Multi-objective Bayesian global optimization for continuous problems and applications

Yang, K.

Author: Yang, Kaifeng
Title: Multi-objective Bayesian global optimization for continuous problems and applications
Date: 2017-12-06

Appendix A

A.1 Symbols

Table A.1: Notations.

Symbol	Type	Description	Ref.
m	$\mathbb{N}^+$	Dimension of a search space	Eq. (2.1)
d	$\mathbb{N}^+$	Dimension of an objective space	Eq. (2.1)
$\mathbb{S}$	$\mathbb{R}^m$	Search space	Eq. (2.1)
$\mathcal{X}$	$\subseteq \mathbb{S}$	Feasible set in $\mathbb{S}$	Eq. (2.1)
$\mathbf{x}$	$\in \mathcal{X}$	Decision vector	Eq. (2.1)
$\mathbf{X}$	$\subset \mathcal{X}$	Decision vector set	Eq. (2.2)
$\mathbf{y}_i$	$\mathbb{R}$	The i-th objective function	Eq. (2.1)
$\mathbf{y}$	$\mathbb{R}^d$	The objective functions	Def. (2.1)
$y_i(\mathbf{x})$	$\mathbb{R}$	The i-th objective value of $\mathbf{x}$	Def. (2.1)
$\mathcal{P}^*$	$(\mathbb{R}^d)^n$	Pareto front	Def. (2.3)
$\mathcal{P}$	$(\mathbb{R}^d)^n$	Pareto front approximation	Def. (2.4)
n	$\mathbb{N}^+$	Number of the points in $\mathcal{P}$ or $\mathcal{P}^*$	Def. (2.5)
$\mathbf{r}$	$\mathbb{R}^d$	Reference point	Def. (2.5)
y_i	$\mathbb{N}^+$	The i-th objective space	Fig. 2.2
μ	$\mathbb{N}^+$	Population size	Alg. 1
p_m	$0 < p_m < 1$	Mutation rate	Alg. 1
p_c	$0 < p_c < 1$	Crossover rate	Alg. 1
R_i	$\mathbb{R}^+$	Branch resistance	Def. 4-8
f_{loss}	$\mathbb{R}^+$	Active power loss	Def. 4-8
P_i	$\mathbb{R}^+$	Active power	Def. 4-8
Q_i	$\mathbb{R}^+$	Inactive power	Def. 4-8
V_i	$\mathbb{R}^+$	Node voltage	Def. 4-8
I_i	$\mathbb{R}^+$	Branch current	Def. 4-8
f_{VDI}	$\mathbb{R}^+$	Voltage deviation	Def. 4-11
$\boldsymbol{\mu}$	$\mathbb{R}^d$	Mean values of predictive distribution	Alg. 3
$\boldsymbol{\sigma}$	$(\mathbb{R}_0^+)^d$	Standard deviations of predictive distribution	Alg. 3
$\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}$	$\mathbb{R}^d$	The vectors in $\mathcal{P}$, where $\mathcal{P} = (\mathbf{y}^{(\mathbf{1})}, \ldots, \mathbf{y}^{(\mathbf{n})})$	Fig. 2.2
S_d	$(\mathbb{R}^d)^2$	Integration slices for d dimension	Def. 3.3.1
$\mathbf{l}_d^{(1)}, \ldots, \mathbf{l}_d^{(N_d)}$	$\mathbb{R}^d$	Lower bound of integration boxes	Def. 3.3.1
$\mathbf{u}_d^{(1)}, \ldots, \mathbf{u}_d^{(N_d)}$	$\mathbb{R}^d$	Upper bound of integration boxes	Def. 3.3.1
N_d	$\mathbb{N}^+$	Number of integration boxes	Def. 3.3.1
$\mathbf{x}^*$	$\mathbb{R}^d$	An optimal point in search space	Alg. 3
D	$\mathbb{R}^{(m+d)}$	Training data set	Alg. 3

A.2 Abbreviations

Table A.2: Abbreviations-I.

Abb.	Full Name
AVL	Adelson-Velskii and Landis
BGO	Bayesian Global Optimization
CDF	Cumulative Density Function
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
DKLV	Dächert, Klamroth, Lacour and Vanderpooten
LKF	Lacour, Klamroth and Fonseca
DF(s)	Desirability Function(s)
DM	Decision Maker
DNRP	Distribution Network Reconfiguration Problem
EA(s)	Evolutionary Algorithm(s)
EGO	Efficient Global Optimization
EHVI	Expected Hypervolume Improvement
EHVIG	Expected Hypervolume Improvement Gradient
EI	Expected Improvement
EMOA(s)	Evolutionary Multi-objective Optimization Algorithm(s)
GA	Genetic Algorithm
GAA	Gradient Ascent Algorithm
GSP	Generalized Schaffer Problem
HV	Hypervolume Indicator
HVI	Hypervolume Improvement
HVC	Hypervolume Contribution
LCB	Lower Confidence Bound
MLI	Most Likely Improvement
MONMPC	Multi-Objective Nonlinear Model Predictive Control
MOO	Multi-objective Optimization
MOPSO	Multi-Objective Particle Swarm Optimization
NOEs	Number of Evaluations
NSGA-II	Non-dominated Sorting Genetic Algorithm II

Table A.3: Abbreviations-II.

Abb.	Full Name
OK	Ordinary Kriging
PCA	Principal Component Analysis
PDF	Probability Density function
PICEA-g	Preference-Inspired Co-Evolutionary Algorithm
PID	Proportional Integral Derivative
PMX	Partial-Mapped Crossover
PO	Percentage Overshoot
PoI	Probability of improvement
PR	Preferred Region
REN21	Renewable Energy Policy Network for the 21st Century
ROI	Region of Interest
SMS-EMOA	S-metric Selection Evolutionary Multi-Optimization Algorithm
SMS-EGO	S-metric Selection Efficient Global Optimization
TCDF	Truncated Cumulative Density Function
TEHVI	Truncated Expected Hypervolume Improvement
THV	Truncated Hypervolume
TPDF	Truncated Probability Density Function
λ	Lebesgue Measure

A.3 EHVIG Formula Derivation

1. $\phi'(x) = -x\phi(x)$ (A-1)

2. $\Phi'(x) = \phi(x)$ (A-2)

3. $\frac{\partial \Phi(\frac{y-\mu}{\sigma})}{\partial \mathbf{x}} = \phi(\frac{y-\mu}{\sigma}) \cdot (\frac{\mu - y}{\sigma^2} \cdot \frac{\partial \sigma}{\partial \mathbf{x}} - \frac{1}{\sigma} \cdot \frac{\partial \mu}{\partial \mathbf{x}})$ (A-3)

   Using the chain rule and quotient rule, considering that y does not depend on $\mathbf{x}$, we get the statement in (A-3):

$$\frac{\partial \Phi(\frac{y-\mu}{\sigma})}{\partial \mathbf{x}} = \phi(\frac{y-\mu}{\sigma}) \cdot \frac{\partial(\frac{y-\mu}{\sigma})}{\partial \mathbf{x}} = \phi(\frac{y-\mu}{\sigma}) \cdot \frac{(\frac{\partial y}{\partial \mathbf{x}} - \frac{\partial \mu}{\partial \mathbf{x}})\sigma - (y-\mu)\frac{\partial \sigma}{\partial \mathbf{x}}}{\sigma^2}$$

   After tidying up, we get a statement in (A-3):

$$\frac{\partial \Phi(\frac{y-\mu}{\sigma})}{\partial \mathbf{x}} = \phi(\frac{y-\mu}{\sigma}) \cdot (\frac{\mu - y}{\sigma^2} \cdot \frac{\partial \sigma}{\partial \mathbf{x}} - \frac{1}{\sigma} \cdot \frac{\partial \mu}{\partial \mathbf{x}})$$

4. $$\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mathbf{x}} = \big(\frac{b-a}{\sigma} \cdot \phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{b-\mu}{\sigma})\big) \cdot \frac{\partial \mu}{\partial \mathbf{x}} + \phi(\frac{b-\mu}{\sigma}) \cdot \big(1 + \frac{(b-\mu)(b-a)}{\sigma^2}\big) \cdot \frac{\partial \sigma}{\partial \mathbf{x}}$$

   Using the product rule and considering a and b do not depend on $\mathbf{x}$, we get the statement:

$$\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mathbf{x}} = \frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mu} \cdot \frac{\partial \mu}{\partial \mathbf{x}} + \frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial \mathbf{x}} \quad \text{(A-4)}$$

   Substituting Equation (A-7) into $\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mu}$ and $\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \sigma}$, using the chain rule, quotient rule, and product rule, the statements of $\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mu}$ and $\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \sigma}$ are:

$$\begin{aligned}
\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mu} &= \frac{\partial[\sigma \cdot \phi(\frac{b-\mu}{\sigma}) + (a-\mu)\cdot \Phi(\frac{b-\mu}{\sigma})]}{\partial \mu} \\
&= \sigma \cdot \frac{\partial \phi(\frac{b-\mu}{\sigma})}{\partial \mu} + (-1)\cdot \Phi(\frac{b-\mu}{\sigma}) + (a-\mu)\cdot \frac{\partial \Phi(\frac{b-\mu}{\sigma})}{\partial \mu} \\
&= \frac{b-\mu}{\sigma}\cdot \phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{b-\mu}{\sigma}) + [-\frac{a-\mu}{\sigma}\cdot \phi(\frac{b-\mu}{\sigma})] \\
&= \frac{b-a}{\sigma}\cdot \phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{b-\mu}{\sigma}) \qquad \text{(A-5)}
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \sigma} = \frac{\partial[\sigma \cdot \phi(\frac{b-\mu}{\sigma}) + (a-\mu)\cdot \Phi(\frac{b-\mu}{\sigma})]}{\partial \sigma} \\
&= \phi(\frac{b-\mu}{\sigma}) + \sigma \cdot \frac{\partial \phi(\frac{b-\mu}{\sigma})}{\partial \sigma} + (a-\mu)\cdot \frac{\partial \Phi(\frac{b-\mu}{\sigma})}{\partial \sigma} \\
&= \phi(\frac{b-\mu}{\sigma}) + (\frac{b-\mu}{\sigma})^2 \cdot \phi(\frac{b-\mu}{\sigma}) + \big(-\frac{(a-\mu)\cdot(b-\mu)}{\sigma^2}\cdot \phi(\frac{b-\mu}{\sigma})\big) \\
&= \phi(\frac{b-\mu}{\sigma}) + \frac{(b-\mu)\cdot(b-a)}{\sigma^2}\cdot \phi(\frac{b-\mu}{\sigma}) \\
&= \phi(\frac{b-\mu}{\sigma})\big(1 + \frac{(b-\mu)\cdot(b-a)}{\sigma^2}\big) \qquad \text{(A-6)}
\end{aligned}$$

After substituting Equations (A-5) and (A-6) into (A-4), then we get statement in (A-4):

$$\begin{aligned}
\frac{\partial \Psi(a,b,\mu,\sigma)}{\partial \mathbf{x}} &= \big(\frac{b-a}{\sigma}\cdot \phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{b-\mu}{\sigma})\big)\cdot \frac{\partial \mu}{\partial \mathbf{x}} + \\
&\phi(\frac{b-\mu}{\sigma})\cdot \big(1 + \frac{(b-\mu)(b-a)}{\sigma^2}\big)\cdot \frac{\partial \sigma}{\partial \mathbf{x}}
\end{aligned}$$

A.4 2-D EHVI Formula (Minimization Case)

Definition A.1 (Ψ function) *Let $\phi(s) = 1/\sqrt{2\pi}e^{-\frac{1}{2}s^2}, s \in \mathbb{R}$ denote the PDF of the standard normal distribution and $\Phi(s) = \frac{1}{2}\left(1 + erf\left(\frac{s}{\sqrt{2}}\right)\right)$ denote its cumulative probability distribution function (CDF). The general normal distribution with mean μ and variance σ has the PDF $\phi_{\mu,\sigma}(s) = \frac{1}{\sigma}\phi(\frac{s-\mu}{\sigma})$ and the CDF $\Phi_{\mu,\sigma}(s) = \Phi(\frac{s-\mu}{\sigma})$. Then the function Ψ is defined as:*

$$\begin{aligned}\Psi(a, b, \mu, \sigma) &= \int_{-\infty}^{b} (a - z)\frac{1}{\sigma}\phi(\frac{z - \mu}{\sigma})dz \\ &= \sigma\phi(\frac{b - \mu}{\sigma}) + (a - \mu)\Phi\left(\frac{b - \mu}{\sigma}\right).\end{aligned} \tag{A-7}$$

It partitions the integration domain into $n + 1$ disjoint rectangular stripes S_1, …, S_{n+1}, see Figure A.1 for an illustration. For this, we augment the set P by two points $\mathbf{y}^{(0)} = (r_1, -\infty)$ and $\mathbf{y}^{(n+1)} = (-\infty, r_2)$. The stripes are now defined by:

$$S_i = \left(\begin{pmatrix} y_1^{(i)} \\ -\infty \end{pmatrix}, \begin{pmatrix} y_1^{(i-1)} \\ y_2^{(i)} \end{pmatrix}\right), \quad i = 1, \ldots, n + 1. \tag{A-8}$$

Suppose Y are the sorted non-dominated vectors of the current Pareto front approximation $\mathcal{P}$. A formula will be derived that consists of $n + 1$ integrals, as indicated in Figure A.1. The HVI of a point $\mathbf{y} \in \mathbb{R}^2$ can be expressed by:

$$\mathrm{HVI}(\mathbf{y}, \mathrm{Y}, \mathbf{r}) = \sum_{i=1}^{n+1} \lambda_2[S_i \cap \Delta(\mathbf{y})]. \tag{A-9}$$

This gives rise to the compact integral for the original EHVI, $\mathbf{y} = (y_1, y_2)$:

$$\begin{aligned}&\mathrm{EHVI}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathrm{Y}, \mathbf{r}) = \\ &\int_{y_1=-\infty}^{\infty} \int_{y_2=-\infty}^{\infty} \sum_{i=1}^{n+1} \lambda_2[S_i \cap \Delta((y_1, y_2))] \cdot PDF_{\boldsymbol{\mu},\boldsymbol{\sigma}}(\mathbf{y})d\mathbf{y}\end{aligned} \tag{A-10}$$

It is observed that the intersection of S_i with $\Delta((y_1, y_2))$ is non-empty if and only if $\mathbf{y} = (y_1, y_2)$ dominates the upper right corner of S_i. In other words, if and only

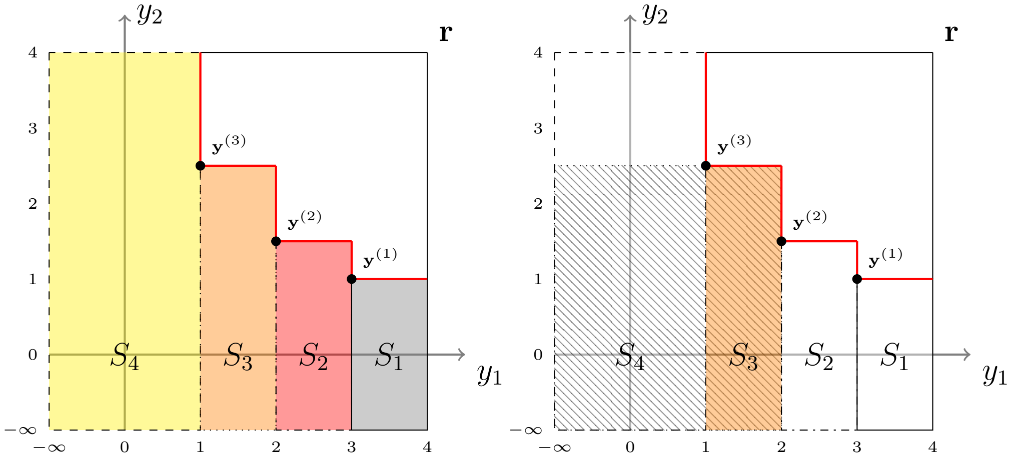

Figure A.1: Partitioning of the integration region into stripes. Right: New partitioning of the reduced integration region after first iteration of the algorithm.

if $\mathbf{y}$ is located in the rectangle with lower left corner $(y_1^{(i)}, -\infty)$ and upper right corner $(y_1^{(i-1)}, y_2^{(i)})$. Therefore:

$$\mathrm{EHVI}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathrm{Y}, \mathbf{r})$$

$$= \sum_{i=1}^{n+1} \int_{y_1=-\infty}^{y_1^{(i-1)}} \int_{y_2=-\infty}^{y_2^{(i)}} \lambda_2[S_i \cap \Delta((y_1, y_2))] \cdot PDF_{\boldsymbol{\mu},\boldsymbol{\sigma}}(\mathbf{y})d\mathbf{y} \tag{A-11}$$

$$= \sum_{i=1}^{n+1} \int_{y_1=-\infty}^{y_1^{(i)}} \int_{y_2=-\infty}^{y_2^{(i)}} \lambda_2[S_i \cap \Delta((y_1, y_2))] \cdot PDF_{\boldsymbol{\mu},\boldsymbol{\sigma}}(\mathbf{y})d\mathbf{y} +$$

$$\sum_{i=1}^{n+1} \int_{y_1=y^{(i)}}^{y_1^{(i-1)}} \int_{y_2=-\infty}^{y_2^{(i)}} \lambda_2[S_i \cap \Delta((y_1, y_2))] \cdot PDF_{\boldsymbol{\mu},\boldsymbol{\sigma}}(\mathbf{y})d\mathbf{y}$$

$$= \sum_{i=1}^{n+1} (y_1^{(i-1)} - y_1^{(i)}) \cdot \Phi\left(\frac{y_1^{(i)} - \mu_1}{\sigma_1}\right) \cdot \Psi(y_2^{(i)}, y_2^{(i)}, \mu_2, \sigma_2) +$$

$$\sum_{i=1}^{n+1} \left(\Psi(y_1^{(i-1)}, y_1^{(i-1)}, \mu_1, \sigma_1) - \Psi(y_1^{(i-1)}, y_1^{(i)}, \mu_1, \sigma_1)\right) \times$$

$$\Psi(y_2^{(i)}, y_2^{(i)}, \mu_2, \sigma_2) \tag{A-12}$$

Since integration is a linear mapping, it is allowed to do the summation after integration in Eq. (A-11). The integrals are now over a rectangular region and can be solved using the function Ψ as detailed in [2].