Exact Formulas for the Computation of Expected Tchebycheff Improvement

Liang Zhao and Qingfu Zhang

Department of Computer Science
City University of Hong Kong

July 3, 2023

Outline

Expected Tchebycheff Improvement (ETI)

- a popular acquisition function in expensive multiobjective optimization.

Existing methods for ETI computation

- approximation methods.

Our work

- exact formulas for the computation of ETI.

Expensive Multiobjective Optimization Problem (MOP):

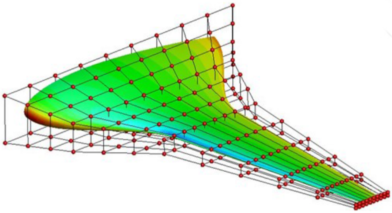
$$\begin{aligned} \text{minimize} \quad & F(x) = (f_1(x), \ldots, f_m(x))^T \\ \text{subject to} \quad & x \in \Omega \subseteq \mathbb{R}^d \end{aligned} \tag{1}$$

where

- evaluation of $F(x)$ is **very expensive**.
- evolutionary algorithms: difficult or impossible to directly use.

Efficient global optimization (EGO)

- ParEGO (J. Knowles, 2006)
- EHVI-EGO (Emmerich et al, 2006)
- MOEA/D-EGO (Zhang et al., 2009)
- ...



Aerodynamic shape optimization

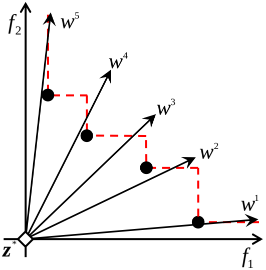
ParEGO VS MOEA/D-EGO

Decomposition

- Tchebycheff scalarization: $g^{te}(F(x)|\boldsymbol{w})$
- MOP -> single-objective subproblem

ParEGO

- randomly generates a weight vector $\boldsymbol{w}^i$
- Gaussian process (GP) model -> $g^{te}(F(x)|\boldsymbol{w}^i)$
- single-objective Expected Improvement (EI)
- select one query point: $x^* = \arg\max_{x\in\Omega} \mathrm{EI}(x)$



MOEA/D-EGO

- GP model -> $f_i(x)$
- Expected Tchebycheff Improvement (ETI): $\mathrm{EI}_T(x|\boldsymbol{w}^i), i = 1, \ldots, N$
- $Q = \arg\max_{x\in\Omega} \mathrm{EI}_T(x|\boldsymbol{w}^i), i = 1, \ldots, N$
- subset selection: select q query points from Q.

Advantages

- readily supports parallel expensive multiobjective optimization.
- achieves a good balance between convergence and diversity.

Disadvantages

- the computation of ETI is complex, due to the maximum of several Gaussians is not a Gaussian [1].

  Let $\eta_i := w_i(y_i - z_i^*)$. $p(\eta_i) = \mathcal{N}(w_i(\hat{\mu}_i(\boldsymbol{x}) - z_i^*), w_i^2\hat{\sigma}_i^2(\boldsymbol{x}))$.

$$\mathrm{EI}_T(x|\boldsymbol{w}) = \mathbb{E}_{\boldsymbol{y}\sim p(\boldsymbol{y}|x,\mathcal{D})}\left[\max(0, g^* - \max_{1\le i\le m}\{w_i(y_i - z_i^*)\})\right], \tag{2}$$

where $g^* = \min_{\boldsymbol{u}\in\mathcal{D}} g^{te}(\boldsymbol{u}|\boldsymbol{w})$, $\boldsymbol{y} = F(x)$, $\boldsymbol{w} = (w_1, \ldots, w_m)^T$.

[1] Nadarajah, S. and Kotz, S., 2008. Exact distribution of the max/min of two Gaussian random variables.

Approximation methods for the computation of ETI

- numerical integration (e.g., Gauss-Hermite quadrature)
- Monte Carlo methods (time-consuming)
- Moment Matching Approximation
  approximate $p(\max_{1 \leq i \leq m} \eta_i)$ with a Gaussian (Zhang et al., 2009)
- approximate $p(\max_{1 \leq i \leq m} \eta_i)$ with a Gumbel distribution (T. Chugh, 2022)

Limitation of approximation methods

- the estimation error of these methods cannot be fully determined without the aid of exact formulas.

Research Goal

- to derive an exact formula for the computation of ETI.
- to analyze the performance of the widely-used approximation method.

Outline

Similarity between Multi-point EI and ETI

- Multi-point EI (i.e., q-EI):

$$q\text{-}EI(X) = \mathbb{E}_{y^i \sim p(y^i|x^i,\mathcal{D})}\left[\max(0, f^* - \min_{1\leq i\leq q} y^i)\right], \quad (3)$$

  where f^* is the minimum of observed objective values.

- Expected Tchebycheff Improvement (ETI):

$$\mathrm{EI}_T(x|\boldsymbol{w}) = \mathbb{E}_{\boldsymbol{y}\sim p(\boldsymbol{y}|x,\mathcal{D})}\left[\max(0, g^* - \max_{1\leq i\leq m}\{w_i(y_i - z_i^*)\})\right], \quad (4)$$

  where g^* is the minimum of $g^{te}(F(x)|\boldsymbol{w})$ on observed samples.

-

$$f^* - \min_{1\leq i\leq q} y^i = f^* + \max_{1\leq i\leq q}\{-y^i\}. \quad (5)$$

Exact Formula of ETI

Inspired by the exact computation method for multi-point EI [2]

$$\mathrm{EI}_T(x|\boldsymbol{w}) = \sum_{k=1}^{m} \mathbb{P}(\boldsymbol{Z}^{(k)} \leq \boldsymbol{b}^k)(g^* - \mathbb{E}[Z_k^{(k)}|\boldsymbol{Z}^{(k)} \leq \boldsymbol{b}^{(k)}]) \tag{6}$$

Proposition (Tallis formulas)

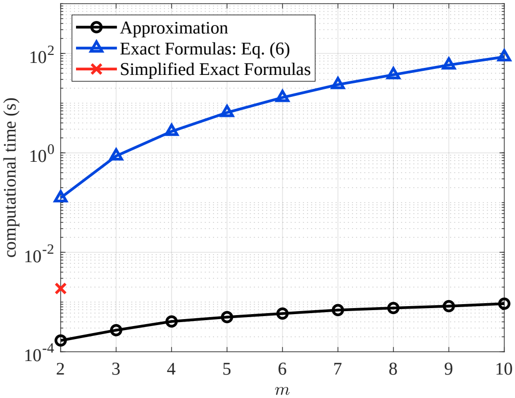
Let $\boldsymbol{Z} := (Z_1, \ldots, Z_m)$ *be a random vector obeying multivariate Gaussian distribution, i.e.,* $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *where* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T \in \mathbb{R}^m$ *and* $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$. *Let* $\boldsymbol{b} = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ *and* $\boldsymbol{Z} \leq \boldsymbol{b}$ *denotes* $\forall i, Z_i \leq b_i$. *For any* $k \in \{1, \ldots, m\}$, *the conditional expectation of* Z_k *given* $\boldsymbol{Z} \leq \boldsymbol{b}$ *is:*

$$\mathbb{E}(Z_k|\boldsymbol{Z} \leq \boldsymbol{b}) = \mu_k - \frac{1}{p} \sum_{i=1}^{m} \frac{\Sigma_{ik}}{\sqrt{\Sigma_{ii}}} \phi(\frac{b_i - \mu_i}{\sqrt{\Sigma_{ii}}}) p_i, \tag{7}$$

[2] C. Chevalier and D. Ginsbourger, 2013. Fast computation of the multi-points expected improvement with applications in batch selection.

Limitation

- relies on the specific technique to compute the multivariable Gaussian cumulative distribution function (cdf).
- is expensive to calculate



Computational time

- Moment Matching Approximation
- Exact formula, i.e., Eq. (6)
- Simplified exact formulas (independent assumption for $\eta_1, \ldots, \eta_m$)

Outline

Experimental settings

Compared methods

- MOEA/D-EGO-A (Moment Matching Approximation)
- MOEA/D-EGO-E (Exact formulas)

Test problems

- ZDT test suite
- F1-F4
- DTLZ test suite

General settings

- number of decision variables d: 8 for $m = 2$, 6 for $m = 3$
- number of initial samples: $11d - 1$
- maximal number of function evaluation: 200
- number of independent runs: 21

Mean IGD+ values ($m = 2$, Batch size: 5)

- perform similarly on the majority of test problems (13 out of 16).
- DTLZ2 and DTLZ5:MOEA/D-EGO-E outperforms MOEA/D-EGO-A.

Problem	m	d	MOEA/D-EGO-A	MOEA/D-EGO-E
ZDT1	2	8	9.9449e-3 (1.94e-3) $\approx$	1.0263e-2 (2.04e-3)
ZDT2	2	8	2.2673e-2 (1.11e-2) $\approx$	2.4315e-2 (7.31e-3)
ZDT3	2	8	1.2522e-2 (2.69e-3) $\approx$	2.1716e-2 (4.60e-2)
ZDT4	2	8	3.4847e+1 (1.14e+1) $+$	4.4925e+1 (1.76e+1)
ZDT6	2	8	4.3851e-1 (2.31e-1) $\approx$	3.8072e-1 (1.24e-1)
F1	2	8	5.5393e-3 (3.80e-4) $\approx$	5.3017e-3 (4.04e-4)
F2	2	8	2.0022e-2 (3.72e-3) $\approx$	2.1768e-2 (1.16e-2)
F3	2	8	1.7296e-2 (5.67e-3) $\approx$	2.4403e-2 (2.39e-2)
F4	2	8	1.8654e-2 (5.68e-3) $\approx$	1.8467e-2 (7.64e-3)
DTLZ1	2	8	9.0484e+1 (2.44e+1) $\approx$	8.9805e+1 (3.00e+1)
DTLZ2	2	8	1.4233e-2 (3.27e-3) $-$	9.7645e-3 (1.93e-3)
DTLZ3	2	8	2.0846e+2 (5.04e+1) $\approx$	2.1182e+2 (4.96e+1)
DTLZ4	2	8	2.6053e-1 (1.00e-1) $\approx$	2.9041e-1 (7.81e-2)
DTLZ5	2	8	1.3539e-2 (2.93e-3) $-$	9.8923e-3 (2.53e-3)
DTLZ6	2	8	8.0241e-1 (4.49e-1) $\approx$	1.0289e+0 (5.90e-1)
DTLZ7	2	8	1.0883e-1 (1.41e-1) $\approx$	7.6586e-2 (1.06e-1)
$+/-/\approx$			1/2/13	

Mean IGD+ values on DTLZ with different batch sizes

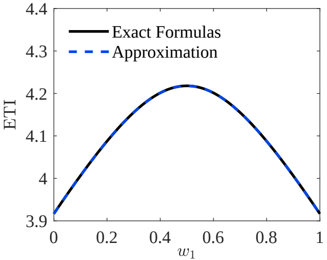
- perform similarly on the majority of test problems.
- DTLZ2 and DTLZ5:MOEA/D-EGO-E outperforms MOEA/D-EGO-A.

Problem	Batch Size	MOEA/D-EGO-A	MOEA/D-EGO-E
DTLZ1	1	9.0191e+1 (1.99e+1) ≈	7.9933e+1 (2.12e+1)
	2	9.1206e+1 (2.52e+1) ≈	8.9113e+1 (2.18e+1)
	5	9.0484e+1 (2.44e+1) ≈	8.9805e+1 (3.00e+1)
	8	9.1492e+1 (2.36e+1) ≈	8.9204e+1 (2.25e+1)
DTLZ2	1	8.9393e-3 (1.95e-3) −	6.3421e-3 (2.75e-3)
	2	9.4724e-3 (1.65e-3) −	7.6098e-3 (2.09e-3)
	5	1.4233e-2 (3.27e-3) −	9.7645e-3 (1.93e-3)
	8	1.3527e-2 (2.98e-3) −	9.7923e-3 (1.67e-3)
DTLZ3	1	2.0715e+2 (3.86e+1) ≈	1.9265e+2 (4.16e+1)
	2	2.0187e+2 (4.18e+1) ≈	2.0160e+2 (5.64e+1)
	5	2.0846e+2 (5.04e+1) ≈	2.1182e+2 (4.96e+1)
	8	2.0796e+2 (6.34e+1) ≈	1.9443e+2 (6.81e+1)
DTLZ4	1	2.0105e-1 (9.16e-2) ≈	2.1201e-1 (7.91e-2)
	2	2.3758e-1 (8.07e-2) ≈	2.2104e-1 (8.95e-2)
	5	2.6053e-1 (1.00e-1) ≈	2.9041e-1 (7.81e-2)
	8	2.9145e-1 (7.14e-2) ≈	3.2500e-1 (6.96e-2)
DTLZ5	1	8.0689e-3 (1.91e-3) −	6.5651e-3 (5.03e-3)
	2	1.0556e-2 (2.47e-3) −	7.2783e-3 (1.77e-3)
	5	1.3539e-2 (2.93e-3) −	9.8923e-3 (2.53e-3)
	8	1.4231e-2 (2.88e-3) −	9.5625e-3 (1.95e-3)
DTLZ6	1	1.1214e+0 (4.87e-1) ≈	1.2910e+0 (6.17e-1)
	2	8.4697e-1 (4.39e-1) ≈	1.1063e+0 (5.53e-1)
	5	8.0241e-1 (4.49e-1) ≈	1.0289e+0 (5.90e-1)
	8	7.8318e-1 (5.76e-1) ≈	1.2757e+0 (7.43e-1)
DTLZ7	1	6.2493e-2 (1.13e-2) ≈	7.0531e-2 (5.87e-2)
	2	9.1616e-2 (6.73e-2) ≈	9.3847e-2 (8.60e-2)
	5	1.0883e-1 (1.41e-1) ≈	7.6586e-2 (1.06e-1)
	8	3.8449e-2 (7.54e-2) ≈	3.2312e-2 (6.18e-2)
+/−/≈		0/8/20	

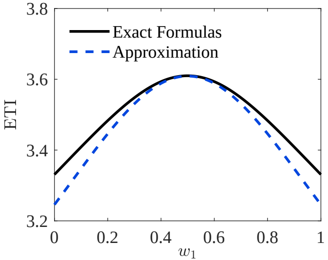
Discussion

Why the approximation method performs similarly with the exact formula on many of the test problems?

- (a) vs. (b): approximation error $\searrow$ as $\hat{\sigma}_i^2 \searrow$.
- predicted variances of the GP models get smaller during optimization.



(a) $\hat{\mu}_1 = 1, \hat{\mu}_2 = 1, \hat{\sigma}_1 = 1, \hat{\sigma}_2 = 1$



(b) $\hat{\mu}_1 = 1, \hat{\mu}_2 = 1, \hat{\sigma}_1 = 10, \hat{\sigma}_2 = 10$

Mean IGD+ values on DTLZ ($m = 3$, Batch size: 5)

- DTLZ2 and DTLZ5:MOEA/D-EGO-A outperforms MOEA/D-EGO-E.

Problem	m	d	MOEA/D-EGO-A	MOEA/D-EGO-E
DTLZ2	3	6	5.0246e-2 (3.53e-3) $+$	7.2728e-2 (1.82e-2)
DTLZ5	3	6	1.0300e-2 (1.35e-3) $+$	1.5099e-2 (4.68e-3)
DTLZ7	3	6	5.4108e-2 (4.88e-3) $\approx$	5.4427e-2 (3.86e-3)
$+/-/\approx$			2/0/1	

Discussion

Why does the approximation method outperform the exact formula?

- Exact formula relies on the specific technique to compute the multivariable Gaussian cumulative distribution function (cdf).
- Gaussian cdf: does not have a closed-form expression when $m \geq 3$.

Code

- https://github.com/mobo-d/MOEAD-EGO/ETI

```
for k=1:M
    L_k =  eye(M);    L_k(:,k)=-1;  L_k(k,k)= 1;
    mu_k = g_mu*L_k'; % The mean of Z^(k), n*m
    b_k = zeros(n,M); b_k(:,k) = Gbest;
    for j=1:1:n
        Sigma_k =L_k*diag(g_sig2(j,:))*L_k'; % The covariance matrix of Z^(k) for j-th query, m*m
        % P(Z^k<=b^k)
        p_k  =   qsimvnv(200*M,Sigma_k, -inf.*ones(M,1), (b_k(j,:)-mu_k(j,:))') ;
        ETI(j) =  ETI(j)+ p_k *(Gbest(j)-mu_k(j,k));

        temp_ci = (b_k(j,:) -mu_k(j,:))'./diag(Sigma_k);%m*1
        temp_Sigma_k_ii = sqrt(diag(Sigma_k));
        phi_ik =   Sigma_k(:,k).*normpdf((b_k(j,:) -mu_k(j,:))'./temp_Sigma_k_ii)./temp_Sigma_k_ii;%m*1
        for i=1:M
            % need c.i^(k) and Sigma.i^(k)
            cik =  b_k(j,:)'  - mu_k(j,:)' -Sigma_k(:,i).* temp_ci ; cik(i)=[]; % m-1*1
            sigmaik =  (Sigma_k - Sigma_k(:,i)*Sigma_k(i,:)./Sigma_k(i,i)) ;  sigmaik(i,:)=[]; sigmaik(:,i)=[]; % m-1*1
            Phi_ik =   qsimvnv(200*M,sigmaik ,-inf.*ones(M-1,1),cik ) ; %1*1
            ETI(j) =  ETI(j) +  phi_ik(i)*Phi_ik;
```

Outline

Conclusion and Future Work

Exact formulas VS Moment matching approximation

- Moment matching approximation: approximation error $\searrow$ as $\hat{\sigma}_i^2 \searrow$.
- Moment matching approximation is good enough in terms of both accuracy and computational time.

Future work

- Similarity between Multi-point EI and ETI:
  moment matching approximation -> multi-point EI

Thank you!

Liang Zhao (liazhao5-c@my.cityu.edu.hk)