Expert Systems With Applications 68 (2017) 123–135

Contents lists available at ScienceDirect

Expert Systems With Applications

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

# Multiobjective Markov chains optimization problem with strong Pareto frontier: Principles of decision making

Julio B. Clempner*, Alexander S. Poznyak

[a] Centro de Investigaciones Economicas, Administrativas y Sociales, Instituto Politecnico Nacional, Lauro Aguirre 120, col. Agricultura, Del. Miguel Hidalgo, 11360, Ciudad de Mexico, Mexico
[b] Department of Control Automatic, Center for Research and Advanced Studies, Av. IPN 2508, Col. San Pedro Zacatenco, 07360 Mexico City, Mexico

ARTICLE INFO

*Article history:*
Received 16 February 2016
Revised 12 October 2016
Accepted 13 October 2016
Available online 13 October 2016

*Keywords:*
Multi-objective
Decision making
Pareto
Tikhonov's regularization
Markov chains

ABSTRACT

In this paper, we present a novel approach for computing the Pareto frontier in Multi-Objective Markov Chains Problems (MOMCPs) that integrates a regularized penalty method for poly-linear functions. In addition, we present a method that make the Pareto frontier more useful as decision support system: it selects the ideal multi-objective option given certain bounds. We restrict our problem to a class of finite, ergodic and controllable Markov chains. The regularized penalty approach is based on the Tikhonov's regularization method and it employs a projection-gradient approach to find the strong Pareto policies along the Pareto frontier. Different from previous regularized methods, where the regularizator parameter needs to be large enough and modify (some times significantly) the initial functional, our approach balanced the value of the functional using a penalization term ($\mu$) and the regularizator parameter ($\delta$) at the same time improving the computation of the strong Pareto policies. The idea is to optimize the parameters $\mu$ and $\delta$ such that the functional conserves the original shape. We set the initial value and then decrease it until each policy approximate to the strong Pareto policy. In this sense, we define exactly how the parameters $\mu$ and $\delta$ tend to zero and we prove the convergence of the gradient regularized penalty algorithm. On the other hand, our policy-gradient multi-objective algorithms exploit a gradient-based approach so that the corresponding image in the objective space gets a Pareto frontier of just strong Pareto policies. We experimentally validate the method presenting a numerical example of a real alternative solution of the vehicle routing planning problem to increase security in transportation of cash and valuables. The decision-making process explored in this work correspond to the most frequent computational intelligent models applied in practice within the *Artificial Intelligence* research area.



## 1. Introduction

### 1.1. Brief review

Multi-criterion optimization is a well-established research area with an extensive collection of solution concepts and methods presented in the literature (for an overview see, for example, Collette & Siarry, 2004; Dutta & Kaya, 2011; Ehrgott & Gandibleux, 2002; Eichfelder, 2008; Mueller-Gritschneder, Graeb, & Schlichtmann, 2009; Roijers, Vamplew, Whiteson, & Dazeley, 2013; Zitzler, Deb, & Thiele, 2000). The most common notion of multi-criterion optimization is that of efficient (or Pareto optimal) solutions, namely those solutions that cannot be improved upon in all coordinates (with strict improvement is at least one coordinate) by another solution. Regrettably efficient solutions are non-unique. Then, different methods have been proposed to identify one optimal solution, such that of computing the scalar combinations of the objective functions. Alternatively, an optimization method that leads to a well-defined solution is the constrained optimization problem, where one criterion is optimized subject to explicit constraints on the others. In practice decision makers often find it hard to specify such preferences, and would prefer a representation method based-on an intelligent system able to exemplify the range of such possible alternatives. Many researchers utilized versatile computational intelligent models such as Pareto based techniques for handling this types of problems (Angelov, Filev, & Kasabov, 2010).

In the context of Markov decision processes (MDP) serious works have been developed to compute efficient solutions. Durinovic, Lee, Katehakis, and Filar (1986) presented a MOMDP based on the average reward function to characterize the complete sets of efficient policies, efficient deterministic policies and

* Corresponding author.
*E-mail addresses:* julio@clempner.name (J.B. Clempner), apoznyak@ctrl.cinvestav.mx (A.S. Poznyak).

http://dx.doi.org/10.1016/j.eswa.2016.10.027
0957-4174/

efficient points using linear programming. Beltrami, Katehakis, and Durinovic (1995) considered a MOMDP formulation for deploying emergency services considering the area average response time and the steady-state deterioration in the ability of each district to handle future alarms originating in its own region. Wakuta and Togawa (1998) suggested a policy iteration algorithms for finding all optimal deterministic stationary policies concerned with MOMDP with no discounting as well as discounting. Chatterjee, Majumdar, and Henzinger (2006) considered MDPs with multiple discounted reward objectives showing that the Pareto curve can be approximated in polynomial time in the size of the MDP. Etessami, Kwiatkowska, and Yannakakis (2007) studied efficient algorithms for multi-objective model checking problems which runs in time polynomial in the size of the MDP. Clempner and Poznyak (2016c) provided a method based on minimizing the Euclidean distance is proposed for generating a well-distributed Pareto set in multi-objective optimization for a class of ergodic controllable Markov chains. Clempner (2016) proposed to follow the Karush–Kuhn–Tucker (KKT) optimization approach where the optimality necessary and sufficient conditions are elicited naturally for a Pareto optimal solution in MOMDP. Clempner and Poznyak (2016a) presented a multi-objective solution of the Pareto front for Markov chains transforming the original multi-objective problem into an equivalent nonlinear programming problem implementing the Lagrange principle. Clempner and Poznyak (2016d) suggested a novel method for computing the multi-objective problem in the case of a metric state space using the Manhattan distance. Learning schemes for constrained MCs were presented by Poznyak, Najim, and Gomez-Ramirez (2000), based on the theory of stochastic learning automata. Pirotta, Parisi, and Restelli (2015) provided an algorithm to exploit a gradient–based approach to optimize the parameters of a function that defines a manifold in the policy parameter space so that the corresponding image in the objective space gets as close as possible to the Pareto frontier. As well as, Vamplew, Dazeley, Barker, and Kelarev (2009) studied the benefits of employing stochastic policies to MO tasks and examined a particular form of stochastic policy known as a mixture policy proposing two different methods.

### 1.2. Motivation

Let us consider a general vector optimization problem as

$$\begin{array}{l} \min f(x) \text{ (with respect to } \mathcal{K}) \\ \qquad \text{subject to} \\ g_i(x) \le 0, \; i = 1, \dots, m \\ h_i(x) = 0, \; i = 1, \dots, p \end{array} \tag{1}$$

where $x \in \mathbb{R}^N$ is the optimization variable, $\mathcal{K} \subseteq \mathbb{R}^q$ is a proper cone, $f(x) : \mathbb{R}^N \to \mathbb{R}^q$ is the objective function, $g_i(x) : \mathbb{R}^N \to \mathbb{R}$ are the inequality constraint functions, and $h_i(x) : \mathbb{R}^N \to \mathbb{R}$ are the equality constraint functions. Here the proper cone $\mathcal{K}$ is used to compare the objective values. The optimization problem presented in Eq. (1) is a convex optimization problem if: i) the objective function $f(x)$ is $\mathcal{K}$-convex, ii) the inequality constraint functions $g_i(x)$ are convex, and iii) the equality constraint functions $h_i(x)$ are affine.

Let us consider the set of achievable objective values of feasible points,

$$\mathcal{O} = \left\{ f(x) \middle| \exists x \in x \in \mathbb{R}^N, g_i(x) \le 0, i = 1, \dots, m, \right. \\ \left. h_i(x) = 0, i = 1, \dots, p \right\} \subseteq \mathbb{R}^q$$

The values $f(x_1) \preccurlyeq_{\mathcal{K}} f(x_2)$ are to be compared using the inequality $\preccurlyeq_{\mathcal{K}}$ meaning that $x_1$ is 'better than or equal' in value to $x_2$. In this sense, if the set $\mathcal{O}$ has a minimum (optimal point of the problem given in Eq. (1)) then there is a feasible point $x^*$ such that $f(x^*) \preccurlyeq_{\mathcal{K}} f(x_i)$ for all feasible $x_i$. When an optimization problem has an optimal point it is unique: if $x^*$ is an optimal point then $f(x^*)$ can be compared to the objective at every other feasible point, and is better than or equal to it.

A point $x^*$ is optimal if and only if it is feasible and

$$\mathcal{O} \subseteq \{f(x^*)\} \oplus \mathcal{K}$$

where $\{f(x^*)\} \oplus \mathcal{K}$ is the set of values that are worse than, or equal to, $f(x^*)$.

We say that a feasible point $x$ is Pareto optimal (or efficient) if $f(x)$ is a minimal element of the set of achievable values $\mathcal{O}$. In this case we say that $f(x)$ is a Pareto optimal point for the optimization problem given in Eq. (1). Thus, a point $x^*$ is Pareto optimal if it is feasible and, for any feasible $x$, $f(x) \preccurlyeq_{\mathcal{K}} f(x^*)$ implies $f(x) = f(x^*)$. A point $x$ is Pareto optimal if and only if it is feasible and

$$(\{f(x^*)\} \ominus \mathcal{K}) \cap \mathcal{O} = \{f(x^*)\} \tag{2}$$

where $\{f(x^*)\} \ominus \mathcal{K}$ is the set of points that are better than or equal to $f(x^*)$, i.e., the achievable value better than or equal to $f(x^*)$ is $f(x^*)$ itself.

An optimization problem can have many Pareto optimal points. Every Pareto optimal point is an achievable objective value that lies in the boundary of the set of feasible points. The set of Pareto optimal points $\mathcal{P}$ satisfies

$$\mathcal{P} \subseteq \mathcal{O} \cap \mathcal{B}(\mathcal{O})$$

where $\mathcal{B}(\mathcal{O})$ denotes the boundary of the set of feasible points.

Let us consider the scalar optimization problem

$$\begin{array}{l} \min \lambda^\intercal f(x) \text{ (with respect to } \mathcal{K}) \\ \qquad \text{subject to} \\ g_i(x) \le 0, \; i = 1, \dots, m \\ h_i(x) = 0, \; i = 1, \dots, p \\ \lambda \in [0, 1], \; \sum_l^q \lambda_l = 1 \end{array} \tag{3}$$

The weight vector $\lambda$ is a free parameter that allows to obtain (possibly) different Pareto optimal solutions of the optimization problem (1). Geometrically, this means that a point $x$ is optimal for the scalarized problem if it minimizes $\lambda^\intercal f(x)$ over the feasible set, if and only if $\lambda^\intercal\{f(x_i) - f(x)\} \ge 0$ for any feasible $x_i$.

If the optimization problem given in Eq. (1) is convex, then the scalarized problem given in Eq. (3) is also convex, since $\lambda^\intercal f(x)$ is a scalar-valued convex function. This means that we can find Pareto optimal points of a convex optimization problem by solving a convex scalar optimization problem.

Note that this problem may have non-unique solution. Tikhonov's regularization (Tikhonov, Goncharsky, Stepanov, & A.G., 1995; Tikhonov & Arsenin, 1977) is one of the most popular approaches to solve discrete ill-posed of the minimization problem

$$f(x) = x_{j_1}^{(1)} A^l x_{j_2}^{(2)}. \tag{4}$$

The method looks for establishing an approximation of $x$ by replacing the minimization problem (4) by a penalized problem of the form

$$f_\delta(x) = \sum_{j_1, j_2} a_{j_1 j_2} x_{j_1}^{(1)} x_{j_2}^{(2)} \\ + \frac{\delta}{2} \left( \left\| x_{j_1}^{(1)} \right\|^2 + \left\| x_{j_2}^{(2)} \right\|^2 \right)$$

with a regularization parameter $\delta > 0$. The term $\frac{\delta}{2} \left( \left\| x_{j_1}^{(1)} \right\|^2 + \left\| x_{j_2}^{(2)} \right\|^2 \right)$ penalizes large values of $x$, and result in a sensible solution in cases when minimizing the first term only does not. The parameter $\delta$ is computed to obtain the right balance making both, the original objective function $x_{j_1}^{(1)} A^l x_{j_2}^{(2)}$ and the term $\frac{\delta}{2} \left( \left\| x_{j_1}^{(1)} \right\|^2 + \left\| x_{j_2}^{(2)} \right\|^2 \right)$ small. We can described the regularization problem as an optimization problem

$$\min_x \left( x_{j_1}^{(1)} A^l x_{j_2}^{(2)}, \delta \left\| x_{j_1}^{(1)} \right\|^2 + \left\| x_{j_2}^{(2)} \right\|^2 \right). \tag{5}$$