European Journal of Operational Research 211 (2011) 112–121

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

The cross-entropy method in multi-objective optimisation: An assessment

James Bekker [a,*], Chris Aldrich [b]

[a] Department of Industrial Engineering, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa
[b] Department of Process Engineering, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

ARTICLE INFO

Article history:
Received 24 September 2010
Accepted 21 October 2010
Available online 5 November 2010

Keywords:
Simulation
Cross-entropy
Stochastic processes
Multi-objective optimisation
Pareto-optimal

ABSTRACT

Solving multi-objective problems requires the evaluation of two or more conflicting objective functions, which often demands a high amount of computational power. This demand increases rapidly when estimating values for objective functions of dynamic, stochastic problems, since a number of observations are needed for each evaluation set, of which there could be many. Computer simulation applications of real-world optimisations often suffer due to this phenomenon. Evolutionary algorithms are often applied to multi-objective problems. In this article, the cross-entropy method is proposed as an alternative, since it has been proven to converge quickly in the case of single-objective optimisation problems. We adapted the basic cross-entropy method for multi-objective optimisation and applied the proposed algorithm to known test problems. This was followed by an application to a dynamic, stochastic problem where a computer simulation model provides the objective function set. The results show that acceptable results can be obtained while doing relatively few evaluations.

1. Introduction

In this article, we propose using the cross-entropy method (CEM) of Rubinstein and Kroese (2004) for multi-objective optimisation (MOO). It follows from the literature that the CEM is simple and generally converges quickly in single-objective problems. Our idea is to extend the CEM to problems with more than one objective, and apply the CEM to computationally expensive time-dependent, stochastic simulation problems where the optimisation of two or more objectives is pursued.

Multi-objective optimisation using evolutionary algorithms (MOEAs) has been actively researched over the past 25 years (see Coello et al., 2007 [p. 64]). Two major references among the many journal publications are the books by Deb (2001) and Coello et al. (2007). In a recent article, Coello (2009) highlighted current research trends and open topics in the field of MOEAs, which include discussion of alternative metaheuristics for solving MOO problems. It is also noted that there is much focus on designing MOEAs that reduce the number of objective function evaluations, because these evaluations can be very expensive when optimising some real-world problems.

Genetic algorithms (GAs) and other biologically inspired metaheuristics (*e.g.* ant colony and particle swarm optimisation) have been widely applied in solving MOO problems. Arguably the best-known evolutionary-based algorithms are the Multi-objective genetic algorithm (MOGA) of Fonseca and Fleming (1993), the Non-dominated sorting genetic algorithm (NSGA-II) of Deb et al. (2002), the Niched-Pareto Genetic Algorithm (NPGA) of Erickson et al. (1999), the Pareto Archived Evolution Strategy (PAES) of Knowles and Corne (2000), the Strength Pareto Evolutionary Algorithm (SPEA) of Zitzler and Thiele (1999), the Multi-objective Messy Genetic Algorithm (MOMGA) of Van Veldhuizen and Lamont (2000), and the Pareto Envelope-based Selection Algorithm (PESA) of Corne et al. (2000). These algorithms and some of their variants are discussed in Coello et al. (2007). The Adaptive Range Multi-objective Genetic Algorithm (ARMOGA) of Sasaki and Obayashi (2005) requires relatively few objective evaluations to find the Pareto front and has been applied in optimisation problems where computationally intensive objective evaluations are needed, for example in transonic wing design (Oyama et al., 2001).

The MOO problem, in general, is a problem of the type:

Minimise

$$f(\mathbf{x}) := [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})], \tag{1}$$

$$g_i(\mathbf{x}) \leqslant 0, \quad i = 1, 2, \ldots, p, \tag{2}$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, q, \tag{3}$$

where the vector of decision variables is denoted by $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, m$, are the objective functions, while the constraint functions are g_i, $h_j : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$ (Coello, 2009).

Since MOO problems usually have at least two conflicting objectives, there exist many acceptable solutions for a given problem. These form the *Pareto optimal set*. A few definitions pertaining to

* Corresponding author. Tel.: +27 21 8084951; fax: +27 21 8084245.
E-mail address: jb2@sun.ac.za (J. Bekker).

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.10.028

Pareto optimality are necessary, and the basic definitions in Coello (2009) are repeated here for convenience:

Definition 1. Given two vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^m$, we say that $\mathbf{u} \leqslant \mathbf{v}$ if $u_i \leqslant v_i$ for $i = 1,2,\ldots,m$, and that $\mathbf{u} < \mathbf{v}$ if $\mathbf{u} \leqslant \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$.

Definition 2. Given two vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^m$, we say that $\mathbf{u}$ dominates $\mathbf{v}$ (denoted by $\mathbf{u} \prec \mathbf{v}$) iff $\mathbf{u} < \mathbf{v}$.

Definition 3. A vector of decision variables $\mathbf{x}^* \in \mathcal{F}$ ($\mathcal{F}$ is the feasible region) is *Pareto optimal* if there does not exist another $\mathbf{x} \in \mathcal{F}$ such that $\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}^*)$.

Definition 4. The *Pareto optimal set* $\mathcal{P}^*$ is defined by $\mathcal{P}^* = \{\mathbf{x} \in \mathcal{F} | \mathbf{x}$ is Pareto optimal$\}$.

Definition 5. The *Pareto front* $\mathcal{P}_T^*$ is defined by $\mathcal{P}_T^* = \{\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n | \mathbf{x} \in \mathcal{P}^*\}$.

Solving an MOO problem requires that the Pareto optimal set be found from the set of all decision variable vectors that satisfy Eqs. (2) and (3).

Multi-objective problems and solutions are widely reported in the literature, for example in inventory management (Tsou, 2008; Tsou, 2009; Mahapatra and Maiti, 2005), while Baesler and Sepúlveda (2001) improved the design of a cancer treatment centre based on four objectives. Li et al. (2009) used a multi-optimisation method in an environmentally conscious design of chemical processes and products. Kleijnen and Wan (2007) studied optimisation of simulated systems by comparing some optimisation methods. These include a brute-force approach, modified response surface methodology (RSM), perturbation analysis (PA) and feasible directions (FD).

Evolutionary Algorithms (EA) are widely used in MOO research and applications. Beausoleil (2006) applies a multiple-objective scatter search (MOSS) to test problems from the literature, and Deb et al. (2002) improve on existing algorithms with their NSGA-II. Coello et al. (2004) apply particle swarm optimisation while incorporating Pareto dominance. Summanwar et al. (2002) solve constrained optimisation problems using multi-objective genetic algorithms, while Zitzler and Thiele (1999) apply the SPEA to the 0/1 knapsack problem. Gil et al. (2007) developed a hybrid method for solving MOO problems by combining *e.g.* PESA and NSGA-II. In other applications, specific methods are developed to solve MOO problems, *e.g.* Lee (2007) developed a trajectory-informed search methodology and applies it to several test problems. Chapter 7 in Coello et al. (2007) presents a comprehensive reference of applications in engineering, science, industry and miscellaneous fields (*e.g.* investment portfolio optimisation and stock ranking). A comprehensive list of references is maintained at the EMOO home page (http://www.lania.mx/~ccoello/).

The cross-entropy method is a relatively recent development by Reuven Rubinstein. He originally developed it for use in the field of importance sampling (Rubinstein, 1997), but it has been extended to many other types of problems, including continuous optimisation and combinatorial optimisation. Specific examples are solving problems like the Rosenbrock function (Rubinstein and Kroese, 2004, p. 89, the max-cut problem (p. 140), the Travelling Salesman Problem (p. 147) and the capacitated vehicle routing problem (p. 238).

We next discuss the cross-entropy method, followed by the proposed algorithm for multi-objective optimisation using the CEM. We show metrics resulting from its application to various deterministic benchmark (test) problems, and finally its application to a simulated, stochastic problem.

2. Optimisation and the cross-entropy method

The CEM for optimisation is briefly outlined in this section, and for detail the reader is referred to Rubinstein and Kroese (2004), the CEM website (http://www.cemethod.org) and Kroese and Rubinstein (2005), the latter being a complete journal issue devoted to the cross-entropy method. The CEM for optimisation has its foundation in *Importance Sampling* and the *Kullback–Leibler distance*, and these are discussed first (Rubinstein and Kroese, 2004).

Let $\mathbf{X} = (X_1,\ldots,X_n)$ be a random vector assuming values from some space $\mathcal{X}$, and let f be some real function on $\mathcal{X}$. Suppose we want to determine the probability that $f(\mathbf{X})$ is greater or equal than a real number γ under a family of probability density functions $h(\cdot;\mathbf{u})$ on $\mathcal{X}$. This probability is

$$l = \mathbb{P}_{\mathbf{u}}(f(\mathbf{X}) \geqslant \gamma) = \mathbb{E}_{\mathbf{u}} I_{\{f(\mathbf{X}) \geqslant \gamma\}}. \tag{4}$$

$f(\mathbf{X} \geqslant \gamma)$ is called a *rare event* if l is very small, and it can be efficiently estimated using importance sampling. To do so, we take a random sample $\mathbf{X}_1,\ldots,\mathbf{X}_N$ from a different density g on $\mathcal{X}$, and estimate l using the likelihood ratio estimator (Rubinstein and Kroese, 2004):

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I_{\{f(\mathbf{X}_i) \geqslant \gamma\}} \frac{h(\mathbf{X}_i;\mathbf{u})}{g(\mathbf{X}_i)}. \tag{5}$$

Now use the change of measure with density

$$g^*(\mathbf{x}) = \frac{I_{\{f(\mathbf{x}) \geqslant \gamma\}} h(\mathbf{x};\mathbf{u})}{l}, \tag{6}$$

then

$$l = \frac{I_{\{f(\mathbf{X}_i) \geqslant \gamma\}} h(\mathbf{X}_i;\mathbf{u})}{g^*(\mathbf{X}_i)}. \tag{7}$$

The value of g^* depends on the unknown l, but g^* can be approximated within the family of densities $\{h(\cdot;\mathbf{v})\}$ with reference parameter $\mathbf{v}$ such that the distance between g^* and $h(\cdot;\mathbf{v})$ is minimal. A measure of this distance is the *Kullback–Leibler distance* or cross-entropy between g and h, and it is defined as:

$$\mathcal{D}(g,h) = \mathbb{E}_g \ln \frac{g(\mathbf{X})}{h(\mathbf{X})} \tag{8}$$

$$= \int g(\mathbf{x}) \ln g(\mathbf{x}) d\mathbf{x} - \int g(\mathbf{x}) \ln h(\mathbf{x}) d\mathbf{x}. \tag{9}$$

To minimise the Kullback–Leibler distance between g^* in Eq. (6) and $h(\cdot;\mathbf{v})$, $\mathbf{v}$ is chosen such that $-\int g^*(\mathbf{x}) \ln h(\mathbf{x};\mathbf{v}) d\mathbf{x}$ is minimised. This can be achieved by solving the maximisation problem

$$\max_{\mathbf{v}} \int g^*(\mathbf{x}) \ln h(\mathbf{x};\mathbf{v}) d\mathbf{x}. \tag{10}$$

When g^* of Eq. (6) is substituted in Eq. (10), the maximisation program

$$\max_{\mathbf{v}} \int \frac{I_{\{f(\mathbf{x}) \geqslant \gamma\}} h(\mathbf{x};\mathbf{u})}{l} \ln h(\mathbf{x};\mathbf{v}) d\mathbf{x} \tag{11}$$

is obtained, which is equivalent to the program

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathbb{E}_{\mathbf{u}} I_{\{f(\mathbf{X}) \geqslant \gamma\}} \ln h(\mathbf{X};\mathbf{v}). \tag{12}$$

With reference to the above, the CEM for optimisation can now be stated. Suppose we wish to find the maximum of some performance function $f(\mathbf{x})$ over all states $\mathbf{x}$ in some set $\mathcal{X}$. Let the maximum be γ^*, then

$$\gamma^* = \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}). \tag{13}$$

The deterministic problem is randomised by defining a family of probability density functions $\{h(\cdot;\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ on the set $\mathcal{X}$. The *associated stochastic problem* of Eq. (13) is the estimation problem