Computers & Industrial Engineering 113 (2017) 1–9

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

An experimental approach for examining solution errors of engineering problems with uncertain parameters

Shangyao Yan [a], James C. Chu [b,*], Sin-Siang Wang [a]

[a] Department of Civil Engineering, National Central University, No. 300, Jhongda Rd., Jhongli, Taoyuan 320, Taiwan
[b] Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan

ARTICLE INFO

Article history:
Received 8 May 2017
Received in revised form 23 August 2017
Accepted 5 September 2017
Available online 7 September 2017

Keywords:
Engineering optimization models
Uncertain parameters
Error analysis
Optimality tolerance
Project scheduling

ABSTRACT

One potential overlook for applying optimization models to solve engineering problems is that their parameters are rarely error-free, implying that their solutions usually contain errors even when the models are solved to optimality. If the deviation between the solution based on parameters containing errors and the true optimal (but unavailable) solution based on error-free parameters is significant, the following decision-making could be meaningless. In this study, an experimental method is developed to evaluate solution errors of optimization models in which uncertain parameters are included in objective functions. A project scheduling problem is used as the case study. The effect of parameter errors and optimality tolerances in solution algorithms on solution errors are studied. The case study shows that the model solution errors increase as the scale of problem increases for the same range of parameter errors. It also shows that the model solution errors are similar for an optimality tolerance of within 4%. Regression models are estimated, which are useful for estimating potential errors between a solution based on parameters containing errors and the true optimal solution before a model is actually solved. They can also be used to determine values of optimality tolerance in solution algorithms that achieve the balance between solution quality and time.

1. Introduction

Engineering optimization models are formulated and solved to help engineers to make decisions for various engineering problems such as project scheduling, transportation scheduling, and manufacturing scheduling problems (Deng, Li, & Yang, 2011). A potential overlook for applying these optimization models is that if their parameters are not error-free, their solutions would also contain errors, even when they are solved to optimality. Since all kinds of uncertainties exist in engineering problems, it is unrealistic to expect these models to generate solutions that lead to optimal actions for real-world problems (Yu, Xu, & Tang, 2016).

Recognizing that errors might exist in model parameters, many engineering optimization problems are proposed to consider errors in parameter values. The related methodologies that have been adopted in these studies include data mining, grey prediction, fuzzy sets, artificial neural networks, stochastic programming, and robust optimization. The examples are Namk and Schaefer (1995), Teodorovic (1999), Hsu and Wen (2000), Lin and Yao (2001), Xue and Norrie (2001), Kenyon and Morton (2003), Wu, Liao, and Wang (2005), Choi (2006), Yan, Chi, and Tang (2006), Better, Glover, and Laguna (2007), Chen (2013), Wang and Phan (2014), Yan, Wang, and Chang (2014), Fereiduni and Shahanaghi (2016), Karande, Zavadskas, and Chakraborty (2016), Moghaddam and Mahlooji (2016), Rahafrooz and Alinaghian (2016), and Parnianifard, Azfanizam, Ariffin, and Ismail (2018). The main contribution of these studies is to accommodate parameters containing errors in the prediction or optimization models. After the optimization or prediction, the results are analyzed and compared. A key point that has been ignored is that since the model parameters contain errors, the model solutions naturally contain errors. Thus, decision-making based on the model solutions would be meaningful only if the deviation between the true optimal (but unavailable) solution and the solution based on parameters containing errors is negligible. On the other hand, if the deviation is significant, the analysis and comparison for model results and the following decision-making could be meaningless. In mathematical programming, whether the solution is "optimal" or not can be checked with sensitivity analysis. The model is often solved with the expected values of the parameters containing errors and sensitivity analysis is conducted to see whether the solution changes when model parameters change (Arsham & Oblak, 1995;

* Corresponding author.
E-mail addresses: t320002@cc.ncu.edu.tw (S. Yan), jameschu@ntu.edu.tw (J.C. Chu), jay05080@gmail.com (S.-S. Wang).

http://dx.doi.org/10.1016/j.cie.2017.09.003
0360-8352/

Dave, 1985; Dye & Hsieh, 2013; Glasserman & Tayur, 1995; Ivorra, Mohammadi, & Ramos, 2009; Karimi, Mostoufi, & Sotudeh-Gharebagh, 2014; Kian & Kargar, 2016; Kumar, Rosenberger, & Iqbal, 2016; Perboli, Ghirardi, Gobbato, & Perfetti, 2015). If the "optimal" solution is stable, it is believed that the solution is safe to use. If the "optimal" solution is not stable, parametric programming (if the number of uncertain parameters is small) and sampling (if the number of uncertain parameters is large) are often used to find alternative optimal solutions. Wallace (2000) pointed out that these approaches assume that the true optimal solution inherits properties from alternative optimal solutions generated by parametric programming or sampling. However, Wallace (2000) also demonstrated with very simple examples that this assumption is generally false and thus the standard methods of sensitivity analysis are of little use in finding true optimal solutions for optimization problems with uncertain parameters. Unlike the aforementioned methods which solve for potentially wrong optimal solutions when errors exist in the model parameters, this study aims to answer the following question:

> "When errors exist in model parameters, what is the solution error between the true optimal solution based on error-free parameters and the solution based on parameters containing errors?"

This is analogous to the role of central limit theorem in the parameter estimation in statistical inference. When the mean of a population is estimated based on a random sample, the true value of the population mean is never known but the central limit theorem allows us to evaluate the potential error of the sample mean. Similarly, answering the above question allows the decision makers to evaluate the quality of the solution when the errors in model parameters are inevitable. To the best of the authors' knowledge, the question has never been studied.

There are two major causes for the uncertainty in model parameters. The first cause of uncertainty in model parameters comes from the imperfect information and understanding of the real world. The second cause comes from the techniques or equipment for collecting data that are used to calculate parameter values. In general, the first cause of uncertainty is uncontrollable, which is usually called *random error*. The second cause is controllable, which is called *systematic error*. These two causes of error cannot be completely avoided in the process of experimental measurements (Weltner et al., 2009). In this study, the impact of the two causes of parameter errors on the model solution error will be studied. In addition, optimality tolerances are usually adopted in solution algorithms for optimization models such as the branch-and-bound algorithm to increase the solution efficiency (Parpinelli, Lopes, & Freitas, 2002; Tam, Taniar, & Smith, 2002; Lee & Tong, 2011; Shanmugasundari & Ganesan, 2013; William, Asunción, & Antonio, 2014). In general, a small value for the optimality tolerance leads to a larger solution time and vice versa. Therefore, selecting an appropriate optimality tolerance that reaches an acceptable balance of solution quality and solution time is beneficial. The effect of optimality tolerances on the model solution error will also be studied along with uncertain parameters.

This study proposes an experimental method to answer the aforementioned question empirically with a case study of a project scheduling model. The uncertain parameters of a mathematical programming model can be separated into two parts: those for the objective function and those for the constraint set. Based on mathematical programming theory, the uncertain parameter values included in the objective function have a more direct influence on the model solution than those included in the constraint set. Thus, this study focuses on the uncertain parameters in the objective function. Different scenarios of systematic error, random error, and optimality tolerances are generated to understand their impact on model solution errors. In addition, regression analysis is performed to obtain the statistical relationship between the parameter errors, optimality tolerances, and solution errors.

The rest of the article is organized as follows. Section 2 reviews the model of project scheduling problem. Section 3 explains the experimental approach. Sections 4 and 5 are the test results and regression analysis. Section 6 describes the conclusions of the study.

2. Review of the case study model

The project scheduling model proposed in Chen, Yan, Wang, and Liu (2015) is selected as a case study to demonstrate the methodology. Readers are referred to Chen et al. (2015) for the complete details of the model. The relevant part of the model is briefly introduced here. First of all, the notations of variables, parameters, and sets used in the model are defined as follows.

Variables:

x_{ijk}	the flow variable of arc (i,j,k). 1 if the combination of activity/mode/period corresponding to arc (i,j,k) is selected and 0 otherwise.

Parameters:

c_{ijk}	the present value of net cash flow of activity arc (i,j,k);
I_k	the cash inflows for activity/mode/period combination *k;*
E_k	the cash outflows for activity/mode/period combination *k;*
$finish(k)$	the finishing time for activity/mode/period combination *k;*
$start(k)$	the starting time for activity/mode/period combination *k;*
α	the interest rate;
r_{ijkl}	the amount of the l^{th} renewable resource required by arc (i,j,k) in the network;
r_{ijko}	the amount of the o^{th} non-renewable resource required by arc (i,j,k) in the network;
a_l	the available amount of the l^{th} renewable resource;
b_0	the available amount of the o^{th} non-renewable resource;
s_{qi}	the number of predecessors of the node pair (q,i) in the network;
u_{qik}	the flow adjustment coefficient of arc (q,i,k) in the network;
v	a supply point in the network;
f	a collection point in the network;
p_{ijk}	the flow upper bound associated with arc (i,j,k) in the network.

Sets:

A_{ij}	the set of all parallel activity arcs of the node pair (i,j);
N	the set of all nodes in the network;
C	the set of all activities in the network;
W	the set of all time-precedence node pairs of activities;
W_a	the set of all node pairs of the a^{th} activity in the network;
B_{qi}	the set of arcs of all predecessors of the node pair (q,i) in the network;
SR	the set of all kinds of renewable resources;
NR	the set of all kinds of non-renewable resources;
T_h	the set of node pairs at the h^{th} time point;
T	the set of all time points during the project duration.