Computers and Chemical Engineering 88 (2016) 50–58

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Linear programming-based scenario reduction using transportation distance

Zhuangzhi Li, Zukui Li*

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G2V4, Canada

ARTICLE INFO

Article history:
Received 29 November 2015
Received in revised form 2 February 2016
Accepted 6 February 2016
Available online 13 February 2016

Keywords:
Scenario reduction
Transportation distance
Linear programming

ABSTRACT

One of the major difficulties for scenario-based decision-making problems (e.g. stochastic programming using scenarios) is that the problem complexity quickly increases as the number of scenarios increases. Scenario reduction aims at selecting a small number of scenarios to represent a large set of scenarios for decision making, so as to significantly reduce the computational complexity while preserving the solution quality of using a large number of scenarios. In this work, a new computationally efficient scenario reduction algorithm is proposed based on transportation distance minimization. The proposed algorithm relies on solving linear programming problems. The scenario subset updating step and the probability value assignment step are performed in an iterative manner until the transportation distance converges. Comparison with existing scenario reduction methods reveals that the proposed method is very efficient for the reduction of large scenario set. Application studies on stochastic optimization problems also demonstrate the effectiveness of the proposed method.

1. Introduction

Uncertainty is pervasive in various decision-making problems and there is a need for making optimal and reliable decisions in the presence of uncertainty. While uncertainty can be incorporated into the decision-making problem, consideration of the exact probability distribution in the optimization model often leads to nonlinearity and computational difficulty of numerical integration. As an alternative, scenario representation is widely used in solving optimization problem under uncertainty because of its advantage in modeling.

In reality, a huge number of scenarios may need to be considered due to the large number of uncertain parameters, which may result in numerically intractable problems because of the limitation of the computational resources. Therefore, it is necessary to find a subset of scenarios which can best approximate the original large number of scenarios. This induces the important topic of scenario-based decision making, scenario reduction. Generally, scenario reduction aims at selecting a few representative scenarios among the original large number of scenarios, and new probabilities will be assigned to each selected scenario.

Although it is an important topic, scenario reduction has received limited attention in the past (Dupačová et al., 2003; Heitsch et al., 2006; Karuppiah et al., 2010). Among the existing methods, the transportation distance (Rachev and Rüschendorf, 1998)-based scenario reduction was shown to be one of the most effective methods. A heuristic-based scenario reduction method for stochastic programming was proposed in Heitsch and Römisch (2003) and a corresponding tool SCENRED2 is available in GAMS. Li and Floudas (2014) proposed a mixed integer linear optimization-(MILP) based scenario reduction method, which rigorously minimizes the transportation distance (i.e. Kantorovich distance) to find the optimal subset to represent the initial superset of scenarios. This method also considers the performance of input and output space simultaneously. Although the MILP-based scenario method can provide the optimal reduced subset of scenarios, the method is limited by the size of the problem. With the state-of-the-art mixed integer linear optimization solver, the method can only address moderate number of scenarios (i.e. up to 5000). Li and Floudas (2015) recently proposed a method for the reduction of the huge number of scenarios (e.g. 5^{30}) generated from the factorial combination. A sequential reduction framework was proposed which significantly reduce the computational complexity. Some criteria for quantifying the quality of scenario reduction were also proposed considering that it is impractical to evaluate huge number of scenarios.

In this work, a linear programming (LP)-based scenario reduction method is proposed. The objective is to address the reduction from a large set of scenarios (much more than 5000, and not necessarily generated from the factorial combination). In the proposed

* Tel.: +1 780 492 1107; fax: +1 780 492 2881.
E-mail address: zukui@ualberta.ca (Z. Li).

http://dx.doi.org/10.1016/j.compchemeng.2016.02.005
0098-1354/

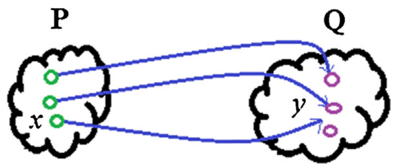

Fig. 1. Transportation problem.

algorithm, the selected scenarios and the corresponding probabilities are updated in two successive steps in an iterative fashion. The first step calculates the probabilities based on a given set of selected scenarios and the second step updates the selection of scenarios. Furthermore, the first step is solved based on a simple linear programming problem and the second step only relies on simple cost calculations. Hence, the computational complexity of the proposed method is much less than the existing MILP-based method. The proposed method works well for scenario set with a large size. Various case studies were investigated to illustrate the proposed algorithm. The effectiveness of the proposed method is also demonstrated through comparison studies with the MILP-based scenario reduction method (Li and Floudas, 2014) and the state-of-the art scenario reduction tool SCENRED2 (Heitsch and Römisch, 2003).

The rest of the paper is organized as follows. Section 2 introduces scenario reduction using transportation distance, and the MILP-based scenario reduction method is reviewed for comparison with the proposed algorithm. A linear programming-based scenario reduction algorithm is presented and illustrated in Section 3. The comparison studies are shown in Section 4. Section 5 studies an application in chance constrained portfolio optimization and the paper is concluded in Section 6.

2. Scenario reduction using mixed integer linear optimization

The scenario reduction problem investigated in this paper can be stated as following: Given a super set of scenarios/samples I, each scenario is associated with a probability (not necessarily equal, but their summation equals 1), remove N (a user-specified number) scenarios from I and keep the rest scenarios as subset S, update the probabilities of the preserved scenarios (so that their summation is 1), so as to minimize certain probability distance metric between the distribution of the superset I and the distribution of subset S.

Transportation distance is used to quantify the cost of the mass movement from one location to another. This is related to the Kantorovich transportation problem (Kantorovich, 1942) as shown in Fig. 1: if P and Q are two distributions of mass and if $c(x, y)$ represents the cost of transporting a unit of mass from the location x to the location y, what is the minimal total transportation cost to transfer P to Q? Transportation distance corresponding to the cost function $c(x, y)$ is defined as the minimal total transportation cost. The above Kantorovich transportation distance is used in this work to quantify the difference between I and S. If the difference between them is less, the corresponding transportation distance will also be smaller. In this paper, a scenario reduction is deemed as "optimal" when the Transportation distance between the super set of scenarios I and the selected subset S (with new probabilities) are minimum.

Li and Floudas (2014) proposed a mixed integer linear optimization formulation-based scenario reduction method based on transportation distance minimization. The method was designed to address uncertainty in optimization problem, it not only considers the input space (i.e. the values of uncertain parameters) but also consider the output space (i.e. the objective value of optimization problem). As a result, it leads to a reduced distribution not only close to the original distribution in terms of the distance in input space, but also captures the performance of the output. A simplified version of the mixed integer programming-based scenario reduction model (Li and Floudas, 2014) is presented as following

$$\min_{y_i, v_{i,i'}, d_i, p_{i'}^{new}} \sum p_i^{orig} d_i \tag{1a}$$

$$s.t. \quad \sum_{i \in I} y_i = N \tag{1b}$$

$$\sum_{i' \in I} v_{i,i'} = y_i, \forall i \in I \tag{1c}$$

$$0 \le v_{i,i'} \le 1 - y_{i'}, \forall i, i' \in I \tag{1d}$$

$$d_i = \sum_{i' \in I} c_{i,i'} v_{i,i'}, \forall i \in I \tag{1e}$$

$$0 \le d_i \le y_i c_{\max}, \forall i \in I \tag{1f}$$

$$y_i \in \{0, 1\}, \forall i \in I$$

In the above model, i and i' represent scenarios, I is the superset of all scenarios, p_i^{orig} represent probability of scenario i in the original discrete distribution (it is not necessarily equal probability $1/|I|$), d_i represent the cost of removing a scenario i and transporting it to other preserved scenarios, binary variables y_i denote whether a scenario is removed ($y_i = 1$) or preserved ($y_i = 0$), continuous variables $v_{i,i'}$ denote the probability mass fraction of scenario i that is transported to scenario i', continuous variables $p_{i'}^{new}$ denote the new probabilities of the scenarios ($p_{i'}^{new} = 0$ means scenario i' is removed), $c_{i,i'}$ is the transportation cost between two scenarios which can be modeled using distance (e.g. Manhattan distance $c_{i,i'} = \sum_{t=1}^{T} \left|\theta_t^i - \theta_t^{i'}\right|$, Euclidean distance $c_{i,i'} = \sqrt{\sum_{t=1}^{T} (\theta_t^i - \theta_t^{i'})^2}$, or other alternatives); θ^i is a realization of uncertain parameters in scenario i, $\theta^i = \{\theta_1^i, \theta_2^i, \ldots, \theta_T^i\}$, $c_{\max}$ is the maximum distance between two scenarios. According to Dupačová et al. (2003), the new probability of a selected scenario can be evaluated with the following equation

$$p_{i'}^{new} = (1 - y_{i'}) p_{i'}^{orig} + \sum_{i \in I} v_{i,i'} p_i^{orig}, \forall i' \in I \tag{2}$$

which means that new probability of a selected scenario is equal to the sum of its former probability and of all probabilities of removed scenarios that are transported to it. As a major limitation of the above MILP-based scenario reduction model, it suffers from combinatorial difficulty for relatively large problems. For example, for problem with more than 5000 scenarios, the MILP problem tends to be intractable for solving with the state-of-the-art solvers.

3. Linear programming-based scenario reduction

To address the computational issue of solving a MILP problem, a new method for scenario reduction is proposed in this work. The proposed method consists of two major steps that work iteratively: the first step calculates the probability assignment for a given set of selected scenarios; the second step updates the selected scenarios by evaluating the transportation cost. The algorithm stops once the transportation cost converges. Detailed algorithm for each step and the overall workflow is described below.

3.1. Probability mass transportation

If a subset of scenarios S is known as shown in Fig. 2, then the optimal probability values of the selected scenarios that minimize