \times n$) $k_{ij}(x_i, x_j) = \prod_{d=1}^{n} k_{ij,d}(x_{i,d}, x_{j,d})$

Correlation vector for x and each in the dataset k_x (size: $n \times 1$)

Since the differentiation calculation is a linear operation,
if the process is mean-square differentiable,

The Expected Gradient is equivalent to the gradient of *the expected function value.* (*the approximate objective function*)

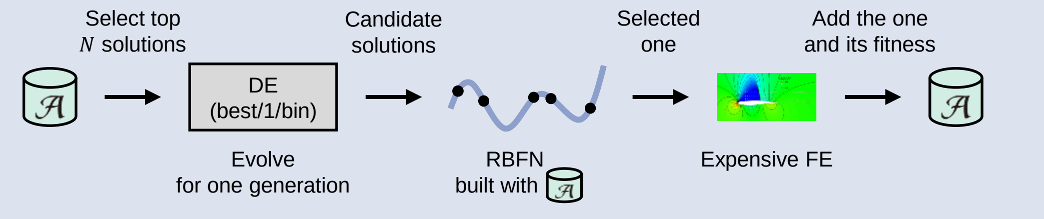
$$\hat{g}(x) = \left[\frac{\partial \hat{f}(x)}{\partial x_1}, \ldots, \frac{\partial \hat{f}(x)}{\partial x_d}, \ldots, \frac{\partial \hat{f}(x)}{\partial x_D}\right] = J(x)^T K^{-1}(f - 1\mu) \qquad J(x)_{i,d} = \frac{\partial k(x_{i,d}, x_d)}{\partial x_d}$$

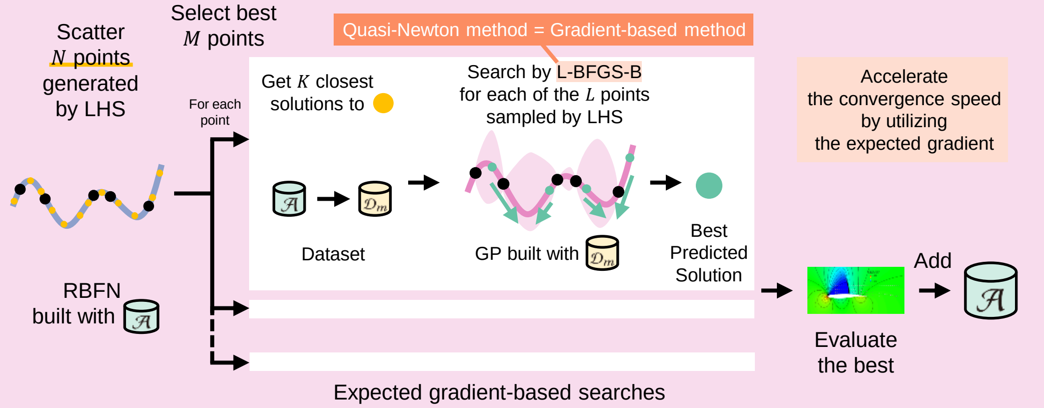
→ Gradient-based searches can be applied!!

Proposal: expected gradient-based SAEA

Initialization Phase
Get N samples with Latin Hypercube Sampling (LHS) and Evaluate them
Construct an archive $\mathcal{A}$ with initial samples and their fitness values

while *terminal criteria are not met*





Experiment

- **Experimental Design**
  - IEEE CEC'13 benchmark suite (Single-obj., Real-coded) [Liang+ 13]

Number of functions	28
Problem dimension D	10, 30
Maximum number of FEs	1,000
Number of runs	15

  - Compared Algorithm (Parameter settings follow the papers.)

GP	RBFN
GPEME [Liu+ 14]	S-JADE* [Cai+ 19]
IKAEA [Zhan+ 21]	SAHO [Pan+ 21]
GSGA* [Cai+ 20], Proposal*	

* : SAEAs that alternate global and local search phases

Parameter settings of our proposal
$N = 100, F = 0.5, CR = 0.9,$
$M = 3, K = 50, L = 5D$

Wilcoxon's rank-sum test
(significance level = 0.05)
+ : our proposal underperforms
- : our proposal outperforms
~ : cannot find significance

best worst

- **Results**

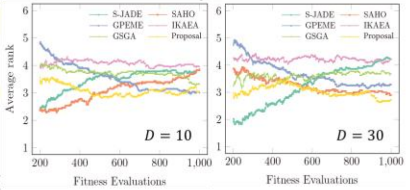
Fitness values (1,000 FEs, D = 30 as an example)

	S-JADE	SAHO	GPEME	IKAEA	GSGA	Proposal
F1	6.92E+00 +	1.88E-15 +	6.71E+02 −	3.12E-02 +	3.48E-04 +	2.75E+02
F2	9.40E+07 −	1.06E+07 +	1.41E+08 −	7.57E+07 −	1.05E+08 −	3.61E+07
F3	2.07E+15 ~	4.05E+17 −	4.59E+11 ~	1.81E+16 ~	2.95E+11 +	5.69E+13
F4	8.40E+04 +	1.25E+05 ~	1.75E+05 −	1.06E+05 ~	1.61E+05 −	1.17E+05
F5	3.12E+03 ~	1.79E+02 +	1.34E+03 +	3.07E+03 ~	2.75E+03 ~	2.53E+03
F6	1.08E+02 ~	4.22E+01 +	7.66E+01 +	2.02E+02 −	1.05E+02 ~	1.28E+02
F7	2.06E+04 −	2.09E+05 −	1.13E+03 ~	1.11E+05 ~	4.49E+02 +	2.61E+03
F8	2.12E+01 ~	2.12E+01 ~	2.12E+01 ~	2.12E+01 ~	2.12E+01 ~	2.12E+01
F9	3.75E+01 −	2.97E+01 ~	2.85E+01 ~	4.40E+01 −	3.87E+01 −	2.96E+01
F10	5.84E+01 ~	1.25E+00 +	2.98E+02 −	9.39E+00 +	1.22E+02 −	6.44E+01
F11	2.87E+02 −	2.80E+02 −	1.69E+02 −	2.97E+02 −	2.52E+02 −	1.24E+02
F12	3.02E+02 −	2.39E+02 −	2.94E+02 −	3.00E+02 −	2.87E+02 −	1.38E+02
F13	3.18E+02 −	3.00E+02 −	2.98E+02 −	2.96E+02 −	3.33E+02 −	2.58E+02
F14	7.90E+03 −	6.14E+03 ~	5.48E+03 ~	6.36E+03 ~	7.05E+03 −	5.30E+03
F15	8.67E+03 −	6.65E+03 ~	8.90E+03 −	8.80E+03 −	8.62E+03 −	7.11E+03
F16	4.51E+00 ~	4.59E+00 ~	4.46E+00 ~	4.74E+00 ~	4.58E+00 ~	4.40E+00
F17	2.74E+02 ~	2.70E+02 ~	2.56E+02 ~	3.14E+02 −	2.85E+02 −	2.44E+02
F18	2.91E+02 +	2.92E+02 ~	3.28E+02 ~	3.24E+02 ~	3.44E+02 ~	3.21E+02
F19	4.67E+04 ~	2.95E+05 −	7.49E+03 +	8.21E+03 +	1.88E+02 +	4.39E+04
F20	1.50E+01 ~	1.50E+01 ~	1.48E+01 ~	1.50E+01 −	1.50E+01 −	1.49E+01
F21	2.41E+03 +	4.34E+03 −	4.66E+03 −	2.43E+03 +	1.56E+03 +	2.75E+03
F22	8.47E+03 −	6.62E+03 ~	5.90E+03 ~	6.74E+03 ~	7.55E+03 −	5.68E+03
F23	9.17E+03 −	6.42E+03 +	9.28E+03 −	9.34E+03 −	9.06E+03 −	7.66E+03
F24	2.99E+02 −	2.88E+02 ~	2.72E+02 +	2.99E+02 −	3.03E+02 −	2.84E+02
F25	3.16E+02 −	3.02E+02 −	2.84E+02 +	3.34E+02 −	3.08E+02 −	2.93E+02
F26	3.35E+02 ~	3.59E+02 ~	3.85E+02 −	3.58E+02 ~	3.64E+02 ~	3.50E+02
F27	1.17E+03 −	1.08E+03 ~	1.03E+03 ~	1.49E+03 −	1.28E+03 −	1.08E+03
F28	4.65E+03 ~	7.51E+03 −	5.38E+03 −	5.37E+03 ~	4.03E+03 ~	4.16E+03
+/−/~	4/13/11	6/9/13	5/12/11	4/13/11	5/16/7	

Wilcoxon's rank-sum test (+/-/~)

D	FE	vs S-JADE	vs SAHO	vs GPEME	vs IKAEA	vs GSGA
10	200	11/ 1/16	12/ 0/16	2/12/14	5/10/13	6/ 5/17
	400	7/ 8/13	9/ 2/17	5/11/12	4/13/11	7/12/ 9
	600	8/11/ 9	6/10/12	7/11/10	4/11/13	7/15/ 6
	800	7/13/ 8	5/13/10	5/11/12	4/12/12	7/13/ 8
	1,000	7/13/ 8	5/13/10	8/ 9/11	5/13/10	7/10/11
30	200	12/ 1/15	4/ 6/18	0/14/14	2/15/11	6/ 5/17
	400	8/ 4/16	9/ 6/13	3/ 9/16	2/ 8/18	4/10/14
	600	6/ 7/15	8/ 7/13	4/ 8/16	4/ 7/17	5/12/11
	800	7/11/10	6/10/12	4/10/14	5/11/12	6/14/ 8
	1,000	4/13/11	6/ 9/13	5/12/11	4/13/11	5/16/ 7

Average rank



An expected gradient-based intensive search succeeded in improving the performance of SAEA.